



**HAL**  
open science

## Crystal Structure of the RNA Polymerase Domain of the West Nile Virus Non-structural Protein 5

Hélène Malet, Marie-Pierre Egloff, Barbara Selisko, Rebecca E Butcher, Peter J Wright, Michael Roberts, Arnaud Gruez, Gerlind Sulzenbacher, Clemens Vonrhein, Gérard Bricogne, et al.

► **To cite this version:**

Hélène Malet, Marie-Pierre Egloff, Barbara Selisko, Rebecca E Butcher, Peter J Wright, et al.. Crystal Structure of the RNA Polymerase Domain of the West Nile Virus Non-structural Protein 5. *Journal of Biological Chemistry*, 2007, 10.1074/jbc.M607273200 . hal-03160533

**HAL Id: hal-03160533**

**<https://hal.science/hal-03160533v1>**

Submitted on 5 Mar 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Crystal Structure of the RNA Polymerase Domain of the West Nile Virus Non-structural Protein 5<sup>\*S</sup>

Received for publication, August 1, 2006, and in revised form, January 29, 2007 Published, JBC Papers in Press, February 7, 2007, DOI 10.1074/jbc.M607273200

Hélène Malet<sup>†1</sup>, Marie-Pierre Egloff<sup>‡1</sup>, Barbara Selisko<sup>‡</sup>, Rebecca E. Butcher<sup>§</sup>, Peter J. Wright<sup>§</sup>, Michael Roberts<sup>¶</sup>, Arnaud Gruez<sup>‡</sup>, Gerlind Sulzenbacher<sup>‡</sup>, Clemens Vornrhein<sup>¶</sup>, Gérard Bricogne<sup>¶</sup>, Jason M. Mackenzie<sup>||</sup>, Alexander A. Khromykh<sup>||</sup>, Andrew D. Davidson<sup>\*\*</sup>, and Bruno Canard<sup>†2</sup>

From the <sup>†</sup>Architecture et Fonction des Macromolécules Biologiques, CNRS, and Universités d'Aix-Marseille I et II, UMR 6098, ESIL Case 925, 13288 Marseille, France, the <sup>‡</sup>Department of Microbiology, Monash University, Clayton, Victoria 3168, Australia, <sup>¶</sup>Global Phasing Ltd., Sheraton House, Cambridge CB3 0AX, United Kingdom, the <sup>||</sup>School of Molecular and Microbial Sciences, University of Queensland, Brisbane, Queensland 4072, Australia, and the <sup>\*\*</sup>Department of Cellular and Molecular Medicine, University of Bristol, Bristol BS8 1TD, United Kingdom

Viruses of the family *Flaviviridae* are important human and animal pathogens. Among them, the *Flaviviruses* dengue (DENV) and West Nile (WNV) cause regular outbreaks with fatal outcomes. The RNA-dependent RNA polymerase (RdRp) activity of the non-structural protein 5 (NS5) is a key activity for viral RNA replication. In this study, crystal structures of enzymatically active and inactive WNV RdRp domains were determined at 3.0- and 2.35-Å resolution, respectively. The determined structures were shown to be mostly similar to the RdRps of the *Flaviviridae* members hepatitis C and bovine viral diarrhoea virus, although with unique elements characteristic for the WNV RdRp. Using a reverse genetic system, residues involved in putative interactions between the RNA-cap methyltransferase (MTase) and the RdRp domain of *Flavivirus* NS5 were identified. This allowed us to propose a model for the structure of the full-length WNV NS5 by *in silico* docking of the WNV MTase domain (modeled from our previously determined structure of the DENV MTase domain) onto the RdRp domain. The *Flavivirus* RdRp domain structure determined here should facilitate both the design of anti-*Flavivirus* drugs and structure-function studies of the *Flavivirus* replication complex in which the multifunctional NS5 protein plays a central role.

The *Flaviviridae* form a large family of single-stranded positive-sense RNA viruses comprising the three genera *Hepacivirus*, *Pestivirus*, and *Flavivirus*. The genus *Flavivirus* contains more than 80 known arthropod-borne viruses, including major

human and animal pathogens such as dengue virus (DENV),<sup>3</sup> yellow fever virus, Japanese encephalitis virus, and West Nile virus (WNV). Both DENV and WNV are considered as emerging pathogens. Dengue fever is one of the most important mosquito-borne viral diseases in the world, with more than 3 billion people at risk in endemic tropical areas (1). Dengue outbreaks are increasingly severe in terms of cases and fatalities in many regions of the world (2). WNV was discovered in the West Nile district in Uganda in 1937 and was subsequently shown to have an extensive worldwide distribution with the exception of the Americas (1). In 1999, WNV was introduced into the Americas in the New York City area and has since spread throughout the mainland United States, southern Canada, and Mexico. WNV epidemics in the United States have resulted in a total of 23,925 cases of human disease and 946 deaths reported to the Centers for Disease Control (CDC) from 1999 to 2006. WNV consists of 2 lineages (I and II). The North American WNV isolates belong to lineage I, which also includes the Australian subtype Kunjin (3). In contrast to other lineage I WNV strains (4), infections with the Kunjin subtype of WNV do not cause fatal disease in humans (5).

The *Flavivirus* positive sense RNA genome contains a single open reading frame encoding a polyprotein that is processed into three structural and seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5). Signature-sequence analysis suggests that the non-structural protein NS5 is comprised of two domains. The DENV NS5 30-kDa N-terminal domain has been shown to possess RNA cap (nucleoside-2'-O) methyltransferase (MTase) activity, and its crystal structure has been determined (6). That of WNV has been shown to mediate

<sup>\*</sup> This work was supported by SPINE project Grant QLRT-2001-00988, VIZIER integrated project Grant LSHG-CT-2004-511960 of the European Union 6th Framework Programme (FP6), as well as the Conseil Régional de la Région Provence-Alpes-Côte d'Azur. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

<sup>§</sup> The on-line version of this article (available at <http://www.jbc.org>) contains supplemental material and Fig. 1S.

The atomic coordinates and structure factors (code 2HCN, 2HCS, and 2HFZ) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

<sup>1</sup> Both authors contributed equally to this work.

<sup>2</sup> To whom correspondence should be addressed. Tel.: 33-491-82-86-44; Fax: 33-491-82-86-46; E-mail: [bruno.canard@afmb.univ-mrs.fr](mailto:bruno.canard@afmb.univ-mrs.fr).

<sup>3</sup> The abbreviations used are: DENV, dengue virus; aa, amino acids; a/bNLS, nuclear localization sequence that recognize importin- $\alpha$ ; bNLS, nuclear localization sequence that recognize importin- $\beta$ ; BVDV, bovine viral diarrhoea virus; DENV-2, dengue virus type 2; ds, double-stranded; HCV, hepatitis C virus; Imp $\alpha$ / $\beta$ , karyopherin- $\alpha$ / $\beta$  heterodimer, also known as importin- $\alpha$ / $\beta$ ; Imp $\alpha$ , karyopherin- $\alpha$ , also known as importin- $\alpha$ ; Imp $\beta$ , karyopherin- $\beta$ , also known as importin- $\beta$ ; MTase, methyltransferase; NLS, nuclear localization sequence; NS3, non-structural protein 3; NS5, non-structural protein 5; POL, West Nile virus polymerase; POL1, construct 273–905 of West Nile virus polymerase; POL2, construct 317–905 of West Nile virus polymerase (in complex with calcium ion); POL3, construct 273–882 of West Nile virus polymerase; Phi6, *Pseudomonas* phage Phi-6; RdRp, RNA-dependent RNA polymerase; r.m.s., root mean square; RT, reverse transcriptase; WNV, West Nile virus.

both (guanine-N7) and (nucleoside-2'-O) methylation of the 5'-cap structure (7). The C-terminal domain of about 70 kDa harbors the RdRp (RNA-dependent RNA polymerase), as initially identified by the presence of signature-sequence motifs A to F (8, 9), and subsequently confirmed in RdRp activity assays (10–12).

The *Flaviviridae* RdRp is essential to viral replication. The RdRp duplicates the single-stranded RNA genome during a single, continuous polymerization event. The RdRp enters at the 3'-end of the genome and is able to copy the whole RNA molecule in a primer-independent fashion (12–17) referred to as *de novo* RNA synthesis. Crystal structures of RdRps from ~9 virus species have been reported, among them those of *Hepaci*- and *Pestivirus* RdRps (18–22), but none from a *Flavivirus* species.

Very little is known about the mechanism of flaviviral RNA capping, or its co-ordination with RNA synthesis. In eukaryotic cells, mRNA capping is a nuclear event, whereas *Flavivirus* RNA synthesis is thought to occur in the cytoplasm of the infected cell. The viral NS3 and NS5 proteins have been demonstrated to possess RNA-triphosphatase and 2'-O-N7-MTase activities, respectively (6, 7, 23–26), however, the guanylyltransferase activity required for a complete capping reaction is still uncharacterized.

Interestingly, during infection, a proportion of the NS5 and, more recently, of the NS3 protein of a number of flaviviruses has been detected in the nucleus (DENV) or at least the perinuclear region (WNV and Japanese encephalitis virus) of infected cells (27–29). Trafficking of the DENV NS5 to the nucleus was proposed to rely on a nuclear import pathway based on the identification of functional nuclear localization sequences (NLSs) in NS5 and their interaction *in vitro* with distinct members of the karyopherin family of intracellular transport proteins (karyopherin- $\alpha/\beta$ , also known as importin- $\alpha/\beta$  or Imp $\alpha/\beta$ ) (30, 31). NLSs are small motifs composed of one or more clusters of basic amino acids that do not conform to a specific sequence consensus. Two adjacent regions encompassing distinct NLSs have been defined in DENV NS5 termed the bNLS (aa 320 to 368) and the a/bNLS (aa 369 to 405) (30, 31). The bNLS and a/bNLS bind with high affinity to Imp $\beta$  and an Imp $\alpha/\beta$  heterodimer *in vitro*, respectively, and both are capable of targeting  $\beta$ -galactosidase to the nucleus in the context of a fusion protein (30, 31). However, whether the bNLS and a/bNLS are required for nuclear import of the intact NS5 protein during viral infection is not known. The bNLS of DENV NS5 has also been shown to interact with NS3 and competition between Imp $\beta$  and NS3 binding has been confirmed using pull-down assays (30).

Here, we report the crystal structures of enzymatically active and inactive WNV RdRp domains. They reveal a classic RdRp fold bearing palm, thumb, and finger domains decorated with *Flavivirus*-specific attributes such as a priming loop adopting a different fold. Structural elements in the WNV RdRp structure corresponding to putative DENV NLSs were found to be integral parts of the RdRp domain. Interestingly, despite sequence conservation in the NLSs of DENV and WNV NS5 the latter was not found to localize to the nucleus. To gain insight into the overall spatial organization of the full-length NS5 protein, a reverse genetic approach was used to identify amino acids in the

DENV MTase and RdRp domains, which potentially interact and could be used to position one domain relative to the other. The structure of the WNV MTase domain was modeled using DENV MTase atomic coordinates. Both WNV MTase and RdRp domains were assembled *in silico* using docking procedures, generating a model of full-length WNV NS5 protein.

## EXPERIMENTAL PROCEDURES

**Expression and Purification of the Polymerase Domains**—Three WNV strain Kunjin *ns5* gene constructs encoding a N-terminal His<sub>6</sub> tag fused to *pol1* (corresponding to aa 273–905), *pol2* (corresponding to aa 317–905), or *pol3* (corresponding to aa 273 to 882) open reading frames were cloned into the pDEST14 vector, and expressed in *Escherichia coli*. The proteins were expressed and purified to homogeneity as follows: the *E. coli* strain C41pRos, transformed with either plasmid construct, was grown at 37 °C to an  $A_{600}$  of 0.6, induced with 0.5 mM isopropyl  $\beta$ -D-thiogalactopyranoside, and further incubated 16–18 h at 17 °C. Cells were harvested by centrifugation. The cell pellet was resuspended in 50 mM Tris buffer, pH 8.0, containing 150 mM NaCl, 10 mM imidazole, DNase I (2  $\mu$ g/ml), a protease inhibitor tablet (Sigma), and sonicated on ice. The sample was centrifuged, the supernatant collected and filtered through a 0.22- $\mu$ m filter. The sample was applied to a 5-ml bed volume HiTrap nickel immobilized metal ion affinity chromatography column (Amersham Biosciences) connected to a FPLC system (Amersham Biosciences). The protein was eluted with 50 mM Tris buffer, pH 8.0, containing 150 mM NaCl and 500 mM imidazole. Protein-containing fractions were then applied onto a preparative Superdex 200 gel filtration column pre-equilibrated in 10 mM Tris buffer, pH 9.0, with 300 mM NaCl and 5% glycerol. Protein was concentrated to 6 (POL1) and 9 mg/ml (POL2 and POL3) using a Vivaspinn 30-kDa molecular mass cut-off centrifugal concentrator (Vivascience).

A selenomethionine-substituted protein was used for structure determination. It was expressed according to standard conditions of the methionine-biosynthesis pathway inhibition (32) and purified following the same procedure as that for the native protein.

**Crystallization**—Initial crystallization trials were set up with a nano-drop dispenser in 96-well sitting drop plates using commercial crystallization kits and purified protein at 6 mg/ml for POL1 and 9 mg/ml for POL2. Initial hits were further optimized by the hanging drop vapor diffusion method in Linbro plates by mixing 1  $\mu$ l of protein solution with 1  $\mu$ l of reservoir solution. Crystals of POL1 were grown from 10% PEG 1000, 0.2 M sodium cacodylate, 0.3 M MgCl<sub>2</sub>, pH 7.0. POL2 was crystallized using two conditions (i) 5% PEG 10,000, 0.2 M imidazole malate, pH 8.0, and (ii) 10% PEG 8000, 0.1 M imidazole pH 7.0, and 0.2 M calcium acetate.

Crystals were briefly soaked in a cryo-protectant solution composed of mother liquor supplemented with 36% glycerol and flash frozen in liquid nitrogen. WNV POL1 and POL2 diffraction intensities were recorded on different beamlines (Table 1) at the European Synchrotron Radiation Facility (Grenoble, France). Integration, scaling, and merging of the intensities were carried out using programs from the CCP4 suite (33) and statistics are provided in Table 1.

**TABLE 1**  
Crystallographic data and refinement statistics

	POL2			POL1, magnesium
	SAD (selenomethionine)	Native	Calcium	
Construct (aa)	317–905	317–905	317–905	273–905
<b>Data collection</b>				
Space group	P4 <sub>3</sub>	P4 <sub>3</sub>	P4 <sub>3</sub>	I222
Cell dimensions (Å)	<i>a</i> = <i>b</i> = 110.0 <i>c</i> = 67.7	<i>a</i> = <i>b</i> = 110.0 <i>c</i> = 69.1	<i>a</i> = <i>b</i> = 110.1 <i>c</i> = 68.6	<i>a</i> = 73.3 <i>b</i> = 103.4 <i>c</i> = 190.1
X-ray source	ESRF ID23-1	ESRF ID14-EH1	ESRF ID14-EH3	ESRF BM14
Wavelength (Å)	0.9795	0.934	0.933	0.9763
Resolution range (Å)	35.0–2.8 (2.95–2.8)	35.0–2.5 (2.65–2.5)	35.0–2.35 (2.48–2.35)	21.77–3.0 (3.18–3.0)
Total reflections	237,932 (34,735)	163,928 (24,051)	194,222 (17,177)	67,265 (10,835)
Unique reflections	20,157 (2934)	28,709 (4172)	34,348 (4930)	14,328 (2140)
Completeness (%)	100 (100)	99.8 (100)	99.9 (99.7)	96.4 (100)
$\langle I/\sigma(I) \rangle$	23.1 (4.5)	18.4 (2.9)	21.8 (1.5)	12.9 (3.4)
$R_{\text{sym}}^a$	0.106 (0.560)	0.069 (0.567)	0.055 (0.567)	0.097 (0.337)
Multiplicity	11.8 (11.8)	5.7 (5.8)	5.7 (3.5)	4.7 (5.1)
Anomalous completeness (%)	99.9 (99.9)			
Anomalous multiplicity	6.1 (6.1)			
<b>Refinement</b>				
$R_{\text{work}}$ (%) <sup>b</sup>		20.6	21.0	26.0
$R_{\text{free}}$ (%) <sup>c</sup>		24.3	25.9	26.5
<b>Number of atoms</b>				
Protein		3,937	3,932	4,915
Water molecules		109	179	12
Ion in the non-catalytic position near the active site		0	1 (Calcium)	1 (Magnesium)
<b>r.m.s. deviations</b>				
Bond lengths (Å)		0.022	0.021	0.007
Bond angles (°)		1.7	1.7	1.2
<b>Ramachandran analysis (%)</b>				
Most favored		91.0	91.9	85.4
Additionally allowed		8.5	7.6	12.4
Generously allowed		0.5	0.5	1.5
Disallowed		0	0	0.7
Sequence assigned to model (aa)		322–337, 362–410, 420–452, 473–576, 603–747, 752–891	322–338, 363–410, 420–452, 474–576, 604–747, 752–892	274–409, 416–458, 471–899
PDB code		2HCS	2HCN	2HFZ

<sup>a</sup>  $R_{\text{sym}} = \sum |I - \langle I \rangle| / \sum I$ , where *I* is the observed intensity and  $\langle I \rangle$  is the average intensity. Values in parentheses refer to the highest resolution shell.

<sup>b</sup>  $R = \sum ||F_o| - |F_c|| / \sum |F_o|$ .

<sup>c</sup>  $R_{\text{free}}$  is calculated as  $R$ , but on 5% of all reflections that are never used in crystallographic refinement.

**Structure Determination**—The structure of WNV POL2 was solved using single wavelength anomalous dispersion data collected at the peak of the selenium absorption edge from a selenomethionine-derivatized crystal. Location of 16 selenium atoms (of the 21 expected) was performed using SHELXD (34). Phases and figures of merit were calculated using SHELXE (35) and SHARP (36). Density modification was performed using RESOLVE (37) and SHARP. About 40% of the model was built in an automatic fashion with RESOLVE and completed manually using COOT (38), with the RdRp structures of HCV and BVDV as guides for connectivity (Protein Data Bank codes 1NB6 and 1S49). The model was refined against two data sets using REFMAC (39): the first, which extends to 2.5-Å resolution, was called “native” as it contains no ion near the active site. The second extends to 2.35-Å resolution and was called “calcium” as a calcium ion was found in a non-catalytic position near the active site (Table 1). The calcium resulting model was subsequently used to solve the structure of POL1 by the molecular replacement method using AMORE (40). Refinement of POL1 was initially performed using simulated annealing as implemented in CNS (41) at 3.0-Å resolution (Table 1). Electron density corresponding to the extra 44 N-terminal amino acids of POL1 (compared with POL2) and to some disordered loops of POL2 became visible. Attempts to model the missing parts were performed using several rounds of manual building

and refinement using CNS (41). BUSTER-TNT (42) was then used, and allowed further completion of the model, particularly for residues 363–367, for which no density had previously been visible. Finally, POL1 was refined using BUSTER-TNT under harmonic restraints to much of the POL2 calcium structure, to avoid “un-refinement” of the POL2 substructure against the lower-resolution POL1 data as detailed in “Supplemental Material.” The *R*-factors and geometry were improved by this harmonically restrained refinement. Refinement statistics are listed in Table 1.

**RdRp Activity Tests**—RdRp activity tests were performed in 50- $\mu$ l reactions containing 50 mM HEPES, pH 8.0, 5 mM dithiothreitol, 10 mM KCl, 5 mM MnCl<sub>2</sub>, 5 mM MgCl<sub>2</sub>, 10  $\mu$ M GTP (0.01  $\mu$ Ci of [<sup>3</sup>H]GTP per  $\mu$ l, 6.1 Ci/mmol, Amersham Biosciences) using 1  $\mu$ M poly(rC) (Amersham Biosciences) and WNV RdRp (264 nM POL1, 760 nM POL2, and 800 nM POL3). Reactions were started with a premix of MnCl<sub>2</sub>, MgCl<sub>2</sub>, and GTP and incubated at 30 °C. Samples of 12  $\mu$ l were taken after 5, 10, and/or 15 min and the reaction stopped by adding 30  $\mu$ l of 50 mM EDTA in 96-well plates. Samples were then transferred to glass fiber filter mats with DEAE active groups (DEAE filter mat, Wallac) using a Filtermat Harvester (Packard Instruments). Filtermats were washed three times with 0.3 M ammonium formate, pH 8.0, twice with water, once with ethanol, dried, and the filter transferred into sample bags. Liquid scin-



TABLE 2

## Structural comparison of WNV POL1 with known crystal structures of viral polymerases

WNV palm domain (aa 499–541 and 610–717), fingers domain (aa 274–498 and 542–609), and thumb domain (aa 718–899) were superimposed onto equivalent domains of crystal structure models of viral polymerases. Polymerase structural models used for these calculations are hepatitis C virus (*Flaviviridae* and *Hepacivirus*, PDB code 1NB6), bovine viral diarrhea virus (*Flaviviridae* and *Pestivirus*, PDB code 1S49), foot and mouth disease virus (*Picornaviridae* and *Aphthovirus*, PDB code 1WNE), human poliovirus 1 (*Picornaviridae* and *Enterovirus*, PDB code 1RA7), rabbit haemorrhagic disease virus (*Caliciviridae* and *Lagovirus*, PDB code 1KHV), norwalk virus (*Caliciviridae* and *Norovirus*, PDB code 1SH3), mammalian orthoreovirus 3 (*Reoviridae* and *Orthoreovirus*, PDB code 1N1H), pseudomonas phage Phi-6 (*Cystoviridae* and *Cystovirus*, PDB code 1HI0), HIV-1 reverse transcriptase (*Retroviridae* and *Alpharetrovirus*, PDB code 1RTD). Among several HIV-1 reverse transcriptase structural models available, 1RTD was chosen because it is in a closed conformation resembling that of viral RdRps. r.m.s. deviations were calculated between C $\alpha$  atoms on matched residues. The number of matched atoms used for the calculation of the r.m.s. deviations are also indicated. The percentage of structural identity is calculated as a fraction of pairs of identical residues among all aligned.

PDB code	Palm domain			Fingers domain			Thumb domain		
	R.m.s. deviation	Number of matched aa	% of structural identity	R.m.s. deviation	Number of matched aa	% of structural identity	R.m.s. deviation	Number of matched aa	% of structural identity
1NB6	1.97	104	19.2	2.71	178	18.5	4.20	120	11.7
1S49	2.14	112	21.4	2.98	178	14.6	2.46	116	11.2
1WNE	2.13	94	16	2.81	181	7.2	3.22	72	8.3
1RA7	2.26	106	14.2	2.83	168	12.5	3.76	78	6.4
1KHV	2.03	106	12.3	3.43	199	10.6	3.65	80	2.5
1SH3	2.16	118	15.3	3.17	193	12.4	3.76	76	7.9
1N1H	2.82	97	7.2	3.76	156	9	3.43	60	1.7
1HI0	3.00	102	11.8	3.85	177	7.3	4.40	64	6.2
1RTD	3.06	80	10	2.11	32	12.5	4.73	60	6.7

tillation fluid was added and incorporation was measured in counts per minute using a Wallac MicroBeta TriLux Liquid Scintillation Counter.

**Introduction of NS5 Mutations into a Genomic Length DENV cDNA Clone**—Mutations encoding changes in NS5 amino acids were introduced into the genomic length DENV serotype 2 strain New Guinea C (DENV-2) cDNA clone pDVWS601, which yields the virus v601 (43, 44). Mutations were initially introduced into subgenomic PCR fragments by overlap extension PCR using mutagenic primers (details available from the authors upon request) and then transferred into pDVWS601 as follows. The mutation K46A/R47A/E49A (each NS5 amino acid targeted for mutagenesis is numbered followed by the substituted amino acid) was engineered into a 1001-bp overlap extension PCR fragment (DENV-2 nucleotides 7165–8165), which was HpaI<sup>7406</sup>/StuI<sup>7874</sup> digested and introduced into the corresponding sites of pDVWS601 to produce pDVWS601-NS5<sub>K46A,R47A,E49A</sub>. The NS5 mutation L512V was engineered into a 2077-bp fragment (DENV-2 nucleotides 8085–10161), which was AatII<sup>8570</sup>/MluI<sup>9732</sup> digested and introduced into the corresponding sites of pDVWS601 and pDVWS601-NS5<sub>K46A,R47A,E49A</sub> to produce pDVWS601-NS5<sub>L512V</sub> and pDVWS601-NS5<sub>K46A,R47A,E49A,L512V</sub>, respectively.

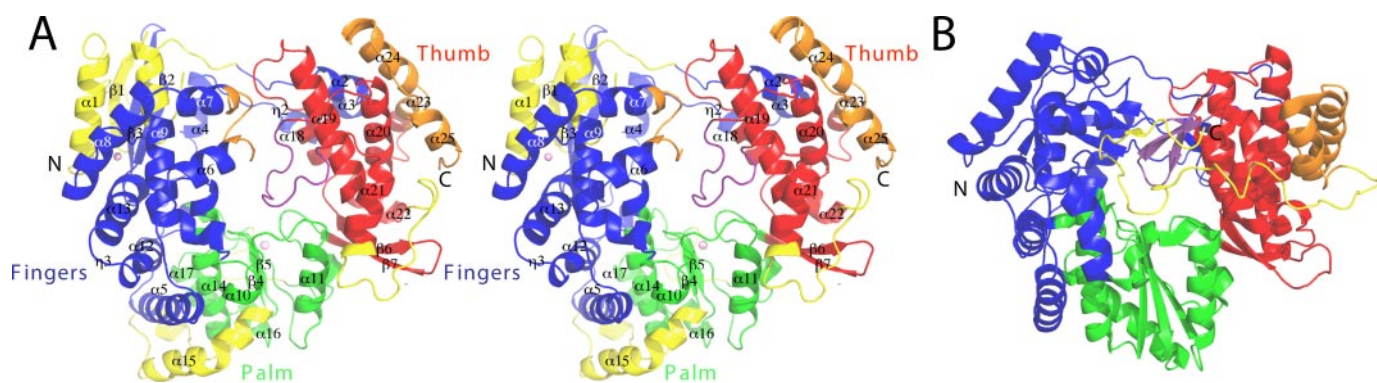
**Recombinant DENV Recovery, Growth, and Sequencing**—Procedures for the recovery of infectious DENV-2 from pDVWS601 including transcription of RNA, electroporation of BHK-21 cells, and passaging and plaque assay of virus in *Aedes albopictus* C6/36 cells have been described previously (43, 44). Confirmation of the mutations in the recovered viruses and the identification of second site mutations were performed by sequencing the entire DENV-2 genome. Viral RNA was extracted from the culture supernatants of infected cells using a QIAamp Viral RNA Extraction kit (Qiagen) and used for the production of six overlapping 2-kilobase pair RT-PCR products spanning the DENV-2 genome using DENV-2-specific primers and the SuperScript One-Step RT-PCR with Platinum Taq System (Invitrogen). Each RT-PCR product was then purified using a Qiagen PCR Purification Kit and used for automated DNA sequencing using DENV-2-specific primers.

**Molecular Modeling and Docking**—Alignment of the MTase sequences of DENV-2 (PDB code 1L9K) and WNV was performed using SPDBV (45) (sequences share 59% identity and 71% similarity). The homology model of WNV MTase was built using the SWISS-MODEL server (45) and stereochemistry was verified using WHATCHECK (46). The docking of the MTase and the POL1 domain was carried out with the protein-protein docking algorithm FTDOCK (47), based on shape complementarities (using a grid spacing of 1 Å) and electrostatic treatment. A search over the complete binding space for both molecules was performed. Results were filtered using the distance restraints between putatively interacting residues as determined by reverse genetic experiments (6 Å filter) and according to the relative positions of the MTase C terminus and POL1 N terminus (45 Å filter for the 13 missing residues), which led to a single solution.

## RESULTS AND DISCUSSION

**Protein Crystallization and Structure Determination**—Crystals of the full-length WNV strain Kunjin NS5 protein could not be obtained despite numerous attempts. However, when a N-terminal His<sub>6</sub> tag was fused to a ns5 polymerase domain coding for proteins lacking either 272 (POL1) or 316 (POL2) N-terminal amino acids, crystals were obtained. The weak identity of the WNV RdRp with existing RdRps structures (maximum of 21%, Table 2) precluded the use of molecular replacement for structure determination. The structure of POL2 was then determined by single wavelength anomalous dispersion of selenomethionylated protein, and the structure of POL1 was solved by molecular replacement using POL2 as a search model. For discussing features common to POL1 and POL2, the term POL will be used. POL has a roughly spherical shape with overall dimensions of 70 × 67 × 50 Å and adopts a “right-hand” overall structure, consisting of fingers ( $\alpha$ 1 to  $\alpha$ 9 from aa 274 to 498, and  $\alpha$ 12 to  $\beta$ 3 from aa 542 to 609), palm ( $\alpha$ 10 and  $\alpha$ 11 from aa 499 to 541, and  $\alpha$ 14 to  $\alpha$ 17 from aa 610 to 717), and thumb ( $\beta$ 6 to  $\alpha$ 25 from aa 718 to 905) domains (Figs. 1A, 2B, and supplemental Fig. 1S). The most closely related structures are those of RdRps from members of the flaviviridae family, the hepacivirus hepatitis C virus (HCV) RdRp (18,

## Crystal Structure of West Nile Virus RNA Polymerase



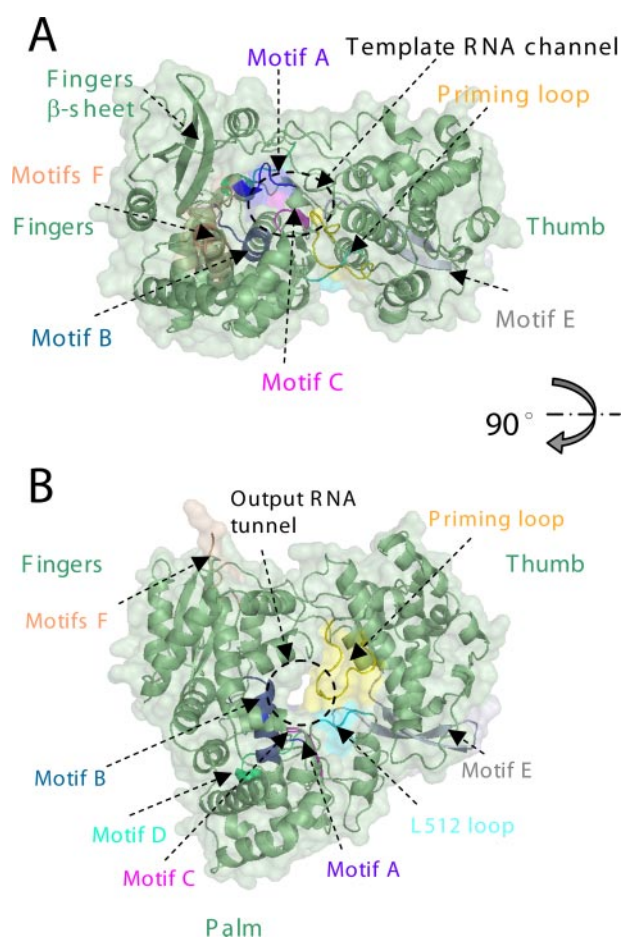
**FIGURE 1. Crystal structure of WNV POL1 and comparison with HCV RdRp.** *A*, stereo view of a ribbon representation of WNV POL1 in its front orientation. The palm, thumb, and fingers domains and the priming loop are colored in green, red, dark blue, and purple, respectively.  $\alpha$ -Helices and  $\beta$ -sheets are indicated. Insertions in WNV POL1 compared with HCV RdRp are displayed in yellow, and major structural differences are shown in orange. These and other figures were prepared with PyMOL. *B*, ribbon representation of HCV RdRp in its front orientation (50) (PDB code 1NB6). The color code is the same as in *A*. Insertions in HCV RdRp compared with WNV are colored in yellow.

19, 21), and the pestivirus bovine viral diarrhea virus (BVDV) RdRp (20), which have structural identities as low as 11–21%, depending on the domain considered (Table 2).

**Overall Structure of WNV POL and Comparison with HCV and BVDV RdRps**—As determined for other RdRps, which initiate RNA synthesis *de novo*, such as the HCV RdRp (Fig. 1*B*), WNV POL displays a closed conformation. The active site is located at the intersection of two tunnels, and is completely encircled by several loops protruding from both the finger and thumb domains (compare Fig. 1, *A* and *B*). Structural superimposition with the *Pseudomonas* phage Phi-6 (Phi6) RdRp initiation complex (PDB code 1HI0) (48) and the human immunodeficiency virus reverse transcriptase in complex with a double-stranded DNA template primer and a nucleotide (PDB code 1RTD) (49) suggests a role for the tunnels. The first tunnel, located at the interface between the thumb and the fingers domains, should allow access of the template single-stranded RNA to the catalytic site (Fig. 2*A*). The second tunnel is roughly perpendicular to the first and goes across the entire protein (labeled *output RNA tunnel*, Fig. 2*B*). Incoming NTPs access the active site via the back aperture of this tunnel. Nascent dsRNA potentially exits via the front part of this tunnel. However, product release would presumably require a conformational change of the loops encircling the active site, as in the present conformation, steric clashes would preclude the accommodation of a dsRNA molecule in the output RNA tunnel (Fig. 2*B*).

**The Palm Domain**—The architecture of the palm domain is a highly conserved feature among all RdRps structurally characterized so far. The WNV POL is no exception, with a r.m.s. deviation (on topologically equivalent C $\alpha$  atoms of the palm domain) of 2.14 and 1.97 Å relative to the BVDV (PDB code 1S49) (20) and HCV RdRps (Fig. 1*B*, PDB code 1NB6) (50), respectively. The palm consists of a central anti-parallel  $\beta$ -sheet ( $\beta$ 4 and  $\beta$ 5) flanked at either side by one ( $\alpha$ 11) and five  $\alpha$ -helices ( $\alpha$ 10 and  $\alpha$ 14 to  $\alpha$ 17) (Fig. 1*A*).

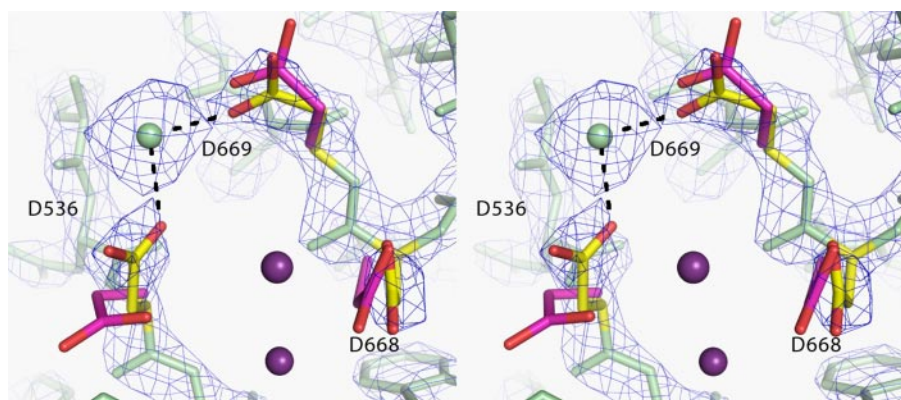
A major difference between the WNV POL palm domain and those of other RdRps is that the  $\beta$ -sheet consists of only 2  $\beta$ -strands instead of 3, as the five residues preceding motif A are not organized in a typical  $\beta$ -strand. Moreover, these 2  $\beta$ -strands are very short ( $\sim$ 10 Å) compared with those of



**FIGURE 2. Motifs and putative RNA input/output channels of WNV POL1.** *A*, top view of WNV POL1 (in green), in which the template RNA channel is indicated. Motifs A, B, C, D, E, and F and the priming loop are indicated in deep purple, blue, purple, green-blue, gray, wheat, and yellow, respectively. The Leu<sup>512</sup> loop, which putatively interacts with the Lys<sup>46</sup>/Arg<sup>47</sup>/Glu<sup>49</sup> loop in the MTase domain (see Fig. 7), is shown in cyan. *B*, front view of *A* (rotated by 90° when compared with *A*), showing the output dsRNA channel.

HCV and BVDV ( $\sim$ 20 Å). In addition, for WNV POL, an extra loop/ $\alpha$ -helix ( $\alpha$ 15) (aa 629 to aa 663) is found in WNV POL in an unusually long insertion connecting motif B to C (Fig. 1*A* and supplemental Fig. 1*S*). This insertion is either





**FIGURE 3. Divalent ion binding site in WNV POL.** Stereo view of the calcium/magnesium ion non-catalytic binding site. The POL2 model is represented in *green sticks* and the corresponding electronic density in *blue*. The  $\text{Ca}^{2+}$  ion is shown as a *green sphere*. The figure is centered on the aspartic acids of motifs A and C colored in *yellow*. Coordination with Asp<sup>536</sup> (motif A) and Asp<sup>669</sup> (motif C) are indicated by *black dotted lines*. Corresponding aspartic acids of motifs A and C in Phi6 (Asp<sup>324</sup>, Asp<sup>453</sup>, and Asp<sup>454</sup>) (PDB code 1HI0), are represented in *magenta sticks*. The position of the two ions in the catalytic position, inferred from the Phi6 RdRp structure, is indicated in *purple*.

absent in the HCV RdRp or quite variable in sequence, length, and conformation in all other RdRp structures. In WNV POL, it packs against an  $\alpha$ -helix ( $\alpha 5$ ) located at the base of the fingers (Fig. 1A).

RdRp palm domains contain three sequence motifs, involved in nucleotide binding and phosphoryl transfer, known to play an essential functional role in RNA synthesis. In particular, motifs A (aa 536 to 541 in WNV POL) and C (aa 667 to 671 in WNV POL) contain three conserved aspartic acids. By structural superimposition with other RdRps, these aspartic acids are identified in WNV POL as Asp<sup>536</sup> (in motif A) and Asp<sup>668</sup> and Asp<sup>669</sup> (in motif C). The role of aspartates corresponding to Asp<sup>536</sup> and Asp<sup>668</sup> in the catalytic mechanism of DNA/RNA polymerases has been well characterized (48, 51). The first aspartate acts as a general base and deprotonates the primer 3' hydroxyl, which subsequently attacks the nucleotide  $\alpha$ -phosphate. The second promotes a suitable geometry for the catalytic reaction. In structurally characterized RdRps, aspartic acids corresponding to Asp<sup>536</sup> and Asp<sup>668</sup> are known to bind two metal ions involved in catalysis. Our search for crystallization conditions suitable for x-ray analysis led to three different buffers. The first contained no divalent ion (native data set), the second 0.3 M  $\text{MgCl}_2$  ("magnesium" data set), and the third 0.2 M calcium acetate (calcium data set). The crystal structure of the WNV POL is devoid of metal ions in catalytic positions. Nevertheless, in the last two conditions only, a metal ion (magnesium and calcium, respectively) could clearly be identified in the same position in the electron density maps. This position is 5.0 and 7.1 Å away from the expected positions of the two catalytic metal ions. Due to the distance from the active site, it is unlikely that it participates in the catalytic reaction (Fig. 3). It is coordinated by Asp<sup>536</sup> and Asp<sup>669</sup>; and consequently, Asp<sup>536</sup> points away from the catalytic position. A similar geometry for a non-catalytic ion and the aspartate equivalent to Asp<sup>536</sup>, with concomitant lack of catalytic metal ions, has been described for other RdRps (51–54). The role of this non-catalytic ion is still elusive. The non-catalytic ion was proposed to be present in an inactive open conformation of rabbit hemorrhagic disease virus RdRp, which affects the active site conformation (52). Never-

theless, WNV POL corresponds to the closed, active conformation of rabbit hemorrhagic disease virus RdRp, which does not have an ion in the non-catalytic position (52) as judged by superimposition of both structures (not shown). Additionally, in Phi6 RdRp (PDB code 1HI0), which adopts also a closed, active conformation, the presence of a non-catalytic ion was shown to be compatible with the binding of ions in the catalytic positions and the formation of an initiation complex. In this case the residue corresponding to Asp<sup>536</sup> rotates toward the catalytic ion position and the residue equivalent to Asp<sup>669</sup> coordinates the non-catalytic ion. It was pro-

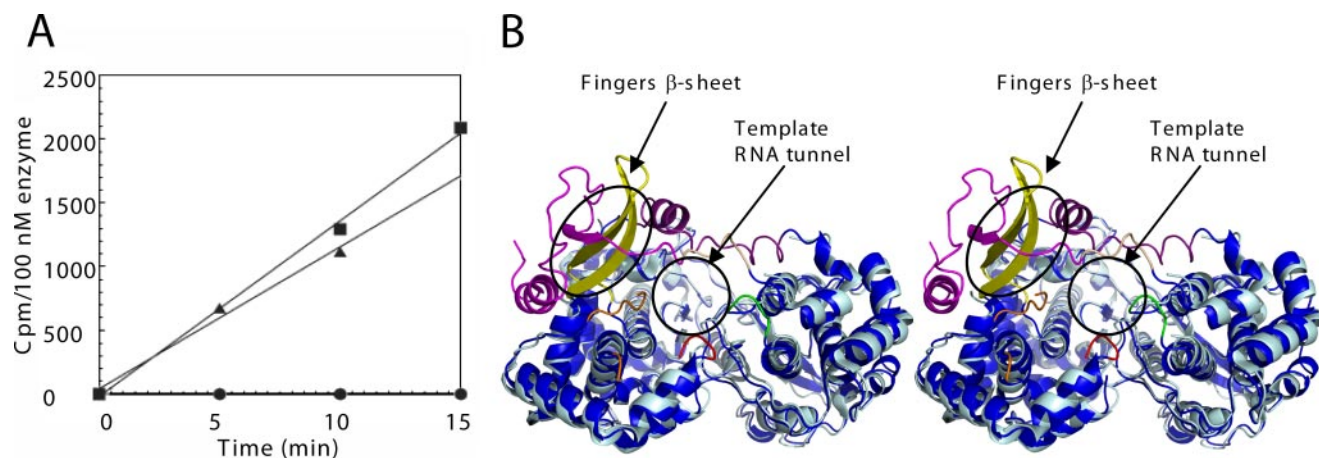
posed that the non-catalytic ion plays a role in the *de novo* initiation mechanism facilitating the movement of the nascent dsRNA after formation of the first dinucleotide out of the active site (48). Further studies have to be undertaken to elucidate the role of the non-catalytic ion for WNV RdRp but we suppose that for the conformation of the WNV POL1 active site observed here, a simple change of the Asp<sup>536</sup> side chain orientation might allow binding of catalytic ions, building of the ternary complex, and RNA synthesis.

**The Fingers Domain**—Comparison of the fingers domain of WNV POL1 (aa 274 to 498 and 542 to 609) to the corresponding domains of the HCV and BVDV RdRps revealed major differences. First, there is an extra N-terminal stretch (aa 274 to 318), which is comprised of an  $\alpha$ -helix ( $\alpha 1$ ) and a  $\beta$ -strand ( $\beta 1$ ) (Fig. 1A in *yellow*).

Second, the loop connecting  $\alpha$ -helices  $\alpha 6$  and  $\alpha 7$  (aa 407 to 420) is longer and adopts a conformation distinct from that of all RdRps structurally characterized so far (Fig. 1A, in *orange*). In primer-dependent RdRps, this loop corresponds to the conserved motif G located at the entrance of the RNA template-binding groove (55), and establishes contacts with the template (53). In HCV and BVDV RdRps, the position of this loop is relatively conserved although the motif G is not. In WNV POL, motif G is not conserved either but, in contrast, the corresponding loop takes an entirely different direction protruding toward the active site and the thumb domain priming loop (see below). Thus it seems to contribute to closure of the active site. Interestingly, its tip is partially disordered (aa 410–415 could not be modeled due to lack of electron density), which suggests conformational flexibility. Its location on the structure seems to coincide with the position of the C termini of the BVDV and HCV RdRps. The C terminus of HCV RdRp has been shown to play a regulatory role for RdRp activity (56, 57). It prevents RNA binding and possibly also the binding of incoming NTPs. The loop connecting  $\alpha$ -helices  $\alpha 6$  and  $\alpha 7$  might play a similar regulatory role in *Flavivirus* RdRps.

Third, the motif F fingertip loop corresponding to a motif named F2 (58) between motifs F1 and F3 (see supplemental Fig. 1S) is partially disordered in the WNV POL (aa 459–470 are

## Crystal Structure of West Nile Virus RNA Polymerase



**FIGURE 4. WNV POL polymerase activity and structural difference between the inactive construct POL2 and the active construct POL1.** *A*, time course of [ $^3\text{H}$ ]GMP incorporation into RNA by WNV POL (POL1, POL2, and POL3) using a homopolymeric RNA template poly(rC). 264 nM POL1 (■), 760 nM POL2 (●), and 800 nM POL3 (▲) were used. Absolute [ $^3\text{H}$ ]GMP incorporation in counts per minute for each enzyme was normalized to 100 nM enzyme concentration. *B*, top view (rotated by  $90^\circ$  when compared with Fig. 1A) of the superimposition of WNV POL1 and POL2. Amino acids visible in the electronic density in both POL1 and POL2 are represented in blue and cyan, respectively. Amino acids that are present in POL1 but absent from POL2 due to the shortening of the construct (aa 273–316) are colored magenta. There are amino acids ordered and visible in POL1, present in the POL2 sequence, but invisible in POL2 electronic density maps. These belong to loops comprising aa 317–321, 339–361, 416–419, 453–458, and 471–472, 577–602, and 748–751, which are represented in wheat, purple, red, orange, yellow, and green, respectively. The unusual 3-stranded  $\beta$ -sheet (made of strands  $\beta$ 1,  $\beta$ 2, and  $\beta$ 3, see Fig. 2A) that stabilizes the whole “back” subdomain of the fingers domain in POL1 is indicated. The template RNA tunnel is also shown.

missing in the density map). Surprisingly, the ordered part (aa 455–458 and 471–476) does not adopt a  $\beta$ -sheet structure as in all other RdRp structures, but F3 appears as an  $\alpha$ -helix. The structure POL1 clearly indicates that the structural elements corresponding to motif F run perpendicular to the motif F  $\beta$ -sheet structure observed in other RdRps, thus parallel to the fingers  $\beta$ -sheet (in WNV POL1  $\beta$ 2 and  $\beta$ 3 combined with  $\beta$ 1 of the N-terminal extension) and exposed to solvent (Fig. 2B). How this unusual conformation relates to the replication mechanism of *Flavivirus* RdRp remains to be elucidated. The basic residues of motif F, especially F2 and F3, have been shown to provide a binding site (named the interrogation site I) for incoming NTPs entering through the tunnel from the back of the molecule (18, 20, 21, 48, 51, 53, 59). The residues involved in NTP binding are conserved in motif F of WNV POL. Should they play the same role during *Flavivirus* RNA synthesis, a conformational change would be necessary to create the binding site in WNV POL. Thus it is expected that upon formation of the ternary initiation complex, a major conformational change of the structural elements constituting motif F of the *Flavivirus* RdRp will take place.

In conclusion, the fingers domain contains elements that, in the putative pre-initiation conformation observed here, are substantially different from that of HCV and BVDV RdRps. A concerted conformational change of both motif F and the loop corresponding to motif G might be necessary to allow RNA synthesis.

**The Thumb Domain**—The thumb domain encompasses the C-terminal 187 aa of WNV POL (aa 718–905). Structural motif E, which consists of two anti-parallel  $\beta$ -strands ( $\beta$ 6 and  $\beta$ 7, aa 714–736), forms the interface between the thumb and the palm domain (Fig. 1A). The rest of the thumb consists of eight  $\alpha$ -helices connected by large loops. Its overall topology is very different from that of the HCV RdRp thumb domain, and is more related to that of BVDV RdRp. The five N-terminal  $\alpha$ -helices of the WNV POL thumb domain can be superimposed on the

corresponding  $\alpha$ -helices of BVDV RdRp and the connectivity between them is conserved. However, the length, conformation, and orientation of the structural elements connecting these  $\alpha$ -helices are very different. Moreover, the three C-terminal  $\alpha$ -helices of WNV POL cannot be superimposed onto the corresponding  $\alpha$ -helices of either the HCV or BVDV RdRps, as there is a 45 degree difference in their orientation. The overall structural diversity of the thumb domain is even more salient when the WNV POL is compared with primer-dependent RNA polymerases, like *Picornavirus* and *Calicivirus* RdRps, which have a much smaller thumb domain (52, 53, 59, 60). To date, two structural elements have been correlated with the difference between primer-dependent and *de novo* RNA synthesis by RdRps: (i) the presence of a characteristic loop (also called the priming loop,  $\beta$ -hairpin in HCV,  $\beta$ -thumb region in BVDV, and critical loop region in Phi6 RdRps) and (ii) the size, overall fold, and location of the C-terminal segment. For example, in the HCV RdRp, the 40 residues upstream of the hydrophobic C-terminal 21-amino acid stretch fold back toward the active site cavity and the RNA binding groove constituting a regulator for RNA synthesis (56, 57). Several residues within this stretch contact the  $\beta$ -hairpin loop, which serves as a priming platform for initiation and prevents back-primed initiation (61). The situation is similar for BVDV NS5B, even though the structure of the C-terminal upstream stretch has not been fully modeled. In WNV POL, the C-terminal six amino acids are not defined in either of the two space groups in which the crystal structures of POL1 and POL2 have been solved. However, the last modeled amino acid, Leu<sup>899</sup>, is more than 40 Å away from the active site, clearly indicating that the C terminus of WNV POL cannot fold back into the active site. Moreover, there is no difference in RdRp activity between WNV POL1 (aa 273–905) and a 23-aa C-terminal truncated form (POL3, aa 273–882) (Fig. 4A), confirming that the C-terminal extremity of *Flaviviruses* does not play a regulatory role in the polymerization process, as it is the case for *Pestivirus* and *Hepaciviruses*. Our structure proposes that



the element that might overtake this function in *Flavivirus* RdRps may be the loop connecting  $\alpha$ -helices  $\alpha 6$  and  $\alpha 7$  (see above).

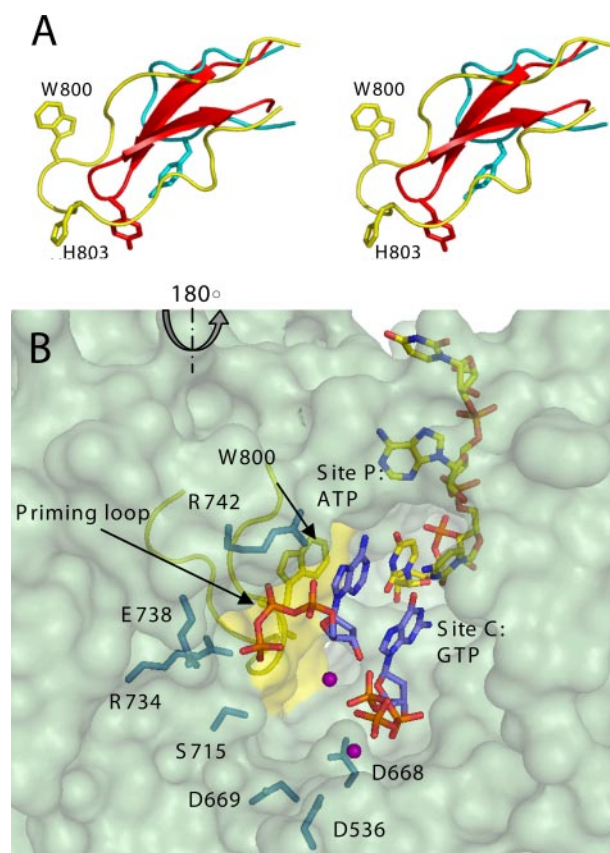
**The Interdomain Region between the MTase and POL Domains**—As proposed previously (12), the region corresponding to aa 322–407 does not form an interdomain linker between the MTase and POL domains, but belongs to the POL domain itself. Rather, this study and our previous structural characterization of DENV NS5 MTase, designate amino acid stretch 260–270 as the interdomain linker. Interestingly, expression and purification of full-length NS5 from several flaviviruses resulted in the identification of a proteolytic fragment, starting at amino acid 265 (identified by N-terminal amino acid sequencing and mass spectrometry),<sup>4</sup> suggesting that the surrounding region might be solvent exposed.

**The Importance of the N-terminal 44 Amino Acids for Activity**—Polymerase activity assays demonstrated that POL2, in contrast to POL1, is unable to incorporate [<sup>3</sup>H]GMP using the homopolymeric RNA template poly(rC) (Fig. 4A). The two structures are, however, very similar with a r.m.s. deviation on equivalent C $\alpha$  atoms as low as 1.16 Å (even if POL1 is in slightly more closed conformation than POL2 (see Supplemental Material)). There are no differences in the orientation of the residues within the active site between POL2 and POL1. The major difference between the two structures is the presence of 44 additional amino acids at the N terminus of POL1. Interestingly, these 44 residues adopt a fold not found in other RdRp structures, except that of the foot-and-mouth disease virus RdRp (53). The additional residues induce the formation of a three-stranded  $\beta$ -sheet (made of strands  $\beta 1$ ,  $\beta 2$ , and  $\beta 3$ , see Figs. 2 and 4B). The presence of this ordered three-stranded  $\beta$ -sheet might stabilize the whole “back” subdomain of the fingers domain. Indeed, 70 amino acids of this subdomain, disordered in POL2, are clearly visible in the POL1 electron density map (aa 317–321, aa 339–361, aa 416–419, aa 453–458, aa 471–472, aa 577–602, and aa 748–751). This back subdomain closes the groove accommodating the incoming template RNA and so contributes to the formation of the RNA template tunnel, whereas in POL2, this groove might remain open due to partial disorder (Fig. 4B).

**The Priming Loop**—Several structural and biochemical studies on HCV, BVDV, and Phi6 RdRps have demonstrated that the priming loop constitutes a platform stabilizing the RNA synthesis initiation complex (22, 48, 53, 56, 57, 62). Conformational changes, not yet characterized by structural studies, are presumably required to shift to an RNA synthesis elongation mode. Comparative enzymatic characterization of WNV and HCV RdRps suggests that a similar conformational change occurs for WNV POL, even though some structural differences could account for the different processivity of the two RdRps (12). The WNV POL structure reveals that the priming loop consists of aa 796–809 (Fig. 1A in purple and Fig. 2 in yellow), which was difficult to infer from sequence alignment. It originates from the same part of the thumb domain as for the BVDV and HCV RdRps (it connects  $\alpha$ -helices  $\alpha 20$  to  $\alpha 21$ ) but is larger.

<sup>4</sup> B. Coutard, unpublished results.

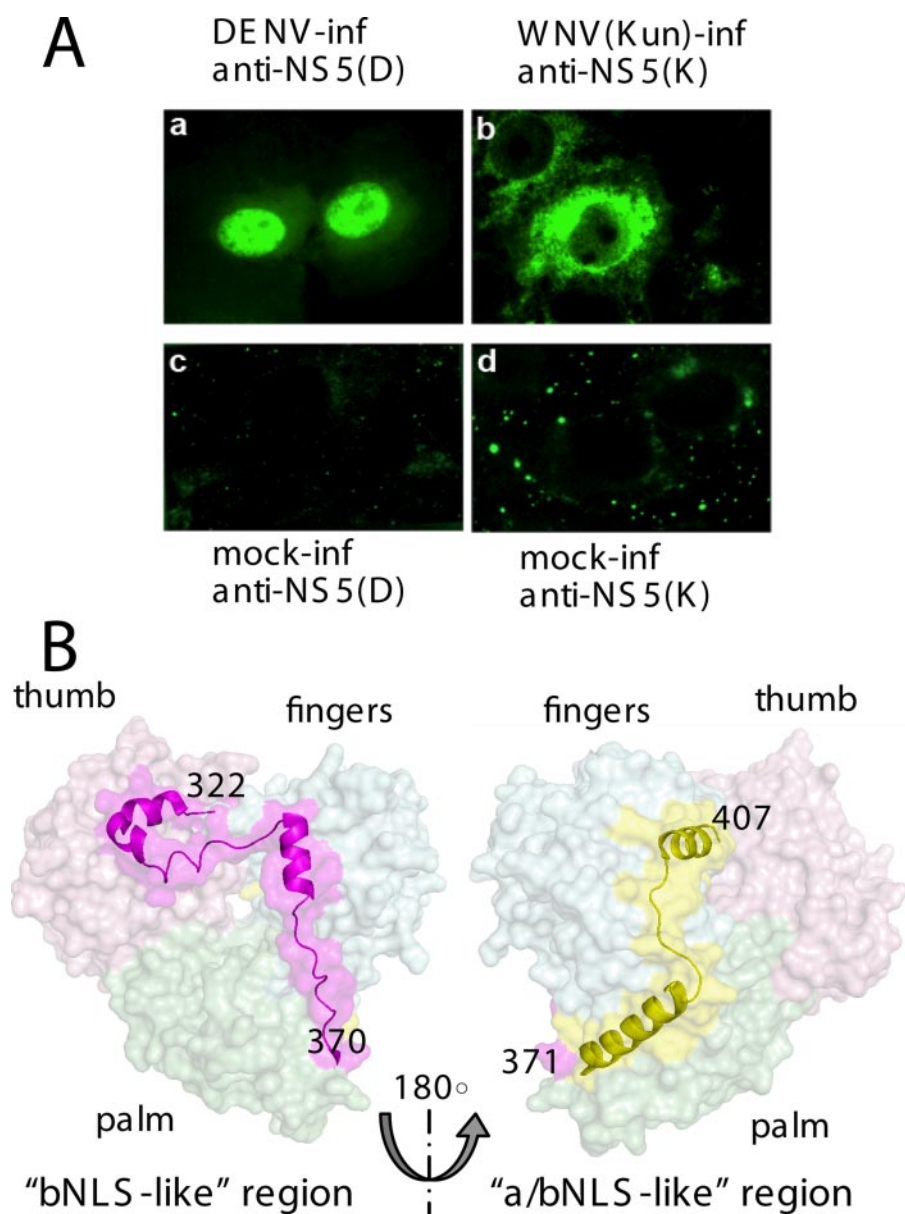
## Crystal Structure of West Nile Virus RNA Polymerase



**FIGURE 5. A model for RNA polymerization initiation by WNV POL1.** A, stereo view of superimposition of the priming loops of WNV (yellow), HCV (red), and BVDV (cyan) RdRps seen from the front view described in the legend to Fig. 1. Tyrosine residues involved in the stabilization of the priming nucleotide in both HCV and BVDV RdRps are indicated, as well as Trp<sup>800</sup> and His<sup>803</sup> of WNV POL. B, the RNA polymerization initiation model, seen from the back, i.e. rotated by 180° compared with A. The RNA template (3' of the negative strand genomic RNA of WNV corresponding to 5'-UACU-3') is represented according to atom types, catalytic ions as purple spheres, and priming and catalytic nucleotides in blue. WNV POL residues that stabilize the nucleotides and that are conserved in sequence and orientation when compared with HCV RdRp are shown in green sticks. The priming loop is colored in yellow and Trp<sup>800</sup> of the priming loop is represented in yellow sticks.

Its overall size and positioning is most similar to that of Phi6 RdRp, whereas its folding is unique, bearing no secondary structural elements (Figs. 1A, 2B, and 5). It is stabilized by hydrogen bonds with side chains belonging to residues located in  $\alpha$ -helices  $\alpha 11$  (Gln<sup>515</sup>),  $\alpha 14$  (Tyr<sup>610</sup>), and  $\alpha 19$  (Lys<sup>761</sup>, Tyr<sup>763</sup>, and Gln<sup>765</sup>), and by van der Waals interactions with side chains belonging to residues in  $\alpha$ -helices  $\alpha 18$  (Arg<sup>742</sup>) and  $\alpha 19$  (Arg<sup>754</sup> and Met<sup>766</sup>). Most of these residues are conserved in the DENV, Japanese encephalitis virus, and yellow fever virus RdRps. In the priming loops of the Phi6, HCV and BVDV RdRps, a tyrosine side chain (Tyr<sup>630</sup> in Phi6, Tyr<sup>448</sup> in HCV, and Tyr<sup>581</sup> in BVDV) is used to orient the base of the priming nucleotide via a stacking interaction. In WNV POL, there is no equivalent tyrosine. However, two conserved residues Trp<sup>800</sup> and His<sup>803</sup>, which may perform similar function, are found in close proximity.

**A Model for Initiation of RNA Synthesis**—To characterize the initiation complex of WNV POL, many attempts of co-crystallization with short sequence-specific RNAs, oligonucleotides, and divalent ions were made. As these attempts were unsuccessful, we used the similarity with the HCV and Phi6 RdRps to



**FIGURE 6. Regions of WNV NS5 corresponding to DENV putative nuclear localization sequences and differences in nuclear import between the DENV and WNV NS5 proteins.** *A*, DENV but not WNV NS5 localizes to the nucleus of virus-infected cells. Vero cells were infected as indicated with dengue type 2 virus (DENV-2) for 40 h (*a*), or Kunjin subtype of WNV (WNV (*Kun*)) for 24 h (*b*), or mock-infected (*c* and *d*). Cells were fixed for nuclear staining by treatment with 4% paraformaldehyde containing 0.1% Triton X-100 and proteins were detected with rabbit polyclonal antibodies specific for either WNV (*Kun*) or DENV NS5 protein (28, 63) and fluorescein isothiocyanate-conjugated anti-rabbit IgG. Rabbit polyclonal antibodies to DENV-2 NS5 (*anti-NS5(D)*) were a generous gift from Dr. R. Padmanabhan. *B*, on the *left side* of the panel, a back view of the WNV POL1 shows in *pink* the region corresponding to the putative DENV bNLS. On the *right side*, a front view of the WNV POL1 showing in *yellow* the position of structural elements corresponding to the putative DENV a/bNLS (rotation of 180° compared with the back view). The palm, thumb, and fingers are represented in *light green, red, and blue*, respectively.

propose a polymerization initiation model (Fig. 5*B*). rNTPs were positioned at both the priming (site P) and catalytic (site C) sites according to the HCV (PDB code 1GX5) and Phi6 (PDB code 1HI0) models. The interrogation site I, missing in WNV POL, is sufficiently far away from the active site, such that a meaningful initiation complex could be built. It became obvious that a simple change in the Trp<sup>800</sup> side chain conformation resulted in the ideal positioning of its aromatic base to stack against the priming nucleotide, thereby playing a role similar to

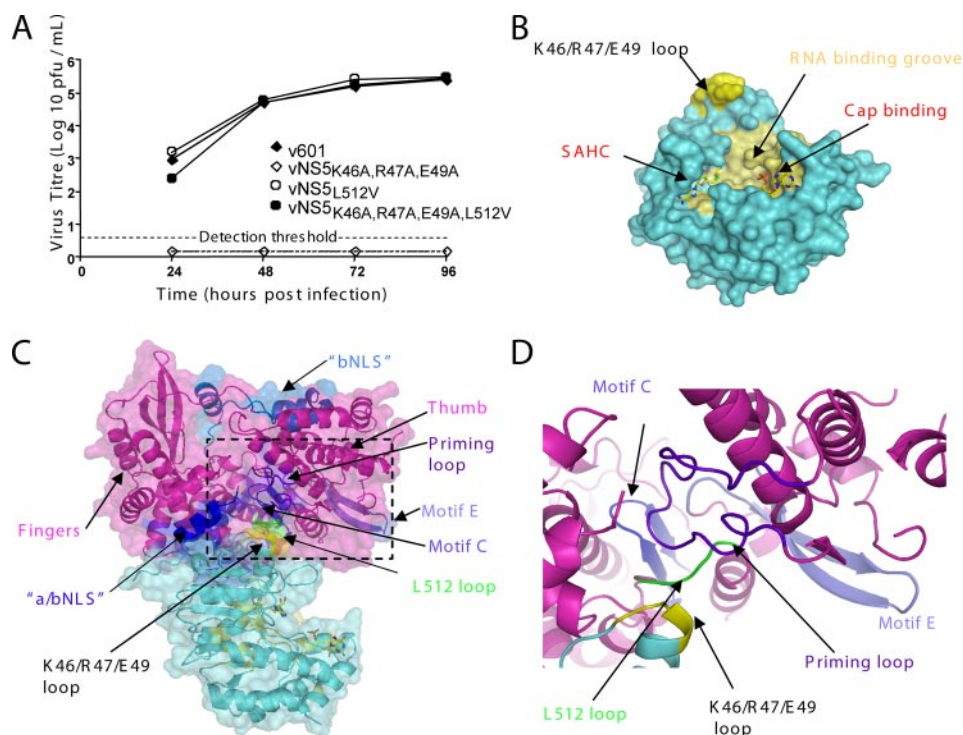
Phi6 Tyr<sup>630</sup> and HCV Tyr<sup>448</sup> (Fig. 5*A*). The side chain conformation of Asp<sup>536</sup>, the position of the two catalytic magnesium ions, and the position of a 4-mer RNA corresponding to the 3' end of the (–)-strand genomic RNA of WNV, was inferred from the initiation complex structure of Phi6 RdRp. No major steric clash between the RNA, the nucleotides modeled, and the solved structure was observed. The resulting model, after minimization, is shown in Fig. 5*B*. In conclusion, residue Trp<sup>800</sup>, which is strictly conserved in *Flavivirus* NS5 proteins, seems to be the key residue in the priming loop providing the initiation platform for *de novo* initiation of RNA synthesis by *Flavivirus* RdRps.

*The NLS Region*—Two adjacent putative NLSs have been identified in DENV NS5, a bNLS (aa 320–368 in DENV NS5, which corresponds to aa 322–370 in WNV) and the a/bNLS (aa 369–405 in DENV NS5, which corresponds to aa 371–407 in WNV). The NLSs present in DENV NS5 were shown to be functional *in vitro* although their role in NS5 nuclear transport in virus-infected cells has not yet been demonstrated. Corresponding studies have not been reported for WNV NS5. Thus, we first examined whether WNV NS5 translocates to the nucleus in WNV-infected cells by immunofluorescence analysis with anti-NS5 antibodies. WNV NS5 was found primarily in the cytoplasm and perinuclear region, but not in the nucleus (Fig. 6*A*). In contrast, DENV NS5 clearly showed a nuclear localization (Fig. 6*A*), as reported previously (28).

Although WNV NS5 was not nuclear localized, the structure of WNV POL reveals the fold and

localization of the regions corresponding to the DENV NLSs on the WNV POL surface (Fig. 6*B*). An equivalent structural scaffold can be found in both BVDV and HCV RdRps (aa 132–217 of BVDV and aa 20–94 of HCV), with a well conserved overall shape and secondary structure, but not charge. The DENV bNLS has been shown to interact with the 300 C-terminal amino acids of NS3 (30). No precise mapping of this interaction is available to date. However, assuming that the corresponding region of WNV NS5 (situated on the top of the back of the





**FIGURE 7. Reverse genetics and structural modeling studies to propose a *Flavivirus* full-length NS5 model.** *A*, growth curve analysis of the parental (v601) and mutant viruses in Vero cells. Vero cells were infected at a multiplicity of infection of 1 with respective viruses. At 24-h intervals after infection the culture supernatants were sampled and the viral titer determined by a plaque assay on C6/36 cells. The horizontal dashed line represents the limits of viral detection. *B*, surface representation of the MTase domain of WNV (cyan) and its putative RNA binding groove (pale yellow). The position of *S*-adenosyl-L-homocysteine (SAHC) within the cofactor-binding site and the cap-binding site with a bound GTP analogue (ligands are taken from the DENV MTase structure, PDB code 1L9K). The position of the Lys<sup>46</sup>/Arg<sup>47</sup>/Glu<sup>49</sup> loop is shown in yellow. *C*, top view of the WNV POL1 (pink) and the MTase (cyan) domains after docking. *a/bNLS* and *bNLS* refer to the *a/bNLS*-like and *bNLS*-like subdomains described in the text, corresponding to DENV NLS regions. *D*, close-up view of *C* around the interacting region.

polymerase as shown in Figs. 6*B* and 7*C*) interacts with NS3, this NS5-NS3 interaction may facilitate localization of NS3 close to the entrance of the RNA template tunnel, consistent with the putative role of its helicase activity in dsRNA unwinding during RNA synthesis.

**Model of Full-length WNV NS5**—The relative spatial organization of the MTase and RdRp domains in *Flavivirus* NS5 is unknown. However, the availability of the crystal structures of the DENV MTase domain (PDB code 1L9K) and of the WNV RdRp domain, as well as our results from DENV reverse genetic experiments, provided us with the opportunity to propose a model of full-length WNV NS5.

Using a DENV-2 infectious cDNA clone, we identified a genetic interaction between the MTase and RdRp domains of DENV-2 NS5. RNA transcribed from the infectious DENV-2 clone yields infectious virus (v601) when introduced into permissive cells. Based on the crystal structure of the DENV-2 MTase domain, a series of mutations designed to disrupt surface interactions involving charged residues was introduced into the *ns5* gene. In particular, amino acids Lys<sup>46</sup>, Arg<sup>47</sup>, and Glu<sup>49</sup> were chosen as they are charged and surface exposed on the MTase. Mutation of these three charged amino acids to alanine could induce a non-complementary charge between the POL and MTase, disrupting domain interactions, and inhibiting virus growth and/or infectivity. RNA transcripts

encoding the K46A/R47A/E49A mutation in the DENV-2 MTase domain failed to routinely produce infectious virus when transfected into permissive cells. However, after repeated attempts, virus was twice recovered from cells transfected with the mutant transcripts. As RT-PCR and sequencing confirmed the presence of the K46A/R47A/E49A mutation in each of the recovered viruses, the entire viral genomes were sequenced to determine whether a second site mutation had arisen that compensated the effects of the K46A/R47A/E49A mutation. Both viruses contained a Leu to Val substitution at surface residue 512 (L512V). The L512V mutation was then tested for its ability to compensate for the effects of the initially introduced K46A/R47A/E49A mutation. By contrast, recombinant viruses were recovered from

both the parental RNA transcripts and transcripts containing the mutations L512V (vNS5<sub>L512V</sub>) and K46A/R47A/E49A/L512V (vNS5<sub>K46A,R47A,E49A,L512V</sub>). The entire genome of all viruses was sequenced, confirming that only the expected mutations were present. Growth curve analysis of the viruses in Vero cells (Fig. 7*A*) demonstrated that vNS5<sub>K46A,R47A,E49A,L512V</sub> and vNS5<sub>L512V</sub> had similar replication kinetics and peak viral titers to v601. Overall, the presence of mutation L512V either alone, or in combination with mutation K46A/R47A/E49A appeared to maintain or even slightly increase virus replicative ability compared with the parental v601. These results suggest that L512V, in the RdRp domain, acts to compensate for the K46A/R47A/E49A mutation, either directly (*i.e.* V512 contacts the K46A/R47A/E49A cluster) or indirectly (V512 restores functionality by a long-range conformational change) in the MTase domain. Our model of full-length NS5 is based on the first hypothesis.

A model of the WNV MTase domain was generated by homology modeling based on the structure of the DENV MTase (PDB code 1L9K, 59% identity). Fig. 7*B* shows the WNV MTase domain with bound co-product *S*-adenosylhomocysteine, the RNA substrate binding groove, and the site where the RNA cap is bound (with a bound GTP analogue) during methyltransferase at the 2'-*O*-position of the first nucleotide. Subsequently, the entire binding space for both the MTase and POL1



domains was screened. A 6-Å filter between the amino acids corresponding to DENV NS5 512 in the POL1 domain and DENV NS5 46, 47, and 49 in the MTase domain was applied. A second filter based on restraints on the distance between the C-terminal amino acid of the MTase domain and the N-terminal amino acid of the POL1 domain reduced the number of solutions from 6 to 1. In the corresponding model (Fig. 7C), the C terminus of the MTase domain is 36 Å away from the N terminus of the POL1 domain, implying that the 13-residue linker between both domains (from aa 264 of the MTase domain to aa 278 of the POL domain) is expected to adopt an extended, stretched conformation. The MTase “Lys<sup>46</sup>/Arg<sup>47</sup>/Glu<sup>49</sup> loop” fits well into a large groove located between the thumb and the fingers domains (Fig. 7, C and D). The bottom of this groove is made of the short “Leu<sup>512</sup> loop” of the palm domain, which connects  $\alpha$ -helices  $\alpha$ 10 to  $\alpha$ 11 and which neighbors the C and E motifs (Figs. 2B and 7D). This loop also delineates “the bottom” of the tunnel through which the newly synthesized dsRNA should come out from the active site (Fig. 2B).

From a structural perspective it is not possible to provide a precise explanation about the nature of the interaction between the K46A/R47A/E49A cluster and L512V in DENV NS5. It might be possible that Val<sup>512</sup> protrudes less from the surface than Leu<sup>512</sup>, and Val<sup>512</sup> allows the formation of a shallow hydrophobic surface region with the neighboring glycines. This region might then interact better with the three alanines of the MTase domain. Both the direct contact hypothesis and our model remain speculative. The possibility exists that L512V promotes an allosteric change compensating the alanine mutations. Thus, the model should be taken with caution as a preliminary guide in the current efforts to understand how RNA synthesis and RNA capping are coordinated. Nevertheless, it is interesting to note that the RNA-substrate binding groove of the MTase domain is positioned near the RdRp dsRNA exit tunnel defined above. The model seems to be structurally and mechanistically consistent with RNA cap methylation of the nascent genome occurring after the (+) RNA product leaves the RdRp domain.

**Concluding Remarks**—The *Flaviviridae* virus family comprises important and emerging human pathogens. The crystal structure of the WNV RdRp domain presented here reveals a number of novel *Flavivirus*-specific features in comparison to known structures of *Hepacivirus* and *Pestivirus* RdRps. It will constitute a basis for structure-based drug design studies against flaviviruses. Our model of full-length WNV NS5 represents a structural hypothesis for further investigation of the spatial organization of the *Flavivirus* replication complex.

**Acknowledgments**—We thank the Joint Structural Biology Group staff at the European Synchrotron Radiation Facility for technical assistance, and in particular D. Nurrizzo. Special thanks to H. Dutartre and C. Bussetta for helpful discussions and C. Cambillau, V. Campanacci, and B. Coutard.

## REFERENCES

- Mackenzie, J. S., Gubler, D. J., and Petersen, L. R. (2004) *Nat. Med.* **10**, S98–S109
- Guzman, M. G., and Kouri, G. (2002) *Lancet Infect. Dis.* **2**, 33–42
- Scherret, J. H., Poidinger, M., Mackenzie, J. S., Broom, A. K., Deubel, V., Lipkin, W. I., Briese, T., Gould, E. A., and Hall, R. A. (2001) *Emerg. Infect. Dis.* **7**, 697–705
- Lanciotti, R. S., Ebel, G. D., Deubel, V., Kerst, A. J., Murri, S., Meyer, R., Bowen, M., McKinney, N., Morrill, W. E., Crabtree, M. B., Kramer, L. D., and Roehrig, J. T. (2002) *Virology* **298**, 96–105
- Hall, R. A., Broom, A. K., Smith, D. W., and Mackenzie, J. S. (2002) *Curr. Top. Microbiol. Immunol.* **267**, 253–269
- Egloff, M. P., Benarroch, D., Selisko, B., Romette, J. L., and Canard, B. (2002) *EMBO J.* **21**, 2757–2768
- Ray, D., Shah, A., Tilgner, M., Guo, Y., Zhao, Y., Dong, H., Deas, T. S., Zhou, Y., Li, H., and Shi, P. Y. (2006) *J. Virol.* **80**, 8362–8370
- O'Reilly, E. K., and Kao, C. C. (1998) *Virology* **252**, 287–303
- Koonin, E. V. (1991) *J. Gen. Virol.* **72**, 2197–2206
- Steffens, S., Thiel, H. J., and Behrens, S. E. (1999) *J. Gen. Virol.* **80**, 2583–2590
- Guyatt, K. J., Westaway, E. G., and Khromykh, A. A. (2001) *J. Virol. Methods* **92**, 37–44
- Selisko, B., Dutartre, H., Guillemot, J. C., Debarnot, C., Benarroch, D., Khromykh, A., Despres, P., Egloff, M. P., and Canard, B. (2006) *Virology* **351**, 145–158
- Ackermann, M., and Padmanabhan, R. (2001) *J. Biol. Chem.* **276**, 39926–39937
- Nomaguchi, M., Ackermann, M., Yon, C., You, S., and Padmanabhan, R. (2003) *J. Virol.* **77**, 8831–8842
- Ranjith-Kumar, C. T., Gutshall, L., Kim, M. J., Sarisky, R. T., and Kao, C. C. (2002) *J. Virol.* **76**, 12526–12536
- Ranjith-Kumar, C. T., Kim, Y. C., Gutshall, L., Silverman, C., Khandekar, S., Sarisky, R. T., and Kao, C. C. (2002) *J. Virol.* **76**, 12513–12525
- Kao, C. C., Singh, P., and Ecker, D. J. (2001) *Virology* **287**, 251–260
- Ago, H., Adachi, T., Yoshida, A., Yamamoto, M., Habuka, N., Yatsunami, K., and Miyano, M. (1999) *Structure* **7**, 1417–1426
- Bressanelli, S., Tomei, L., Roussel, A., Incitti, I., Vitale, R. L., Mathieu, M., De Francesco, R., and Rey, F. A. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96**, 13034–13039
- Choi, K. H., Groarke, J. M., Young, D. C., Kuhn, R. J., Smith, J. L., Pevear, D. C., and Rossmann, M. G. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **101**, 4425–4430
- Lesburg, C. A., Cable, M. B., Ferrari, E., Hong, Z., Mannarino, A. F., and Weber, P. C. (1999) *Nat. Struct. Biol.* **6**, 937–943
- Bressanelli, S., Tomei, L., Rey, F. A., and De Francesco, R. (2002) *J. Virol.* **76**, 3482–3492
- Benarroch, D., Selisko, B., Locatelli, G. A., Maga, G., Romette, J. L., and Canard, B. (2004) *Virology* **328**, 208–218
- Wengler, G. (1993) *Virology* **197**, 265–273
- Bartelma, G., and Padmanabhan, R. (2002) *Virology* **299**, 122–132
- Yon, C., Teramoto, T., Mueller, N., Phelan, J., Ganesh, V. K., Murthy, K. H., and Padmanabhan, R. (2005) *J. Biol. Chem.* **280**, 27412–27419
- Buckley, A., Gaidamovich, S., Turchinskaya, A., and Gould, E. A. (1992) *J. Gen. Virol.* **73**, 1125–1130
- Kapoor, M., Zhang, L., Ramachandra, M., Kusukawa, J., Ebner, K. E., and Padmanabhan, R. (1995) *J. Biol. Chem.* **270**, 19100–19106
- Uchil, P. D., Kumar, A. V., and Satchidanandam, V. (2006) *J. Virol.* **80**, 5451–5464
- Johansson, M., Brooks, A. J., Jans, D. A., and Vasudevan, S. G. (2001) *J. Gen. Virol.* **82**, 735–745
- Brooks, A. J., Johansson, M., John, A. V., Xu, Y., Jans, D. A., and Vasudevan, S. G. (2002) *J. Biol. Chem.* **277**, 36399–36407
- Doublet, S. (1997) *Methods Enzymol.* **276**, 523–530
- Collaborative Computational Project, N. (1994) *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763
- Schneider, T. R., and Sheldrick, G. M. (2002) *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1772–1779
- Sheldrick, G. M. (2002) *Z. Kristallogr.* **217**, 644–650
- Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M., and Paciorek, W. (2003) *Acta Crystallogr. D Biol. Crystallogr.* **59**, 2023–2030
- Terwilliger, T. C. (2002) *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1937–1940

38. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132
39. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255
40. Navaza, J. (2001) *Acta Crystallogr. D Biol. Crystallogr.* **57**, 1367–1372
41. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905–921
42. Blanc, E., Roversi, P., Vornrhein, C., Flensburg, C., Lea, S. M., and Bricogne, G. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2210–2221
43. Gualano, R. C., Pryor, M. J., Cauchi, M. R., Wright, P. J., and Davidson, A. D. (1998) *J. Gen. Virol.* **79**, 437–446
44. Pryor, M. J., Gualano, R. C., Lin, B., Davidson, A. D., and Wright, P. J. (1998) *J. Gen. Virol.* **79**, 2631–2639
45. Guex, N., and Peitsch, M. C. (1997) *Electrophoresis* **18**, 2714–2723
46. Hooft, R. W., Vriend, G., Sander, C., and Abola, E. E. (1996) *Nature* **381**, 272
47. Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A. A., Aflalo, C., and Vakser, I. A. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 2195–2199
48. Butcher, S. J., Grimes, J. M., Makeyev, E. V., Bamford, D. H., and Stuart, D. I. (2001) *Nature* **410**, 235–240
49. Huang, H., Chopra, R., Verdine, G. L., and Harrison, S. C. (1998) *Science* **282**, 1669–1675
50. O'Farrell, D., Trowbridge, R., Rowlands, D., and Jager, J. (2003) *J. Mol. Biol.* **326**, 1025–1035
51. Tao, Y., Farsetta, D. L., Nibert, M. L., and Harrison, S. C. (2002) *Cell* **111**, 733–745
52. Ng, K. K., Cherney, M. M., Vazquez, A. L., Machin, A., Alonso, J. M., Parra, F., and James, M. N. (2002) *J. Biol. Chem.* **277**, 1381–1387
53. Ferrer-Orta, C., Arias, A., Perez-Luque, R., Escarmis, C., Domingo, E., and Verdaguer, N. (2004) *J. Biol. Chem.* **279**, 47212–47221
54. Hansen, J. L., Long, A. M., and Schultz, S. C. (1997) *Structure* **5**, 1109–1122
55. Gorbalenya, A. E., Pringle, F. M., Zeddiam, J. L., Luke, B. T., Cameron, C. E., Kalmakoff, J., Hanzlik, T. N., Gordon, K. H., and Ward, V. K. (2002) *J. Mol. Biol.* **324**, 47–62
56. Adachi, T., Ago, H., Habuka, N., Okuda, K., Komatsu, M., Ikeda, S., and Yatsunami, K. (2002) *Biochim. Biophys. Acta* **1601**, 38–48
57. Leveque, V. J., Johnson, R. B., Parsons, S., Ren, J., Xie, C., Zhang, F., and Wang, Q. M. (2003) *J. Virol.* **77**, 9020–9028
58. Bruenn, J. A. (2003) *Nucleic Acids Res.* **31**, 1821–1829
59. Thompson, A. A., and Peersen, O. B. (2004) *EMBO J.* **23**, 3462–3471
60. Ferrer-Orta, C., Arias, A., Escarmis, C., and Verdaguer, N. (2006) *Curr. Opin. Struct. Biol.* **16**, 27–34
61. Hong, Z., Cameron, C. E., Walker, M. P., Castro, C., Yao, N., Lau, J. Y., and Zhong, W. (2001) *Virology* **285**, 6–11
62. van Dijk, A. A., Makeyev, E. V., and Bamford, D. H. (2004) *J. Gen. Virol.* **85**, 1077–1093
63. Khromykh, A. A., Sedlak, P. L., Guyatt, K. J., Hall, R. A., and Westaway, E. G. (1999) *J. Virol.* **73**, 10272–10280