Structure Function Analysis of West Nile Virus RNA Dependent RNA Polymerase: Molecular Model and Implications for Drug Design

Arezki Azzi and Sheng-Xiang Lin*

Oncology and Molecular Endocrinology Research Center, Laval University Medical Center (CHUQ), 2705 Boul. Laurier, Quebec, Canada, G1V 4G2

Abstract: West Nile virus (WNV) is a mosquito-borne disease that emerged in North America. In 2002 it caused the largest arboviral meningoencephalitis outbreak ever recorded in the US and Canada. The key enzyme responsible for the replication of the virus is the RNA-dependent RNA polymerase (RdRp) enzyme represented by nonstructural protein NS5 in WNV.

To understand the structural basis and enzymatic activity of WNV RdRp as well as potential drug susceptibility, we have built a homology model of WNV NS5 RdRp based on a combination of motif search, fold recognition and sequence alignments orchestrated by 3D-jury system. We have located conserved sequence motifs shared by all RdRps and described the potential functional role of these motifs and specific residues in the polymerization and in the recognition of potential inhibitors.

Virtual docking of several substrates and inhibitors in our WNV RdRp model shows that a non-nucleoside inhibitor such as 2'-methyl-ribofuranosyl-guanosine triphosphate has a higher binding energy indicated by a low free energy obtained upon binding. These results provide a preliminary basis for the development of anti-WNV agents based on RNA dependant RNA polymerase inhibition.

Key Words: West nile virus, RNA dependent RNA polymerase, molecular modeling, inhibitors.

INTRODUCTION

West Nile virus (WNV) is a mosquito-borne disease that emerged in North America. In 2002 it was responsible for the largest arboviral meningoencephalitis outbreak ever recorded in the US and Canada. Since the emergence of the virus in the Western Hemisphere, specifically in New York City in 1999 [1], it has spread rapidly throughout the United States. The West Nile virus has established itself firmly in North America [2] and has spread throughout the 48 contiguous states and 7 Canadian provinces, as well as Mexico, the Caribbean islands and Colombia suggesting an establishment in the neotropics [3]. A significant public health risk in this region is possible because the mosquito population is active year-round with transmission of the virus to humans, to the established avian reservoir and to migratory birds. West Nile Virus, like other flaviviruses such as Yellow Fever, Murray Valley encephalitis, Japanese encephalitis, and Dengue viruses, is becoming an important human pathogen especially for the elderly population where the case-fatality rate ranges from 15 to 29 % [4]. There are currently no specific treatments or vaccines approved for clinical usage against the West Nile virus.

The West Nile viral genome consists of a 42S viral RNA that encodes both structural and non structural proteins. Viral replication proceeds *via* synthesis of a complementary minus strand RNA using the genome as a template and the subsequent synthesis of genomic plus strand RNA from the minus strand RNA template. The key enzyme responsible for both steps is the RNA-dependent RNA polymerase (RdRp) represented by the 600 C-terminal residues of nonstructural protein NS5 in WNV. Since RdRp activity is not essential for the host, its inhibition is not anticipated to result in undesirable side effects during therapy.

To understand the structural basis of WNV RdRp enzymatic activity and potential drug susceptibility, we have built a homology model. In our model we have located the conserved sequence and structural motif shared by all RdRps. We describe the potential functional role of these conserved motifs and specific residues in the polymerization mechanism and in the recognition of potential inhibitors. Structural analysis of WNV RdRp is likely to aid the development of anti-WNV agents and facilitate the design of future molecular biology and biochemical experiments [5].

MATERIAL AND METHODS

Sequence Aligment

Sequences of WNV RdRp and representative RdRp sequences from the flavivirus tick-borne encephalitis virus (TBEV; NCBI, # NP_775511), hepatitis C virus (HCV; PDB Id 1QUV), bromovirus Tomato aspermy virus (TAV; Swiss Prot accession # P29035) and coronavirus rabbit hemorrhagic disease virus (RHDV: PDB Id 1KHV_A) were aligned using the program ClustalW [6].

Method for Model Building

A 3D template for building a WNV RdRp model was selected by a combination of motif search, fold recognition

*Address correspondence to this author at the Oncology and Molecular Endocrinology Research Center, Laval University Medical Center (CHUQ), 2705 Boul. Laurier, Quebec, Canada, G1V 4G2;
E-mail: sheng-xiang.lin@crchul.ulaval.ca

1573-4064/07 $50.00+.00

and structure based sequence alignment orchestrated by the protein structure prediction Meta Server and the 3D Jury method [7]. The server used Sequence Only Methods utilize exclusively sequence information available about the target and the templates to find similarities. They can be used for fold recognition and detection of homologs. The input for 3D-Jury is groups of models generated by a set of prediction servers. All models are compared with each other and a similarity score, equal to the number of C-alpha atom pairs that are within 3.5 Å, is assigned to each pair. The 3D-Jury score of a model is the sum of all similarity scores of considered model pairs divided by the number of considered pairs plus one [7]. The sequence from 300 to 905 of West Nile virus NS5 protein (NP_776022) which contains all of the RdRp functional domains was modeled.

Model Refinement

The model generated from the 3D-jury search contains a few missing residues corresponding to sequence that did not have a motif, a fold or sequence similarities with other proteins. Residue side-chains and loops were built with MODELLER [8]. The final model was obtained after building the most probable rotamer conformation for each side chain where required. This model was refined and energy minimized using GROMOS 43B1 force field [9] from the SWISS-PDB DeepView [10]. All computations were done *in vacuo* without reaction field. Validation of the stereochemistry of the final model structure was done with Procheck [11].

Ligand Docking

Uridine triphosphate as well as known RdRp nucleosidic inhibitors 2'-C-methyladenosine triphosphate, 2'-O-methyl-cytidine triphosphate, 2'-methyl-ribofuranosyl-guanosine triphosphate were selected for docking. Ligand docking in the active site was carried out with the Genetic Optimization Ligand Docking program (GOLD) version 3.0 [12]. Ligand and side-chain flexibility was allowed during docking. The ChemScore function [13] estimates the free energy of ligand binding to a protein as follows:

$$\Delta G_{binding} = \Delta Go + \Delta G_{hbond} S_{hbond} + \Delta G_{metal} S_{metal} + \Delta G_{lipo} S_{lipo} + \Delta G_{rot} H_{rot}$$ [13],

where S_{hbond}, S_{metal}, and S_{lipo} are scores for hydrogen bonding, acceptor-metal, and lipophilic interactions, respectively. H_{rot} represents the loss of conformational entropy of the ligand upon binding to the protein.

RESULTS AND DISCUSSION

Building a Homology Model for RdRp

PDB-BLAST analysis of the NS5 protein sequence revealed a 58 % identity for the first N-terminal 300 residues with Dengue virus L-methionine *S*-methyltransferase (MMT). The remaining C-terminal 605 residues have no homologous structure but sequence similarity search with BLAST reveals a very close protein homolog with 65 % sequence identity to tick-borne encephalitis virus (TBEV) RNA dependent RNA polymerase which belongs to the mammalian tick-born virus group [14]. We decided to focus on modeling the structure of the RdRp enzyme because of its crucial role in the virus replication. We built a homology model for the WNV RdRp using the protein structure prediction Meta Server and the 3D Jury method for model selection. This resulted in the identification of Hepatitis C virus NS5B RdRp X-ray template structure (1QUV) [15] resolved at 2.5 Å and a 3D-jury score of 267 as a template for creating a full atom 3D model of WNV RdRp. This model, obtained from the fold and function assignment system server (FFAS03), also generated the longest modeled sequence among the template selection provided by 3D-jury.

After subsequent examination and optimization of side-chain rotamers, the RdRp model was further analyzed. The final RdRp model has a correct backbone conformation as indicated by the fact that 99.8 % of residues were located in the Ramachandran plot allowed region (plot not shown). Only residue Ala862 located in a loop near the C-terminal end is outside the allowed region. The WNV RdRp sequence contains the strictly conserved motifs A to F, which are well aligned with the motifs present in others RdR polymerases Fig. (**1**). Alignments of the model with other known RdRps reveal the presence of structurally conserved motifs Fig. (**2A**). WNV RdRp folds in a classic "right hand" shape that contains motifs shared by all RdRps. The characteristic fingers, palm, and thumb domains with the structural motifs are indicated in Fig. (**2A**).

Structural Features of WNV RdRp

Analysis of the domain structural motifs and their potential functions are presented in Table **1**. Among viral RdRps, the sequence and structure of the fingers domain are less conserved than those of the palm domain. The WNV RdRp fingers domain is composed of four α-helices (residues 414-446, 479-502, 509-531 and 544-556), the RdRp characteristic α-helix-loop-α-helix (residues 544-578), two stranded β-sheets (residues 453-459 and 473-477) and one conserved sequence motif F shared by all RdRps that plays an important functional role in the mechanism of polymerization. Interrogation, priming and catalytic sites for nucleotide binding share overlapping basic residues in this region. Based on homology with HCV RdRp, we propose that the WNV RdRp interrogation site includes basic conserved residues K459, K471 and R474 found in motif F.

Similar to HCV and RDHV RdRps, the finger domain contains an N-terminal portion from residue 300 to 346 that forms a long loop emanating from the fingertip that bridges the finger and thumb domains and thus forms an encircled nucleic acid binding site.

The palm domain of WNV RdRp forms the catalytic core of RdRps and contains four highly conserved sequence motifs (A-E). Residues forming the catalytic active site are found within motifs A and C. The palm domain is conserved among all classes of polymerase and is composed of a three stranded β-sheet and a unique α-helix flanked by two α-helices on one side and a β-sheet and one α-helix on the other. Interestingly, there is an additional insertion in WNV RdRp, forming an α-helix (645-660) between motif B and C indicated as α-helix BC in Fig. (**2A**). The role of this insertion, which is also present in the tick-borne encephalitis virus RdRp, remains unclear, but its close proximity to the active site may indicate a possible participation in the reaction

RdRp		Motif F
WNV	448	ECNTCIYNMMGKREKKP--GEFGKAKGSRAIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLG
TBEV	447	RCAHCVYNMMGKREKKL--GEFGVAKGSRAIWYMWLGSRFLEFEALGFLNEDHWASRESSGAGVEGISLNYLG
HCV	130	TVTPIDTTIMAKNEVFCPEKGGR--KPARLIVFPDLGVRVCEKMALYDVVSTLPQVVMGSSYGFQYSPGQRVE
TAV	433	AENLQFYEHMVKSDVKP---SVTDLNIDRPATITFHRKQLTSQFSPLFTALFQRFQRCLTKRVILPVGKISSL
RHDV	162	KALHHIYACGLKDELRPLDKVKE-GKK-RLLWGCDVGVAVCAAAVFHNICYKLKMVARFGPIAVGVDMTSRDV

		Motif A
WNV	519	YILKEVGTKPGGKVYADDTAGWDTRITKADLENEAKVLELLDGEHRRLARSIIELTYRHKVVKVMRPAADGKTV
TBEV	518	WHLKKLSTLNGGLFYADDTAGWDTKVTNADLEDEEQILRYMEGEHKQLATTIMQKAYHAKVVKVARPSRDGGTV
HCV	203	FLVNTWKSKKNPMGFSYDTRCFDSTVTENDIRVEESIYQCCDLAPE–ARQAIKSLTER-LYIGG--PLTNSKG-
TAV	506	EIKDF–-SVLNKFCLEIDLSKFDKSQGELHLMIQEGILNRLGCPVHISKWWCDFHRMSYIKDKRAGVSMPISF-
RHDV	233	DVIINNLTSKASDFLCLDYSKWDSTMSPCVVRLAIDILADCCEQTELTKSVVLTLKSHPMTILDAMIVQTK---

		Motif B	Motif C
WNV	593	MDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLGKGKGPKVRTWLFENGEERLSRMAVS	
TBEV	592	MDVITRRDQRGSGQVVTYALNTFTNIKVQLIRMMEGEGVIEAADAHNPRLLRVERWLK---EHGEERLGRMLVS	
HCV	272	-QNCGYRRCRASGVLTTSCGNTLTCY-LKASAACRAAKL----------QDCTMLVN	
TAV	578	--------QRRTGDAFTYFGNTLVTMSMFAWCYDTSQF----------DKLIFS	
RHDV	304	-------RGLPSG PFTSVINSICHWLLWSAAVYKSCAEIGLHCSNLYE----------DAPFYTY	

		Motif D	Motif E
WNV	667	GDDCVVKPLDDRFATSLHFLNAMSK–-VRKDIQEWKPSTGWYDWQQVPFCSNHFTELIMKDGRTLVVPCRGQD	
TBEV	663	GDDCVVRPLDDRFGKALYFLNAMSK--TRKDIGRWEHSAGLSSWEEVPFCSHHFHELVMKDGRTLVVPCRDQD	
HCV	317	GDDLVVICESAGTQEDAASLRVFTE--TRYSAPPGDPPQPEYDLELITSCSSNVSVAHDASGKRVYYLTRDPT	
CSFV	613	GDDSLGFSIKAPVGDPSLFTSLFNM–EAKVM-E---PS-VPY------ICSKFLLTDDFGNTFSVPDPLREIQ	
RHDV	353	GDDGVYAMTPMMVSLLPAIIENLRDYGLSPTAADKTEFIDVCPLNKISFLKRTFELTDIGWVSKLDKSSILRQ	

Fig. (1). Partial sequence alignment of WNV RdRp with those of representative RdRps from viruses. The representative RdRps are from the flavivirus tick-borne encephalitis virus (TBEV; NCBI accession # NP_775511), hepatitis C virus (HCV; PDB Id 1QUV), bromovirus Tomato aspermy virus (TAV; Swiss prot accession # P29035) and coronavirus rabbit hemorrhagic disease virus (RHDV: PDB Id 1KHV_A). Strictly conserved residues are represented as white on black background; those partially conserved or characteristic of a motif are black on a grey background. WNV RdRp sequence contains motifs A to F which are confidently aligned with other RdRps.

mechanism. Further analyses by molecular biology are necessary to clarify the role of this α-helix.

Motif A of WNV RdRp contains two highly conserved aspartic residues denoted Asp 536 and Asp 541. Asp536, together with the two strictly conserved aspartates Asp 667 and 668 in motif C, form the catalytic center of WNV RdRp.

The thumb domain consists essentially of 5 α-helices (residues 762-785, 811-821, 832-841, 848-856 and 864-882)

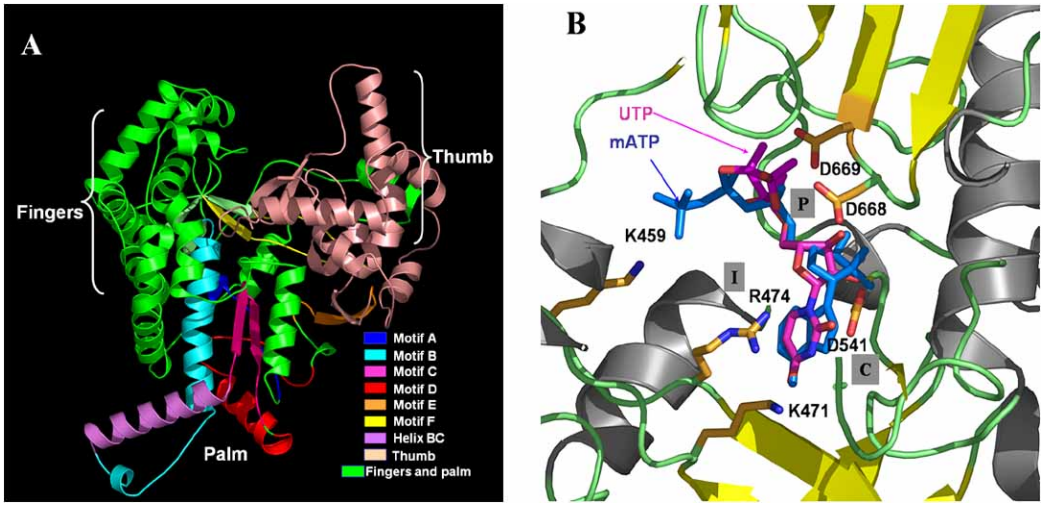


Fig. (2). A. Full view of the molecular model of WNV RdR polymerase. Structural motifs are indicated by color coding. **B.** Docking of uridine triphosphate (UTP) and 2'-C-methyladenosine triphosphate (mATP). The two ligands occupy similar positions in the active site. Nucleotide binding I, C, P sites and conserved residues shown here are all within the enzyme active site region.

Table 1. Potential Function and Interactions of the Conserved Sequence Motifs in the West Nile Virus RNA-dependent RNA Polymerase. Sequence Motifs are Highlighted with Grey Shading Boxes

Motif	Amino acid sequence of the motif	Possible functions and nature of interactions	References
A	530 GKVYADDTAGWDTRI	Asp536: coordination of metals. Asp541: Hydrogen bond to ribose	[21-23]
B	601 QRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKL	Ser604, Thr609 and Asn613: Recognition of the template/primer and sugar selection of rNTP	[15, 22-23]
C	659 RLSRMAVSGDDCVVKPLD	Asp668 and Asp669: Coordination of metals	[21-23]
D	678 RFATSLHFLNAMSKVRKDIQEWKPSTGWY	Stabilization of the core structure.	[22]
E	712 PFCSNHFTELIM	Control the flexibility of the thumb. Ser715 Hydrogen bond to the UTP -β phosphate	[24]
F	458 GKREKKPGEFGKAKGSRAIWF	K459, K471 and R474: rNTP binding and template/primer	[21-24]

and two β-sheets (residues 716-723 and 729-733). The sequence of the thumb domain is less conserved among RdRps. In WNV RdRp, the thumb domain is predicted to contain part of the nucleic acid binding cleft obstructed by the β-hairpin similar to that found in HCV RdRp. This obstruction has been proposed to ensure replication of the 3' portion of the genome during initiation [16]. In addition, the strong interaction between the fingers and thumb domains limits their ability to change conformation independently of each other.

Table 2. Docking of Selected Nucleosides to the WNV RdR Polymerase Active Site. Ligand Docking was Carried Out Using GOLD 3.0 Allowing Flexibility to Ligand and Side-Chains During Docking. The Binding Energy was Estimated from ChemScore Function

Ligand	$\Delta G_{binding}$ kJ/mol
Uridine triphosphate	-26.42
2'-C-Methyladenosine triphosphate	-26.27
2'-O-Methylcytidine triphosphate	-21.62
2'-methyl-ribofuranosyl-guanosine triphosphate	-19.67

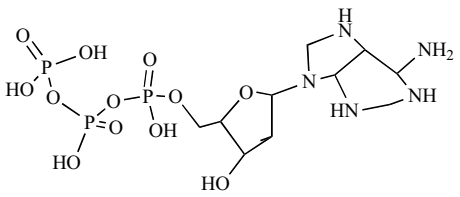
Potential for Inhibiting the RdRp Activity

The RdRp model was further utilized to evaluate small molecules that are expected to block an RdRp catalytic pocket thus inhibiting the enzymatic activity.

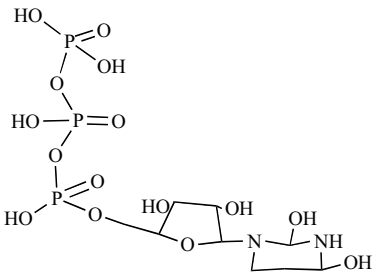
There are two major classes of antiviral agents that target polymerases, nucleoside analogs and non-nucleoside analog inhibitors. Nucleoside analogs such as 2'-C-Methyladenosine and 2'-O-Methylcytidine are substrate analogs that inhibit the enzyme by competing with the substrate. They undergo a phosphorylation of their corresponding nucleotide in the cytoplasm of infected cells in order to become active against the viral polymerases. The nucleotide may be incorporated by the polymerase during processive nucleic acid synthesis leading to early termination of the elongation reaction thus stopping the virus life cycle. Di-nucleotide analogs are also effective in the inhibition of the viral replication initiation, but their chemical instability in the cell remains a challenge that limits their further use [17]. Non-nucleoside inhibitors bind in a hydrophobic pocket on the protein surface distant from the active site and inhibit RNA synthesis noncompetitively with respect to GTP [18]. Examination of the surface of WNV RdRp found no similar hydrophobic pocket. We can therefore predict that non-nucleoside inhibitors would be ineffective for WNV RdRp inhibition.

Starting from our WNV RdRp model, we examined the docking of the substrate UTP and nucleoside analogs Fig. (**3**). The binding energies of nucleoside analog were computed and compared to UTP for identifying the most likely candidates as potential inhibitor molecules. Substrates CTP, UTP and GTP have a similar Km as determined for the hepatitis C virus NS5B RNA-dependent RNA polymerase [19]. A substrate analog should have at least a similar or more favorable binding energy than the substrate to efficiently block the substrate binding site on the enzyme. An oral nucleoside that is potentially useful for the treatment of all HCV genotypes [20], 2'-methyl-ribofuranosyl-guanosine triphosphate, was also docked in WNV RdRp. The $\Delta G_{binding}$ value shown in Table **1** indicates that 2'-C-methyladenosine triphosphate would be a potentially good competitive inhibitor of the WNV RdRp activity, as demonstrated by a similar binding energy to the substrate (Table **2**). Our docking shows that it binds to the same site as UTP Fig. (**2B**). The docking result shows 20 % less binding potential for 2'-O-methylcytidine triphosphate than for 2'-C-methyladenosine triphosphate, implying that it may not be as efficient at inhibiting WNV RdRp as the possibility 2'-C-methyladenosine triphosphate.

The WNV RdRp structural model presented here together with docking data provides important guidelines for the design of novel nucleotide analogs that would specifically inhibit WNV-catalyzed RNA synthesis.



mATP



UTP

Fig. (3). Chemical structure of UTP and 2'-C-methyladenosine triphosphate.

REFERENCES

[1] Briese, T.; Jia, X.-J.; Huang, C.; Grady, L. J.; Lipkin W. I. *Lancet*, **1999***, 354*, 1261.

[2] Drebot, M.A.; Artsob, H. *Can Fam Physician*., **2005** *, 51*, 1094.

[3] Debiasi, R.L.; Tyler, K.L. *Nat. Clin. Pract. Neurol*., **2006***, 2*, 264.

[4] Petersen, L.R.; Martin, A.A. *Ann. Intern. Med.,* **2002***, 137*, 173.

[5] Chou, K.C., Current Medicinal Chemistry, In *Frontiers in Medicinal Chemistry* (Atta-ur-Rahman & Reitz, A.B. eds.), Bentham Science Publishers, The Netherlands, **2006**, Vol. *3*, pp. 455-502.

[6] Higgins, D.; Thompson, J.; Gibson, T.;Thompson J.D.; Higgins, D.G.; Gibson T.J. *Nucleic Acids Res.,* **1994***, 22,* 4673.

[7] Ginalski, K.; Elofsson, A.; Fischer, D.; Rychlewski, L. *Bioinformatics,* **2003***, 19,* 1015.

[8] Sali, A; Potterton, L.; Yuan, F.; van Vlijmen, H.; Karplus, M., *Proteins,* **1995**, *3,* 318.

[9] van Gunsteren, W.F.; Billeter, S.R.; Eising, A.A.; Hünenberger, P.H.; Krüger, P.; Mark, A.E.; Scott, W.R.P.; Tironi, I.G. In *Biomolecular simulation: the GROMOS96 manual and user guide*. Vdf Hochschulverlag ETHZ, Zurich, Switzerland, **1996**; pp. 1-1042.

[10] Guex, N.; Peitsch, M.C. *Electrophoresis,* **1997***; 18,* 2714.

[11] Laskowski, R.A.; MacArthur, M.W.; Moss D.S.; Thornton, J.M. *J. Appl. Cryst.,* **1993***, 26*, 283.

[12] Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R, *J. Mol. Biol.*, **1997**, *267*, 727.

[13] Eldridge, M.D.; Murray, C.W.; Auton, T.R.; Paolini, G.V.; Mee, R.P. *J. Comput. Aided Mol; Des*., **1997***, 11*, 425.

[14] Labuda, M.; Nuttall, P.A. *Parasitology*, **2004***, 129*, 221.

[15] Ago, H.; Adachi, T.; Yoshida, A.; Yamamoto, M.; Habuka, N.; Yatsunami, K.; Miyano, M. *Structure Fold Des.,* **1999***, 7,*1417.

[16] Hong, Z; Cameron, C.E.; Walker, M.P.; Castro, C.; Yao, N.; Lau, J.Y.; Zhong, W. *Virology*, **2001***, 285*, 6.

[17] Zhong, W.; An, H.; Barawkar, D.; Hong, Z. *Antimicrob. Agents Chemother*., **2003***, 47*, 2674.

[18] Love, R.A.; Parge, H.E.; Yu, X.; Hickey, M.J.; Diehl, W.; Gao, J.; Wriggers, H.; Ekker, A.; Wang, L.; Thomson, J.A. *J. Virol.*, **2003***, 77,* 7575.

[19] Lohmann, V.; Roos, A. ; Korner, F.; Koch, JO.; Bartenschlager, R. *J. Viral Hepat*., **2000***, 7*, 167.

[20] Sommadossi, J.P.; Lacolla, P. *Int. Patent Appl.,* WO 01/190121, **2001**. Idenix Pharmaceuticals.

[21] Bressanelli, S.; Tomei,L.; Rey, F.A.; De Francesco, R. *J. Virol.,* **2002***, 76*, 3482.

[22] Tao, Y.; Farsetta, D.L.; Nibert, M.L.; Harrison, S.C. *Cell*, **2002***, 111*, 733.

[23] Butcher, S.J.; Grimes, J.M.; Makeyev, E.V.; Bamford, D.H.; Stuart, D.I. *Nature*, **2001***, 410*, 235.

[24] Huang, H.; Chopra, R.; Verdine, G.L.; Harrison, S.C. *Science,* **1998***, 282*, 1669.

Received: 26 September, 2006 Revised: 17 November, 2006 Accepted: 17 November, 2006