Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
De novo RNA synthesis and homology modeling of the classical swine fever virus RNA polymerase

Pengwei Zhang\textsuperscript{a,b,1}, Jian Xie\textsuperscript{c,1}, Guanghui Yi\textsuperscript{d}, Chuyu Zhang\textsuperscript{a,*}, Rong Zhou\textsuperscript{b,*}

\textsuperscript{a} Institute of Virology, College of Life Sciences, Wuhan University, Wuhan, Hubei 430072, PR China
\textsuperscript{b} Central Laboratory, Guangzhou Children Hospital, Guangzhou, Guangdong, PR China
\textsuperscript{c} Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, Hubei, PR China

Received 26 October 2004; received in revised form 4 March 2005; accepted 7 March 2005
Available online 15 April 2005

Abstract

Classical swine fever virus (CSFV) non-structural protein 5B (NS5B) encodes an RNA-dependent RNA polymerase (RdRp), a key enzyme which initiates RNA replication by a de novo mechanism without a primer and is a potential target for anti-virus therapy. We expressed the NS5B protein in \textit{Escherichia coli}. The rGTP can stimulate de novo initiation of RNA synthesis and mutation of the GDD motif to Gly–Asp–Asp (GAA) abolishes the RNA synthesis. To better understand the mechanism of viral RNA synthesis in CSFV, a three-dimensional model was built by homology modeling based on the alignment with several virus RdRps. The model contains 605 residues folded in the characteristic fingers, palm and thumb domains. The fingers domain contains an N-terminal region that plays an important role in conformational change. We propose that the experimentally observed promotion of polymerase efficiency by rGTP is probably due to the conformational changes of the polymerase caused by binding the rGTP. Mutation of the GDD to GAA interferes with the interaction between the residues at the polymerase active site and metal ions, and thus renders the polymerase inactive.

© 2005 Elsevier B.V. All rights reserved.

Keywords: De novo RNA synthesis; Homology modeling; Classical swine fever; RNA polymerase

1. Introduction

Classical swine fever virus (CSFV) is highly contagious and often causes a fatal disease (classical swine fever (CSF)), previously referred to as hog cholera, in susceptible pigs of all ages. The epidemic of CSF always leads to devastating financial losses, especially in a high pig density area, so it is classified as an Office International des Epizooties (OIE) List A disease. European Union (EU) pursues a straightforward, non-vaccination policy via stamping out suspected and infected herds since there is no efficient anti-CSFV vaccines (de Smit \textit{et al.}, 2001). However, this policy is disputed because in 1997 in the Netherlands outbreaks of CSF brought total financial consequences of US$ 2.3 billion for combatting the major CSF epizootic (Meuwissen \textit{et al.}, 1999). In the future, controlling CSF outbreaks might be based on therapy. At the same time, CSFV shows greater similarity in genome structure and RNA synthesis strategy to hepaciviruses than to flaviviruses (Lindenbach and Rice, 2001). Because hepaciviruses are difficult to grow in cell culture, CSFV, a normative representative in pestiviruses, has been used as a model system for studying hepaciviruses. So there is an urgent need to develop infection mechanism and antiviral research about CSFV.

CSFV belongs to the genus pestivirus of the family Flaviviridae (van Regenmortel \textit{et al.}, 2000). Like other members of the family, including the genus flavivirus and the genus hepacivirus, pestiviruses are small enveloped viruses containing a positive single-strand RNA genome. The CSFV genome of about 12.5 kbp contains a single open reading frame (ORF) encoding for a polyprotein of approximately 4000 amino acids and initiating at an internal ribosome

\textsuperscript{1} Corresponding authors. Tel.: +86 27 87682833; fax: +86 27 87883833. E-mail address: chuyuzhang@yeah.net (C. Zhang).

\textsuperscript{2} Authors contributed equally to this work.
entry site (IREs). The polyprotein is processed into a total of 12 viral structural and non-structural proteins. The gene product order along the ORF is NH2–Npro–C–E1–E2–p7–NS2–NS3–NS4A–NS4B–NS5A–NS5B–COOH (Rice, 1996). The 5′ UTR contains the IRES for cap-independent translation of the viral polyprotein (Rijnbrand et al., 1997; Fletcher and Jackson, 2002). In the 3′ UTR may contain replication signals involved in minus-strand RNA synthesis (Yu et al., 1999). Non-structural proteins are speculated to be components in the viral replication cycle. Replication of CSFV genome proceeds in two steps: synthesis of complementary minus-strand RNA using the genome as template and the subsequent synthesis of genomic RNA using this minus-strand RNA template. The key enzyme involved in both of these steps is a virally encoded RNA-dependent RNA polymerase (RdRp). This viral protein (non-structural protein 5B (NS5B)), which is located at the extreme C-terminus of the polyprotein, contains motifs shared by RdRps, such as the Gly–Asp–Asp (GAA) motif, which is highly conserved among RdRps (Koomin, 1991) and has been demonstrated to possess RdRp enzymatic activity. CSFV RdRp is able to initiate replication de novo without the requirement for protein or nucleic acid primers (Kao and Sun, 1996; Kao et al., 1999, 2001).

Sequence alignments of viral RNA-directed polymerase (reverse transcriptases and RdRps) have identified several conserved sequence motifs that are important for biological functions and shared among these enzymes. Now the data of crystal structures of RdRps from different RNA viruses can be acquired, including reovirus (Tao et al., 2002), calicivirus (Ng et al., 2002), poliovirus (Hansen et al., 1997), Φ6 (Butcher et al., 2001) and hepatitis C virus (HCV) (Ago et al., 1999; Bressanelli et al., 2002). Those studies have elucidated for key aspects of the structural biology of RdRps and confirmed the hypothesis that RdRps share a common architecture and mechanism of polymerase catalysis (Kamer and Argos, 1984). So, we can make the CSFV RdRp protein tertiary structure model based on the homology-modeling method by using these data to get the deeper and wider information of RdRp.

RdRps function as the catalytic subunit of the viral replicase. They are required for the replication of the viruses, so it is an often sought target in the search for antiviral. The development of effective drugs directed against the reverse transcriptase of HIV-1 RT highlights the importance of polymerases as drug targets (Kohlstadt et al., 1992; Larder and Stammers, 1999; Merlazzi et al., 1990; Mitsuya et al., 1990; O’Reilly and Kao, 1998; Smerdon et al., 1994). Moreover, the success obtained with polymerase inhibitors in the treatment of viral infections of human hepatitis B virus (HBV) and HCV provide a basis for designing a reasonable antiviral drug. Thus, CSFV RdRp is an attractive target for development of anti-CSFV drugs.

The work presented here includes expressing, purifying and functional analysis of a recombinant NS5BΔ24 fusion protein from Escherichia coli BL21 (DE3). The fusion protein was demonstrated to have the ability to initiate de novo either plus- or minus-strand viral RNA synthesis in a primer-independent manner and to specifically interact with viral RNA templates. In order to understand the structural basis of RdRp enzymatic activity and potential drug susceptibility, we also compared the sequence of CSFV polymerase with those of HIV-1, reovirus, calicivirus, poliovirus, Φ6 and HCV polymerases, situated in the conserved sequence motifs that are shared among RdRps. In that respect, CSFV shows great similarity in genome structure and RNA synthesis strategy to hepaciviruses and there are prolific structural and biochemical data on HCV polymerase, we built a three-dimensional model of the catalytic domain of CSFV RNA polymerase based on the conserved motifs in RdRps and HCV polymerase crystal structures. The validity of the model developed in the present study is supported by its ability to explain some of the key biochemical data. Analyzing conformational changes of the model and roles of the specific residues in the polymerization mechanism, we addressed some results which are likely to provide guidance in the design of future biochemical experiments and aid the development of anti-CSFV agents.

2. Materials and methods

2.1. Expression and purification

The CSFV (strain Shimen) NS5B construct employed for this study has 24-amino acid C-terminal deletion (NS5BΔ24) and a six-histidine C-terminal tag in a recombinant soluble form. The truncated gene product was cloned into the expression vector pET-28 to construct the pET–NS5BΔ24. Site-directed mutagenesis of GDD to GAA, containing the double substitution of both Asp448 and Asp449 to alanine, generated the mutant pET–NS5BΔ24GAA. CSFV NS5B was expressed from E. coli BL21 (DE3) and purified following the procedure described in the previous report (Xiao et al., 2002).

2.2. SDS-PAGE and Western blot

Protein fractions from the HisTrap affinity column were separated by 12% SDS-PAGE and electrophoresed to a nitrocellulose membrane. The membrane was blocked with 3% BSA in NaCI/Pi and treated with rabbit anti-swine serum infected with CSFV. Alkaline phosphatase (ALP)-conjugated goat anti-rabbit IgG was used as the secondary antibody. Membrane-bound antibodies were detected with Nitro Blue tetrazolium/5-bromo-4-chloroindol-2-y1 phosphate.

2.3. In vitro RdRp

RNA templates were prepared by in vitro transcription as described previously (Wu et al., 2003). The in vitro RdRp standard assay was performed in a total volume of 50 μL, containing 0.25 mM of each NTP, 0.3 μg of RNA template, 0.1 μg of purified protein, etc. The reaction mixtures were
incubated at 25°C for 2h and stopped by the addition of 20 mM EDTA. The RNA products were extracted with acid phenol/chloroform (1:1, v/v) followed by ethanol precipitation. Then, the precipitates were dissolved with either 20 μL of diethyl pyrocarbonate-treated water or denaturing buffer.

2.4. Northern blot analysis

The precipitated RdRp products were dissolved in a denaturing buffer and separated by PAGE (8% gel). Then, the precipitates were dissolved with either 20 μL of diethyl pyrocarbonate-treated water or denaturing buffer. The membrane was dried and exposed to ultraviolet irradiation. Hybridization was performed overnight in a solution containing the appropriate DIG-labeled RNA transcripts. The excess probes were eliminated gradually by washing the membrane from low-stringency to high-stringency. Then, the bound RNA was treated with ALP-conjugated anti-DIG Ig (1:5000) for 30 min. The reaction complexes were visualized using Nitro Blue tetrazolium/5-bromo-4-chloroindol-2-yl phosphate, according to the manufacturers of the DIG RNA Detection Kit (Roche).

2.5. The sequence analysis

Nucleotide and amino acid sequences were compiled and analyzed using the DNASTAR and Vector NTI Suite 6 (Informax, North Bethesda, MD) programs. Amino acid sequences were scanned for known active site motifs and protein family signatures (PROSITE 13.0; Bairoch, 1991). The DNA and the deduced amino acid sequences were compared with the updated GenBank/EMBL/DDBJ, SWISSPROT and PIR databases using FASTA and BLAST network service with the updated GenBank/EMBL/DDBJ, SWISSPROT and PIR databases using FASTA and BLAST network service (Altschul et al., 1997). Protein alignments were generated using the CLUSTAL program (Higgins and Sharp, 1988).

The second structure was predicted with the PHD (Rost et al., 1994) and JPred (Cuff et al., 1998).

2.6. Sequence alignments

Step 1 (Templates selecting and motifs identifying). We chose the HCV RdRp structures (Agrawal et al., 1999; Bressanelli et al., 1999, 2002; Wang et al., 2003) as our templates. But the CSFV RdRp sequence has low-sequence similarity to the HCV RdRp, so any single conventional method of sequence alignment is not enough.

Comparison of the crystal structures of HCV, poliovirus, reovirus, FMDV and HIV-1 polymerases allowed us to align both the structures and primary sequences and identify the conserved regions and insertions into the HCV RdRp. In cases where constraint space programming (CSP) does not give a satisfying solution and for loops above 10 residues, the Accesser/MSI loop database was used to change or improve the improperly created regions in the CSFV RdRp model. The

Step 2 (The building of the target–template alignment).

(a) Matching. The sequence of CSFV RdRp (718 amino acid residues; strain shimen; NCBI Accession No. AF157635) was aligned with that of representatives of the other three groups of Flaviviridae, and of five viral RdRps whose crystal structures are known. All sequences were aligned by seven alignment programs. These programs are: ClustalW (Francois et al., 1998), SIM4 (Huang and Miller, 1991), SAM-T02 (Kevin et al., 1998) and 3D-PSSM (Lawrence et al., 2000), CD-search (Aron et al., 2002), Superfamily (Gough et al., 2001; Julian and Cyus, 2002) and Block-Maker (Hendekoff et al., 1995). All methods were used with the default parameters provided by the authors. Then, the pairwise alignments between the target and template sequences were extracted, leading to different pairwise alignments between the target and the template.

(b) Database building. Each position of the alignments was stored in a database, all the redundant results, i.e., the same amino acid placed at the same position by different programs, being scored in a frequency table.

(c) Screening. Extracting the consensus between different methods may increase the overall confidence of the predictions tremendously. The position with the highest score was taken as the first anchor point to build the final target–template alignment. Incompatible results, aligning regions located up- and down-stream anchor points, were removed from the database. The process was pursued, new anchor positions being determined, and incompatible regions being eliminated, until all results were selected or removed.

(d) The final target–template alignment anchor points were thus composed by the most frequent aligned positions, under the condition of compatibility. These structurally conserved regions (SCRs) will match residues of template according to the polymerases structure motifs. In less conserved regions or regions containing insertions or deletions, we adjusted the alignment of the predicted secondary structure of CSFV RdRp with the secondary structures of other viral RdRps and the properties of amino acids (hydrophobic or hydrophilic character).

2.7. Model building and analysis

The Swiss-model was used to convert the CSFV RdRp sequence into 3D structures based on the above alignment. The best loop was selected using a scoring scheme, which accounts for force field energy, steric hindrance and favorable interactions like hydrogen bond formation. If no suitable loop can be identified, the flanking residues were included to the rebuilt fragment to allow for more flexibility. In cases where constraint space programming (CSF) does not give a satisfying solution and for loops above 10 residues, the Accesser/MSI loop database was used to change or improve the improperly created regions in the CSFV RdRp model.
reconstruction of the model side chains was based on the weighted positions of corresponding residues in the template structures. Starting with conserved residues, the model side chains were built by iso-sterically replacing template structure side chains. Possible side chain conformations were selected from a backbone dependent rotamer library, which has been constructed carefully taking into account the quality of the source structures. A scoring function assessing favorable interactions (hydrogen bonds and disulfide bridges) and unfavorably close contacts was applied to select the most likely conformation. A short (200 steps) minimization procedure was performed in GROMOS 96 (Gunsteren et al., 1995) to remove undesirable interactions that had been generated by the modeling process. Validation of the structures was done by using the PROCHECK (Laskowski et al., 1993), WHATCHECK (Vriend, 1990), VERIFY-3D (Eisenberg et al., 1997) and ANOLEA (Francisco and Ernest, 1998) programs. These check results indicated that the molecular geometry of the model is of good quality. Secondary structure assignments for the final model agreed well with the secondary structure predicted from the sequence using the PHD program.

Solvent accessible surface areas were computed using the XPLOR program (Brunger, 1992) with default parameters. The electrostatic potential was calculated and mapped to the surface using numerical integration of the Poisson–Boltzmann equation as implemented in the Swiss-Pdb Viewer.

Models of the RNA–RNA template-primer, rNTP were built based on the structures of HCV RdRp in its complexes with nucleic acid and NTP or dNTP substrates. The corresponding structures of these complexes were superimposed onto the structural model of CSFV RdRp based on structural alignments of the palm subdomains.

3. Results

3.1. Expression and purification of the recombinant NS5B proteins

Two vectors including pET–NS5B Δ24 and pET–NS5B Δ24GAA expression plasmids were examined to obtain the soluble recombinant NS5B protein from E. coli transformants. However, the full-length NS5B was expressed at a low level and difficult to purify.

NS5B is responsible for genome replication as a part of a larger, membrane associated, replicase complex (Lindenbach and Rice, 2001). With sequence analysis and analysis of the hydrophathy profile, we found that the C-terminal part of NS5B contains a highly hydrophobic region which is predicted to be an anchoring domain (Lindenbach and Rice, 2001). It has been reported that their deletion increases the solubility of NS5B expressed in E. coli (Ferrari et al., 1999; Lai et al., 1999; Tomei et al., 2000; Yamashita et al., 1998; Zhong et al., 2000). Therefore, the effect of the C-terminal region was examined by constructing the expression plasmids, pET–NS5B Δ24 and pET–NS5B Δ24GAA, which lacked the C-terminal 24 amino acid residues. An approximately 75-kDa protein corresponding to the NS5B Δ24 was expressed in the E. coli transformants (Fig. 1A). The protein was identified as the recombinant NS5B Δ24 by Western blot analysis using CSFV-infected pig serum as primary antibody (Fig. 1B). The other mutant protein, NS5B Δ24GAA, was expressed and purified in parallel to the NS5B Δ24 protein.

3.2. RNA synthesis activity of the recombinant NS5B proteins

The activities of the NS5B proteins (NS5B Δ24 and NS5B Δ24GAA) were tested using different templates including 3′-end of minus-strand and plus-strand RNA transcripts. The RNA products synthesized by CSFV NS5B Δ24 were separated by denaturing PAGE (8% gel) and detected using a Northern blot assay. Whichever RNA template was employed, the activity of the NS5B Δ24 has been shown to be primer-independent (Fig. 2). The predominant RNA products migrated similarly to the respective RNA templates (373 nucleotides for synthesized plus-strand RNA and 228 nucleotides for synthesized minus-strand RNA). The mutated NS5B protein, NS5B Δ24GAA, in which the GDD sequence was substituted to GAA did not exhibit any RdRp activity in the presence of templates (Fig. 2). This result shows that
Fig. 2. De novo initiation of viral RNA synthesis by CSFV RdRp. Northern blot assay with viral plus-strand RNA templates (A) and viral minus-strand RNA template (B). Lane 1, RdRp assay with NS5B/Delta124GAA as a control and lane 2, RdRp assay with NS5B/Delta124 and RNA template.

3.3. Preference of de novo RNA synthesis

We used the 3' end of minus-strand RNA transcripts as the template and changed its concentrations with a gradient for RdRp assay. The result shows that a high concentration of the RNA template, above a certain level, had no significant promotion of RNA synthesis (Fig. 3A).

The possibility that NTP may affect the activity of NS5B on the initiation of RNA synthesis was addressed by preincubating the protein with 0.5 mM of each NTP for 30 min at first, and further incubating for 90 min with 0.25 mM NTP as substrates. Compared with the RdRp activity for reaction without preincubation, higher activity of de novo RNA synthesis was obtained when the enzyme with plus-strand or minus-strand RNA template were preincubated with NTP (Fig. 3B and C). Futhermore, preincubation with 0.5 mM GTP can activate CSFV RdRp regardless of template properties, whereas preincubation with 0.5 mM ATP or UTP resulted in higher activity using 3'-end of plus-strand or minus-strand RNA template, respectively.

3.4. Molecular architecture

We built three-dimensional models through homology modeling based on the conserved motifs in RdRps and HCV crystal structures (Figs. 4 and 5). The NS5B of CSFV is a multi-domain protein whose sequence of domains modeled contains 605 residues. The protein contains an N-terminal deletion of 90 and C-terminal deletion of 23 amino acid residues. The function of this N-terminal domain is not known. The C-terminal residues are the highly hydrophobic and predicted to be membrane-anchoring region. Like poliovirus polymerase, HCV polymerase, Phi6 and calicivirus RNA polymerase, its structure shows an elaborate arrangement of polymerase domains that have been termed “fingers,” “palm,” and “thumb,” on the basis of its resemblance to a right hand (Fig. 6A). The CSFV RdRp overall fold contains 23 α-helices and 17 β-strands. It presents a deep cleft in the middle with the palm at the base. The fingers and thumb domains are both important for correctly positioning the substrates for catalysis by the palm domain.

The fingers domain contains amino acids 91–314 and 351–409. Although the fingers domain of the CSFV polymerase contains about 30 more residues than the fingers of HCV, both polymerases still appears to be a conservation of tertiary structure rather than primary sequence. It con-
P. Zhang et al. / Virus Research 112 (2005) 9–23

Fig. 4. Sequence alignment of polymerases of CSFV and HCV. The CSFV NS5B sequence has been truncated in the N-terminus and C-terminus. The HCV sequence has been truncated in the C-terminus. These sequences were aligned as described in Section 2. Asterisks below alignment indicate identical residues and dots indicate similar residues. Lowercase letters indicate secondary structure assignment according to the PHD predicted results and PDB data (h, helix and s, strand). The sequence in the fingers and palm domains containing the conserved motif is underlined with different colors. Strictly conserved residues are marked with solid red triangles. Residues mostly conserved are marked with hollow red triangles.
Figure 5. Stereoview of the polymerase active site and the template/rNTP binding site. The model has been rotated 30° clockwise around a parallel axis. The conserved sequence motifs (A–I) are highlighted by different colors. rNTP substrates is shown in violet and the template is in brown. Three conserved aspartates, Asp345, Asp448 and Asp449 (shown with side chains) locate at the catalytic active site and coordinate with divalent metal ions (shown as gray spheres).
Fig. 6. Overall views of CSFV NS5B polymerase and comparison with other polymerases. (A) Ribbon representation of NS5B with domains colored according to N-terminal (pink), fingers (red), palm (yellow) and thumb (blue). Secondary structural elements are labeled. The orientation is the standard front view of the enzyme. Comparison of RdRps about the N-terminal (B), palm (C) and thumb (D) domains in CSFV, HCV, poliovirus and calicivirus RdRps. Part of the fingers domain of the poliovirus polymerase structure is mostly disordered.

Motif D, like motif B, is conserved in sequence only in the RNA-directed polymerases. Motif E occurs only in RNA-dependent RNA and RT (O’Reilly and Kao, 1998). Motifs A–E correspond to residues 338–357, 400–427, 439–456, 466–483 and 489–507 in CSFV RdRp, respectively. Motif A contains two aspartate residues (Asp345 and Asp350). Asp345 is responsible for binding the catalytic metal ions; Asp350 is primary substrate discriminator for rNTP over dNTP. The Asp345 near the end of the β-strand is completely conserved in all classes of polymerases and positions the two aspartates (Asp448 and Asp449) close to the conserved aspartate of motif A. In motif E, the hydrophobic residues are important for the interactions with the palm core structure and account for the conservation of several hydrophobic residues in motifs A, C and D of RNA-dependent polymerases (Hansen et al., 1997; Jacobo-Molina et al., 1993).

The thumb domain of CSFV RdRp is composed of seven α-helices (labeled Q to W) and four β-strands (labeled 14–17) at the most C-terminal position. It contains a motif I (511–526) consisting of an α-helix and a β-strand. Arg517
is strictly conserved residue in the motif, and is essential for GTP binding which can raise the enzymatic efficiency. The corresponding thumb region in HCV is mainly α-helical, containing eight α-helices and four β-strands (Fig. 6D). The more α-helix (residues 460–466) in HCV RdRp is located between β16 and α17, while in CSFV RdRp a long loop (residues 585–598) replaces it. Although the difference exists, the structure of thumb domain is not affected because there are three α-helices (αQ, αS and αU), which are comparatively conserved in RdRps, form the core of the thumb in CSFV RdRp. This domain reveals a striking difference to that of the thumb domain of calicivirus, poliovirus and Φ6 polymerases, which only contain the core of the domain, thus have a smaller structure (Fig. 6D). In calicivirus, the thumb domain (residues 418–501) comprises four α-helices and two long loops. The first loop links together the first and second α-helices and the second loop links together the third and fourth α-helices (Ng et al., 2002). The long loop (residues 428–446) connecting the first and second helix is folded away from the template cleft and packs against the N-terminal of the fingers domain. In contrast, the corresponding parts of CSFV (residues 529–540) and Φ6 polymerases (residues 526–547) are long loops rich in positively charged residues near the gate of the template tunnel, while in HCV (residues 401–407) and poliovirus (residues 405–409) the parallel structures are shorter turns. Moreover, CSFV RdRp, similar to HCV RdRp, has a long β-hairpin (β16–17) connecting the third and fourth helices (αS and αT) and occluding the active site cleft. This hairpin acts as a discriminator for distinguishing the single-stranded RNA templates from double-stranded RNA (Hong et al., 2001). Without the β-hairpin structure, calicivirus, poliovirus and Φ6 enzymes are consistent with the ability of these enzymes to utilize double-stranded RNA as templates (López Vázquez et al., 2001; Butcher et al., 2001). Another function of this domain is to form a hydrophobic binding pocket near the domain core (αQ, αS and αU) with the help of the palm domain and two long loops of the fingers domain. The rNTPs are bound in a wedge-like fashion to the hydrophobic binding pocket.

### 3.5. N-terminal regions

The long loop at the N-terminal of CSFV NS5B forms a bridge between the fingers and thumb domains, which do not interact with each other directly. This region in caliciviral RdRp (residues 1–64), HCV RdRp (residues 1–49), poliovirus RdRp (residues 12–37 and 67–97) and Φ6 RdRp (residues 1–105) also form a bridge between the fingers and thumb domains (Fig. 6B). The conformational flexibility of the region, containing several loops, shapes a cavity with the help of the fingers and thumb domains for the binding of RNA, and what is more, this region for template selectivity may also limit the conformational change of the RdRp.

In CSFV RdRp, the structure of the polypeptide chain and interaction with the fingers and thumb domains differ substantially from other RdRps. The exact region (helix A) connecting the fingers and thumb domains is fixed in a groove among parallel α-helices Q, S and U of the thumb domain in CSFV RdRp, while the corresponding segment of calicivirus RdRp is part of a loop connecting two short β-strands of N-terminal region (Fig. 6B). In HCV RdRp, the conformation of the last several residues (loops 41–45) differs from what is seen in the CSFV RdRp whose corresponding segment (loops 136–147) protrudes from the fingers domain surface similar to caliciviral RdRp.

The N-terminal polypeptide region of poliovirus polymerase contains two ordered parts: residues 12–37 at the back of the thumb domain and residues 67–97 beneath the fingers domain. The residues (residues 38–66) that join these two segments are disordered in the crystals. Residues 12–23 extend as a single polypeptide-strand from the active site cleft up across the top of the thumb domain. In contrast, the corresponding segment of CSFV RdRp extends away from the active site cleft, leaving room for incoming RNA. The N-terminal of Φ6 RdRp is most complex among these polymerases, because it consists of three α-helices and five β-strands. An α-helix and two β-strands are fixed at the back of the thumb domain.

### 4. Discussion

#### 4.1. Model building

The CSFV RdRp shares <30% sequence identity with other viral RdRps and RTs. Normally, such a low level of homology would not permit reliable sequence alignment and homology modeling. However, we applied a stepwise protocol that relied on manual identification of key conserved motifs and used them as landmarks to guide subsequent alignment of primary sequence.

There are four main methods to solve multiple alignment computational complexity (Thompson et al., 1999): progressive global alignment; iterative methods; alignments based on locally conserved patterns (PSSM/profile or block); statistical methods and probabilistic models (such as Gibbs sampler and Hidden Markov models). Now no alignment method can be qualified as the absolute most reliable one. Every alignment method seeks a balance between sensitivity and selectivity. Extracting the consensus between different methods may increase the overall confidence of the predictions tremendously (Birnstiel et al., 1999; Julie et al., 1999). So, in our subsequent alignment, alignments are obtained by combining, weighting and screening the results of seven multiple alignment programs which include all aspects described above. These methods include: ClustalW is a progressive global alignment method; PSI-BLAST is an iterative search tool (Stephen et al., 1997). 3D-PSSM, CD-search, Superfamily, Block-Maker are based on locally conserved patterns (PSSM/profile or block); SAM-T02 is a HMM tool and Block-Maker consists of Gibbs sampler.
Fig. 7. The rGTP complex. The position of the rGTP sites were derived by superposition of the HCV polymerase structure. The CSFV NS5B/rGTP complex has been rotated 180° clockwise around a parallel axis in (A) and (B). (A) A ribbon diagram of the complex structure. The rGTP is in brown. (B) A stereoview of the rGTP-binding site of NS5B/rGTP complex aligned with the structure of NS5B. The residues involved in the rGTP-binding site and the rGTP are depicted as stick models. The rGTP and Mn2+ ion are shown in brown and grey, respectively. Active site residues are labeled and shown in cyan. The hydrogen bonding pattern between them is shown by green dot line. (C) CSFV NS5B without any ligand. (D) CSFV NS5B/rGTP complex structure. The rGTP’s van der Waals surfaces are in brown. These structures have been rotated 60° forwards around a parallel axis. The C-terminal has been removed for clarity. The protein is colored according to Fig. 6A.

At last, crucial for the sequence alignments were the prediction of the secondary structure of CSFV RdRp and the appropriate alignment of the predicted secondary elements of CSFV RdRp with the secondary structures of PV, HCV, RHDV, RV and Φ6 RdRps.

After model building, several methods are used for the detection of errors in protein models. Two of them, PROCHECK and WHAT-CHECK, are packages that check the stereochemical quality of the model, the requirement for a good structural prediction. Another method is widely used: VERIFY-3D is based on several statistically derived preferences and on the accessible surfaces of the amino acid residues. The method uses amino acid interactions, and it is very accurate during the initial stages of the model building process. A complementary method able to improve the quality of the models produced by VERIFY-3D is ANOLEA (Melo and Feytmans, 1998). The method is based on a statistical atomic mean force potential (AMFP) that involves only short-range and non-local interactions between heavy atoms of the standard amino acid residues. The last models were evaluated using these check programs and were found to satisfy all criteria.

4.2. The rGTP binding site

Some RNA polymerases (Butcher et al., 2001; Kao and Sun, 1996; Luo et al., 2000; Nomaguchi et al., 2003; Ranjith-Kumar et al., 2002) require rGTP for initiating RNA synthesis. The experimental results reported above show that rGTP resulted in higher activities of CSFV RdRp when the minus-strand or plus-strand RNA template was used. To explain these results, we made the superposition of the HCV NS5B/rGTP complex onto the model of CSFV RNA polymerase. There is an rGTP binding site at the back of the thumb domain as HCV. The active surface of rGTP is about 31 Å from the catalytic part which is on the bottom of the template-binding groove. Residues, which are located at the loop A1 of fingers, the helices of
The Mn$^{2+}$ ions are displayed as for Fig. 7. The color code by domains of the rGTP (brown) and the two rNTPs (green) are for the H9251. In addition, the hydrogen bonds between the loop wall, with which Arg127 and Trp623 constitute a pocket for a coping cooperates with Gly621 and Ala622 in building a guanine base. Near the ribose of the rGTP, Ser618 and Thr627 provide a hydrophobic platform for orientating the guanine perpendicular to the plane of Trp623 side chain. Thr624 and hydrogen bond to the N2 of the guanine, which is roughly of Trp623 on the loop between the Bressanelli et al., 2002). Its bidentate hydrogen bonds to the Arg32 seems to be an important specificity determinant of interactions involving fingers, thumb and rGTP. In HCV, the fingertips (Cys126 and His130) gives rise to a network and only Thr624 and Thr627 replace the two proline rings (P495 and P496) and Val499. In contrast, the polymerase turns into an opened conformation when the concentration reached a certain level (Fig. 3A). This limitation of template concentration may be caused by the specific structure of CSFV RdRp. The position of the template tunnel is based on the $\Phi 6$ polymerase-template complex. Superposition of the central palm domain of CSFV RdRp with the corresponding region in $\Phi 6$ polymerase brings the fingers domains of two polymerases into good agreement. The entrance to the template channel in CSFV RdRp lies between the fingers and the thumb where $\Omega$ fingers and fingertips encircle a U-shape valley (Fig. 8). Motif H and the preceding long loop floats near the gate of the RNA template had no significant assist on RNA synthesis as a result (Bressanelli et al., 2002). The rGTP binding at this site, in company with template RNA binding groove, helps to induce a conformational change of the enzyme that would be important for initiation (Ago et al., 1999; Blumenthal and Hill, 1980; Bressanelli et al., 1999, 2002; Luo et al., 2000; Oh et al., 1999). We constructed another model of the polymerase with rGTP according to the structure of HCV RdRp containing an rGTP. Comparing this model with the unliganded polymerase model (see Section 2), we find that before the polymerase contacts with the rGTP, the $\beta$-hairpin (i16–17), motif H and the preceding loop stay at the gate of the template tunnel and part of the loop A2 project into the tunnel, and thus this closed posture prevents templates from coming into the catalytic site (Fig. 7C). On the contrary, the polymerase turns into an opened conformation by binding an rGTP at the site. The rGTP grips the $\Omega$-helices of the thumb domain and $\alpha$-helix A. As a result, the $\beta$-hairpin is drawn from the tunnel in succession. Because of the joint, the loop A1 containing the $\alpha$A, the conformation changing of the thumb domain causes motif H and the preceding loop of the fingers domain to extend away from the tunnel, leaving room for RNA template (Fig. 7D). Put another way, when the unliganded polymerase links to an rGTP at the site, it experiences the process of closing the hand to opening the hand.

4.3. Template tunnel

We observe that the increase in template concentration of the RNA template had no significant assist on RNA synthesis when the concentration reached a certain level (Fig. 3A). This limitation of template concentration may be caused by the specific structure of CSFV RdRp. The position of the template tunnel is based on the $\Phi 6$ polymerase-template complex. Superposition of the central palm domain of CSFV RdRp with the corresponding region in $\Phi 6$ polymerase brings the fingers domains of two polymerases into good agreement. The entrance to the template channel in CSFV RdRp lies between the fingers and the thumb where $\Omega$ fingers and fingertips encircle a U-shape valley (Fig. 8). Motif H and the preceding loop have eight more residues including the three positively charged arginines than HCV containing a loop instead. As a result, the gate of the cavity appears smaller than HCV polymerase. The long loop floats near the gate of the RNA template and part of the loop A2 occludes the template tunnel from the binding of template-rNTP duplexes (Fig. 8). When the enzyme begins to react, this segment moves out of the template cleft moderately to allow RNA and nucleoside triphosphate to bind to the enzyme. These characters explain that the efficiency of replication cannot be heightened by changing the template concentration above a certain level. HIV-RT, calicivirus,
poliovirus and Phi6 polymerases using an rNTP or protein-linked primer and double-stranded template, have a smaller thumb domain, and thus, the capability of the tunnel is greater then CSFV RdRp. The fingers and thumb domains within these polymerases differ significantly and bring on facilitation of different initiation mechanisms (Hobson et al., 2001).

In CSFV RdRp, motif H, motif F, the loop A1 in the fingers domain, the beta-hairpin and the N-terminal loop of the uR in the thumb domain form a long shallow trench fitting the phosphate backbone and ribose moieties, then direct the single-stranded RNA towards the active site cavity. Some positively charged side chains should be good acceptors for the negative charge of the template backbone. Four side chains (Arg218, Arg219, Lys263 and Arg285) among these residues in the fingers domain appear strictly conserved in most of the pestiviruses and all HCV genotypes (Bressanelli et al., 1999; Collett et al., 1988). The groove in CSFV is smaller than HCV, but it is still large enough to allow the template-strand to enter the active site.

The selectivity of RdRps for template binding and initiating RNA synthesis de novo may be controlled by the certain conformational changes. The thumb/fingertips interaction could play an important role in maintaining the fingers domain in the closed form in the absence of ligands via helix A, helix G, motif F and loop A1 because the loop A1 is at the tip of a long loop that could adapt to this type of change. When the RNA template presents, because of the flexibility of the loop A1, it is in the opened form for the bigger entrance and concave. This modest movement on binding of template seems necessary because there would not be enough room for an RNA single-stranded molecule to fit in the structure. The sequence alignment data suggested that the fingers domain also represents a common structural discriminator, which distinguishes RNA from DNA in (+) ssRNA virus RdRps (Ago et al., 1999).

4.4. Catalytic site

Superposition of the conserved central palm domain of Phi6 polymerase-template-rNTP complex onto CSFV RdRp shows the stereo plot of all the ligands to the nucleotide in the catalytic site. In Phi6 polymerase-template-rNTP complex, there are two rNTP sites, i (the priming site) and i + 1 (the catalytic site) surrounding the GDD residues essential for RNA synthesis. According to this structure data, we finished a model of CSFV RdRp with an RNA template and two rNTPs that are base-paired with the template bases. One rNTP is at the i site selected by base pairing with the 3'-end terminal base of the template (Fig. 8).

The rNTP contacts the residues, forming a platform on which an initiation complex could be constructed, with electrostatic, hydrophobic interactions and hydrogen bond at the i site (Fig. 8). In CSFV RdRp, dimensional position of these residues contributes to the orientation of the rNTP. Thus, the rNTP may mimic a primer for initiating the RNA synthesis and is held at the i site in right orientation for making a Watson-Crick base pair with the 3'-base of the template (Figs. 8 and 9A). When the nucleotide is orientated as the primer at the i site, the 3'-OH group of the nucleotide would be placed for in-link attack on the α-phosphates of the incoming rNTP at the i + 1 site. Some residues, Ser498, Arg517 and Thr521, appear strictly conserved among Flaviviridae family, and Lys525 is conserved throughout all pestiviruses members (Bressanelli et al., 2002). This conservation of the polymerases is known to initiate replication de novo in some positive-stranded RNA viruses, such as brome mosaic virus (BMV) (Kao et al., 2000). In HCV, hydrogen bonds also occur between the surrounding residues and the rNTP, and these side chains hold the base in the appropriate position. Similarly, the initiation complex of the Phi6 polymerase contains a tyrosine residue (Y630) stacking against the base of the nucleotide at the i site, which acts as a protein platform to affect the base in the right orientation. A similar role for orienting the priming rNTP has been advanced for the "priming loop" of reovirus RdRp (Tao et al., 2002).

RdRps are metal-activated enzymes that use divalent metal ions for nucleotide polymerization. GDD is highly conserved in RdRps (Kamer and Argos, 1984; Koszin, 1991). Two aspartates of the GDD motif are involved in coordination of divalent cations. There appears to be a fairly strict requirement for these aspartates at this position in the RdRps (Inokuchi and Hirashima, 1987; Jablonski and Morrow, 1993; 1995; Lohmann et al., 1997; Longstaff et al., 1993; Sankar and Porter, 1992). We carried out a site-directed mutagenesis of GDD to GAA, containing the double substitution of both Asp448 and Asp449 to alanine by overlapping PCR. Mutation of these aspartates almost abolished RNA replication. Changing the aspartates of the model to alanines, we found the side chains of alanines are too far to coordinate with the Mn2+ ions.
cations, and thus the interaction between GDD and cations is destroyed (Fig. 9).

4.5. The mechanism for synthesizing RNA

RNA synthesis by RdRp involves template binding, initiation complex formation and transition from initiation to elongation. All these reactions go with the conformation changes, and are completed by the three domains of the polymerase. The fingers and thumb domains with elaborate structures are in the closed conformation when template is absent. An rGTP polymerase open the template tunnel, and thus the tunnel is large enough to allow the template-strand to enter the active site.

When the rNTP and the template go through the rNTP tunnel and the template tunnel respectively, they arrive at the catalytic site for forming a de novo initiation complex. The primer rNTP is held by the platform composed of several residues, and forms a Watson–Crick base pair with the 3′-end of the template base. This rNTP will be the first nucleotide of the nascent RNA-strand. The incoming rNTPs are selected by base pairing with the base of the template and discriminated from dNTP by Asp350. The 3′-OH group of the primer rNTP attacks the α-phosphates of the incoming rNTP to form a new phosphodiester bond. Successively, the nucleotidyl transfer reaction is repeated, and rNTPs are added to the nascent RNA-strand until the polymerase finishes copying the template.

Acknowledgements

This work was supported by National Basic Research Developmental Projects (G1999011900) and National Natural Science Foundation of China (3017214).

References

Agio, H., Adachi, T., Yoshida, A., Yamamoto, M., Hüboka, N., Yatsunami, K., Miyano, M., 1999. Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. Structure 7, 1417–1426.

Altshul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Anwari, M.B., Attali, R.P., Benjamin, A.S., Paul, A.T., Lewis, Y.G., Stephen, H.B., 2002. CDD: a database of conserved domain alignments with links to domain three-dimensional structure. Nucleic Acids Res. 30, 284–283.

Barouch, A., 1991. PROSITE: a dictionary of sites and patterns in proteins. Nucleic Acids Res. 19, 2241–2245.

Blumenfeld, T., Hill, D., 1980. Roles of the host polypeptides in Q beta RNA replication. Host factor and ribosomal protein S1 allow initiation at reduced GTP concentration. J. Biol. Chem. 255, 11713–11716.

Bressanelli, S., Tomei, L., Rey, F.A., De Francesco, R., 1999. Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. Proc. Natl. Acad. Sci. U.S.A. 96, 13034–13039.
Ito, J., Braultwaite, D.K., 1991. Compilation and alignment of DNA polymerase sequences. J. Mol. Biol. 222, 373–381.

Jablonski, S.A., Moroz, C.D., 1993. Enzymatic activity of poliovirus RNA polymerases with mutations at the tyrosine residue of the conserved GDD/GVGD motif and poliovirus contaminating RNA polymerases with FGGD and MGGD sequences. J. Virol. 67, 373–381.

Jablonski, S.A., Moroz, C.D., 1995. Mutation of the asparagine residues of the GDD sequence motif of poliovirus RNA-dependent RNA polymerase results in enzymes with altered metal ion requirements for activity. J. Virol. 69, 1532–1539.

Jacob-Molina, A., Ding, J., Nami, R.G., Clark Jr., A.D., Lu, X., Tani, J., Tajima, F., Williams, A., Ferrin, T.E., Clark, P., Hori, A., Hughes, S.M., Arnold, E., 1993. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. Proc. Natl. Acad. Sci. U.S.A. 90, 6230–6234.

Julian, G., Cys, C., 2002. SUPERFAMILY: HMMs representing all protein families of known structure. SCOP.

Julie, D.T., Frederic, P., Olivier, O., 1999. A comprehensive comparison of multiple sequence alignment programs. Nucleic Acids Res. 27, 2682–2690.

Kamer, G., Argos, P., 1994. Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. Nucleic Acid. Res. 12, 7269–7282.

Kao, C.C., Sun, I.H., 1996. Initiation of minus-strand RNA synthesis by the brome mosaic virus RNA-dependent RNA polymerase: use of oligo(homobacitamic acid) primers. J. Virol. 70, 6828–6830.

Kao, C.C., Del Vecchio, A.M., Zheng, W., 1999. De novo initiation of RNA synthesis by a recombinant Flaviviridae RNA-dependent RNA polymerase. Virology 253, 1–7.

Kao, C.C., Yang, X., Kline, A., Wang, Q.M., Barket, D., Heinz, B.A., Zhou, W., 1996. Initiation of minus-strand RNA synthesis by the RNA-dependent RNA polymerase in vitro: nucleotide, primer, and template parameters. J. Virol. 70, 8381–8482.

Oh, I.W., Ito, T., Lai, M.M., 1999. A recombinant hepatitis C virus RNA-dependent RNA polymerase capable of copying the full-length viral RNA. J. Virol. 73, 7694–7702.

O'Reilly, E.K., Kao, C.C., 1998. Analysis of RNA-dependent RNA polymerase structure and function as guided by known polymerase structures and computer predictions of secondary structure. Virology 252, 387–393.

Ramiah-Kumar, C.T., Giriabali, L., Kim, M.-J., Santhy, R.K., Kao, C.C., 2002. Requirements for de novo initiation of RNA synthesis by recombinant Flaviviridae RNA-dependent polymerases. J. Virol. 76, 12526–12536.

Rice, C.M., 1996. Flaviviridae: the viruses and their replication. In: Fields, B.N., Knipe, D.M., Howley, P.M. (Eds.), Fields Virology. Lippincott–Raven Publishers, Philadelphia, pp. 931–960.

Rijnbrand, R., van der Straaten, T., van Rijn, P.A., Spaan, W.J., Bredenbeck, P.J., 1997. Internal entry of ribosomes is directed by the 5’ non-coding region of classical swine fever virus and is dependent on the presence of an RNA pseudoknot up-strand of the initiation codon. J. Virol. 71, 451–457.

Rosi, B., Sander, C., Schneider, R., 1994. PHD—an automatic mail server for secondary structure prediction. Comput. Appl. Biosci. 10, 35–40.

Sankar, S., Poter, A.R., 1992. Point mutations which drastically affect the polymerization activity of ruxaphenomavirus and RNA-dependent RNA polymerase from the reovirus polymerase. J. Biol. Chem. 267, 10168–10176.

Smerdon, S.J., Jags, J.W., Kohlstadt, L.A., Chinito, A.J., Friedman, J.M., Rice, P.A., Steitz, T.A., 1994. Structure of the binding site for non-nucleoside inhibitors of the reverse transcriptase of human immunodeficiency virus type 1. Proc. Natl. Acad. Sci. U.S.A. 91, 3913–3915.

Stephen, F.A., Thomas, L.M., Aleixoas, A., Jengiu, Z., Zheng, Z., Webb, M., Webb, D.L., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Tao, Y., Fazeta, D.L., Nibert, M.L., Harmon, S.C., 2002. RNA synthesis in a capsid-structural studies of reovirus polymerase. J. Virol. 76, 733–745.

Thompson, J.D., Plozinski, F., Pech, O., 1999. A comprehensive comparison of multiple sequence alignment programs. Nucleic Acids Res. 27, 2682–2690.
Tomei, L., Vitale, R.L., Incitti, I., Serafini, S., Altamura, S., Vitelli, A., De Francesco, R., 2000. Biochemical characterization of a hepatitis C virus RNA-dependent RNA polymerase mutant lacking the C-terminal hydrophobic sequence. J. Gen. Virol. 81, 759–767.

van Regenmortel, M.H.V., Fauquet, C.M., Bishop, D.H.L., Carstens, E.B., Estes, M.K., Lemon, S.M., Maniloff, J., Mayo, M.A., McGlinchey, D.J., Pringle, C.R., Wickner, R.B., 2000. Virus Taxonomy: the Classification and Nomenclature of Viruses. The Seventh Report of the International Committee on Taxonomy of Viruses. Academic Press, San Diego, CA.

Vriend, G., 1990. WHAT IF: a molecular modeling and drug design program. J. Mol. Graph. 8, 52–56.

Wang, M., Ng, K.K., Chen, M.M., Chan, L., Yannopoulos, C.G., Bernard, J., Morris, N., Nguyen-Ba, N., Bethell, R.C., James, M.N., 2003. Nucleoside and non-nucleoside inhibitors bind to an allosteric site on HCV NS5B polymerase: crystal structures and mechanism of inhibition. J. Biol. Chem. 278, 9489–9495.

Wu, H.X., Zhang, C.Y., Zheng, C., Wang, J.F., Pan, Z.S., Li, L., Cao, S., Yi, G.H., 2003. Construction of cytopathic PK-15 cell model of classical swine fever virus. Chin. Sci. Bull. 48, 887–891.

Xiao, M., Zhang, C.Y., Pan, Z.S., Wu, H.X., Guo, J.Q., 2002. Classical swine fever virus NS5B-GFP fusion protein possesses an RNA-dependent RNA polymerase activity. Arch. Virol. 147, 1779–1787.

Yamashita, T., Kaneko, S., Shirota, Y., Qin, W., Nomura, T., Kobayashi, K., Mirakami, S., 1998. RNA-dependent RNA polymerase activity of the soluble recombinant hepatitis C virus NS5B protein truncated at the C-terminal region. J. Biol. Chem. 273, 15479–15486.

Yu, H., Graumann, C.W., Bethens, S.-E., 1998. Sequence and structural elements at the 3′ terminus of bovine viral diarrhea virus genomic RNA: functional role during RNA replication. J. Virol. 73, 3638–3644.

Zhong, W., Ingravallo, P., Wright-Minguez, J., Uoo, A.S., Statham, A., Ferrini, E., Lau, Y.Y., Hong, Z., 2005. RNA-dependent RNA polymerase activity encoded by GB virus-B non-structural protein 5B. J. Viral Hepatol. 7, 335–342.