Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus

Hideo Ago[1†], Tsuyoshi Adachi[1†], Atsuhito Yoshida[1], Masaki Yamamoto[2], Noriyuki Habuka[1], Kimio Yatsunami[1] and Masashi Miyano[1*]

Background: Hepatitis C virus (HCV) is the major etiological agent of hepatocellular carcinoma, and HCV RNA-dependent RNA polymerase (RdRp) is one of the main potential targets for anti-HCV agents. HCV RdRp performs run-off copying replication in an RNA-selective manner for the template–primer duplex and the substrate, but the structural basis of this reaction mechanism has still to be elucidated.

Results: The three-dimensional structure of HCV RdRp was determined by X-ray crystallography at 2.5 Å resolution. The compact HCV RdRp structure resembles a right hand, but has more complicated fingers and thumb domains than those of the other known polymerases, with a novel α-helix-rich subdomain (α fingers) as an addition to the fingers domain. The other fingers subdomain (β fingers) is folded in the same manner as the fingers domain of human immunodeficiency virus (HIV) reverse transcriptase (RT), another RNA-dependent polymerase. The ribose-recognition site of HCV RdRp is constructed of hydrophilic residues, unlike those of DNA polymerases. The C-terminal region of HCV RdRp occupies the putative RNA-duplex-binding cleft.

Conclusions: The structural basis of the RNA selectivity of HCV RdRp was elucidated from its crystal structure. The putative substrate-binding site with a shallow hydrophilic cavity should have ribonucleoside triphosphate (rNTP) as the preferred substrate. We propose that the unique α fingers might represent a common structural discriminator of the template–primer duplex that distinguishes between RNA and DNA during the replication of positive single-stranded RNA by viral RdRps. The C-terminal region might exert a regulatory function on the initiation and activity of HCV RdRp.

Addresses: [1]Central Pharmaceutical Research Institute, Japan Tobacco Inc., 1-1 Murasaki-cho, Takatsuki, Osaka 569-1125, Japan and [2]Structural Biophysics Laboratory, The Institute of Physical and Chemical Research (RIKEN), 1-1 Kouto, Mikazuki-cho, Sayo-gun, Hyogo 679-5148, Japan.

[†]These authors contributed equally to this work.

*Corresponding author.
E-mail: masashi.miyano@ims.jti.co.jp

Key words: HCV, hepatitis C virus, NS5B, RNA-dependent RNA polymerase, X-ray crystallography

Received: 11 August 1999
Revisions requested: 2 September 1999
Revisions received: 20 September 1999
Accepted: 1 October 1999

Published: 1 November 1999

Structure November 1999, 7:1417–1426

0969-2126/99/$ – see front matter

Introduction

Hepatitis C virus (HCV) is the major etiological agent of hepatocellular carcinoma. The World Health Organization has estimated that there are 170 million chronic HCV carriers in the world today, although the risk of acquiring HCV by transfusion has been reduced dramatically since the development of assays for the virus [1]. About 80% of HCV-infected patients develop chronic hepatitis, 20% of them progress to cirrhosis and eventually develop hepatocellular carcinoma. At present, no vaccine is available and there is no broadly effective therapy for all genotypes of HCV, although extensive research has been focused on the development of anti-HCV agents.

HCV is a positive single-stranded-RNA [(+)ssRNA] virus and a member of the *Flaviviridae* [2]. It has six major genotypes and eleven subtypes. The genome is composed of about 9600 nucleotides and it contains a single open reading frame encoding a polyprotein of approximately 3010 amino acid residues. Three structural proteins (C, E1 and E2) and six nonstructural proteins (NS2, NS3, NS4A, NS4B, NS5A and NS5B) are released from the polyprotein by host and viral proteolytic enzymes. It has been shown that NS2, NS3 and NS5B are essential enzymes for the replication of HCV. NS2 and the N-terminal domain of NS3 have proteolytic activity that is responsible for intramolecular and intermolecular cleavage of the polyprotein. The C-terminal region of NS3 is an RNA-helicase domain. The three-dimensional (3D) structures of the NS3 protease [3,4] and helicase [5,6] domains have been determined. NS5B is an RNA-dependent RNA polymerase [7,8]. Lohmann *et al.* demonstrated that NS5B replicates the full minus-strand RNA by a copy-back priming mechanism in the absence of additional viral or cellular factors [9–11]. Recently Oh *et al.* reported that the NS5B protein possesses the ability to copy the full-length HCV RNA using both the genomic and antisense strand as the template without any primer or additional factor [12].

Among RNA-dependent RNA polymerases (RdRps), the crystal structure of poliovirus RdRp has been reported by Hansen *et al.* [13]. They elucidated the structure of this

RdRp extensively, but unfortunately one third of the crystal structure was disordered; this region includes the fingers region. Poliovirus, like HCV, is one of the (+)ssRNA viruses; however, there is less than 15% amino acid sequence identity between the RdRps of these viruses. Thus, it will not be possible to obtain a reliable sequence alignment for structural elucidation by homology modeling, except for those motifs conserved among the polymerases. HCV RdRp is composed of 591 amino acid residues and is more than 100 residues longer than poliovirus RdRp (461 residues); the additional residues are found at the C terminus. Determination of the three-dimensional structure of HCV RdRp would be highly desirable as a basis for the design of inhibitors to use in anti-HCV therapy.

Here we report on the three-dimensional structure of HCV RdRp determined by X-ray crystallography using multiple isomorphous replacement with anomalous scattering (MIRAS) phasing, and focus on the structural basis for the RNA selectivity of the template–primer duplex and the substrate.

Results and discussion

Crystallization and structural determination of HCV RdRp

HCV RdRp was overexpressed and purified as a soluble protein by truncation of the highly hydrophobic C-terminal region, where the putative membrane-anchoring region is located [14–16]. HCV RdRp was crystallized under low-salt conditions, and the crystals were found to belong to space group $P4_32_12$ with $a = b = 63.7$ Å, $c = 262.9$ Å, and Matthews' value (V_M) = 2.04 [17]. The crystal structure was determined with MIRAS phasing guided by the 11 selenomethionine (SeMet) positions ($> 3\sigma$) out of 12 methionine (Met) sites in the isomorphous SeMet crystal (Figure 1, Table 1). The current refined model of HCV RdRp was constructed with 553 out of 570 residues, including 284 water molecules (R value = 0.223 and R_{free} = 0.316) using up to 2.5 Å reflection data [18]. There were no residues in the disallowed region of the Ramachandran plot (Table 2) [19].

Overall structure of HCV RdRp

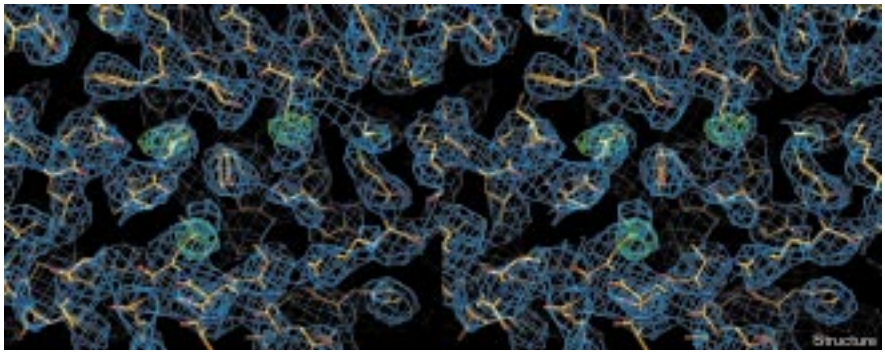
HCV RdRp is a compact globular protein with dimensions of $67 \times 63 \times 68$ Å without any fidelity domains. It can be divided into three domains (i.e., the fingers, palm, and thumb domains) on the basis of its resemblance to a right hand, like other known polymerases including Taq DNA-dependent DNA polymerase I (Taq DNApol I) [20], DNA-dependent DNA polymerase α RB69 [21], human DNA-dependent DNA polymerase β [22], T7 DNA-dependent RNA polymerase [23], human immunodeficiency virus reverse transcriptase (HIV RT) [24], and poliovirus RdRp [13] (Figure 2). The fingers and thumb domains are better constructed than those of the other polymerases. Furthermore, a connection between the fingers and thumb domains is evident.

Fingers domain

The fingers domain of HCV RdRp is divided into two subdomains. One is a β-strand-rich subdomain (β fingers) and the other is an α-helix-rich subdomain (α fingers) (Figure 2). The β-fingers subdomain is composed of four β strands and an α helix, and the α-fingers subdomain has seven α helices.

In the β-fingers subdomain, an antiparallel β sheet (β1 and β2) is supported from behind by α1, β4, β5, and a pseudo β strand of the N-terminal region. Two long loops, Ala9–Thr41 and Asn142–Ala157, extend from the β fingers to the thumb domain. The end of the former loop is a two-turn helix, which is locked into a dimple at the top corner of the thumb domain. The Ala9–Thr41 loop is wrapped upward by the Asn142–Ala157 loop. At the bottom of the latter loop, there is a hole leading to the putative template–primer-binding cleft, which is the putative entry path for incoming ribonucleoside triphosphate (rNTP) substrate. These long loops and the resulting hole seem to be stable because the end of the Ala9–Thr41 loop is locked in the thumb domain and because the loops support each other. The β-fingers surface facing the hole has an electrostatically positive belt created by a line of basic residues (Lys51, Arg48, Arg158, Lys155 and Lys141)

Figure 1



Electron-density map calculated using the MIRAS phase of SHARP [41] after DM [39] density modification (blue) and SeMet-difference Fourier map of the current model between λ = 1.0400 Å and λ = 0.9797 Å (cyan).

Table 1

Summary of the crystal structure determination of HCV RdRp.

	Native	$UO_2(OCOCH_3)_2$		K_2OSO_4		SeMet	
	KEK-PF(BL6B)	Raxis IIc	KEK-PF(BL6B)	Raxis IIc	KEK-PF(BL6B)	SPring-8(BL45XU)	
Statistics of reflection data							
Resolution (Å)	40–2.5	40–3.3	40–2.2	40–2.8	40–2.5	20–2.5	20–2.5
Wavelength (Å)	1.0000	1.5418	1.0000	1.5418	1.0000	1.0400	0.9797
Reflections							
measured	64,714	27,808	236,080	112,365	70,943	47,350	39,535
unique	19,281	7,786	28,544	14,128	17,672	17,385	16,953
R_{merge}* (%)							
all/shell	4.6/6.4	8.0/14.0	11.0/30.8	9.1/23.9	9.5/20.5	8.5/25.5	10.1/25.9
Completeness(%)							
all/shell	98.2/97.9	88.9/81.6	94.3/54.7	98.9/99.1	88.4/97.3	90.9/87.3	88.4/84.7
R_{iso}† (%)	-	22.5	20.1	11.2	12.8	–	13.6‡
Statistics of phase calculation							
Resolution (Å)		40–3.3	40–3.1	40–3.1	40–3.1	–	–
R_{cullis}§							
centric		0.592	0.651	0.700	0.700		
acentric		0.641	0.681	0.639	0.640		
Mean figure of merit							
MIRAS phasing (centric/acentric)		0.628/0.547 (40–2.5 Å)					
after DM		0.725 (40–2.5 Å)					

*$R_{merge} = \Sigma|I-<I>|/\Sigma|I|$. †$R_{iso} = \Sigma(|I_P-I_{PH}|)/\Sigma I_P$. ‡$R_{iso}$ between $SeMet_{\lambda = 1.0400Å}$ and $SeMet_{\lambda = 0.9797Å}$.
§$R_{cullis} = \Sigma||F_{PH}(obs) +/- F_P(obs)|-F(calc)|/\Sigma|F_{PH}(obs) +/- F_P(obs)|$

extending from the outside to the inside of the protein (Figures 3a,4b). This is a suitable structure to accept highly negatively charged triphosphate of rNTP.

The α-fingers subdomain is located at the exit for the RNA duplex. The long loop connecting αA and αB traverses the HCV RdRp molecule, while the region formed by αC, αD, and αE overhangs the palm domain, creating a concave wall (Figures 2,3a). Between the α and β fingers there is a U-shaped valley. The concave wall and the U-shaped valley are electrostatically positive as a result of a line of basic residues (Lys90, Lys98, Lys106, Arg109, Arg168 and Lys172) (Figures 3a,4b).

Surprisingly, structural comparison of polymerases has shown that the β-fingers subdomain of HCV RdRp is similar to the fingers domain of HIV RT [24], except for the highly conserved palm domain; the folding of the fingers and thumb domains has hitherto been quite distinct in different classes of polymerases [20–24], except for the thumb domains of HIV RT as described by Hansen *et al.* in comparison with poliovirus RdRp [13]. When the β fingers and palm of HCV RdRp are superimposed on HIV RT the β fingers show a similar structure to the fingers of HIV RT, despite having no detectable sequence homology [24]. However, no structure corresponding to the HCV α fingers is present in HIV RT, except for a long, extended strand corresponding to the αF helix (Figure 3a). In the β1–β2 region, Lys155 and Arg158, of HCV RdRp correspond geometrically to Lys70 and Arg72 of HIV RT, respectively. Arg158 is presumed to have an important role in rNTP binding, as the guanidino group of the corresponding arginine forms a salt bridge with the triphosphate group of the substrate in the ternary complex of HIV RT [24].

On the other hand, the folding motif of the α-fingers region (residues 74–189) resembles the α-helical regions of the DNA-binding domain of Mrf-2, which binds to duplex DNA (Figure 3c) [25]. A similar folding motif has been reported in the fingers domains of DNApol I and repair enzymes [25], but the DNA-binding mode of DNApol I is distinct from that of the Mrf-2 DNA-binding domain [20]. When the HIV RT ternary complex is superimposed over HCV RdRp, the superimposed DNA duplex of the HIV RT is located in a similar orientation to the Mrf-2-bound DNA. The fingers domain of poliovirus

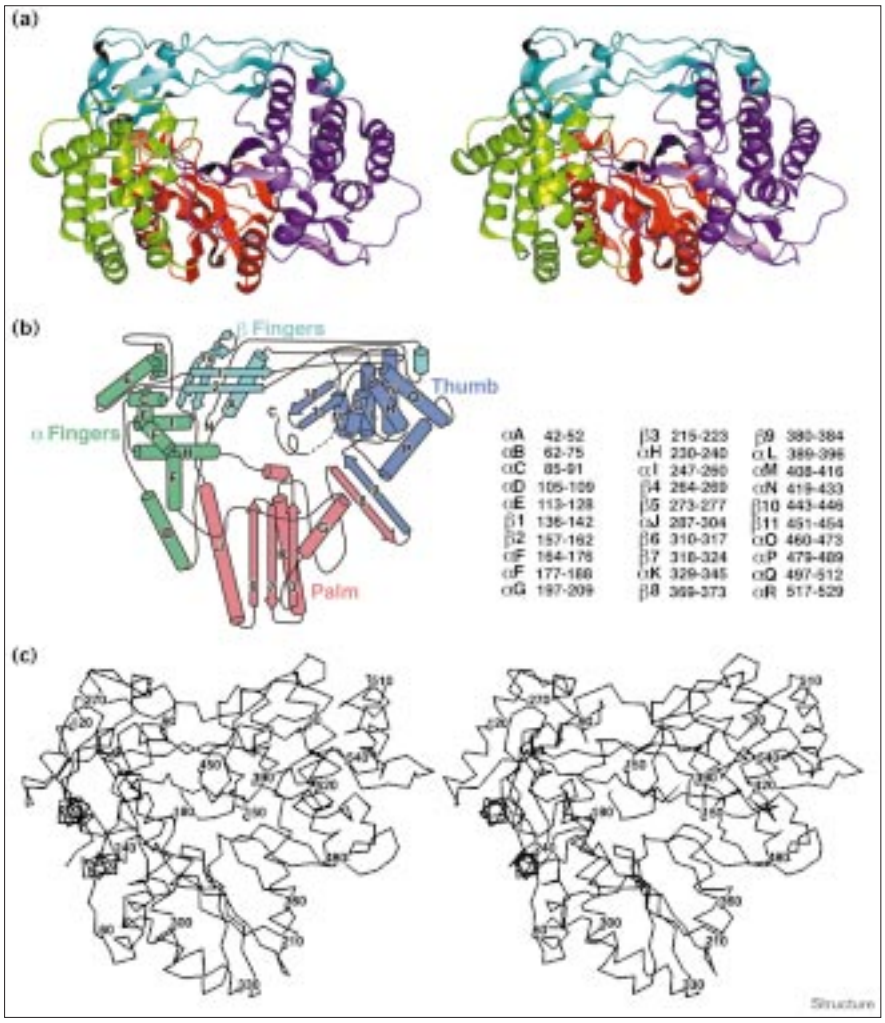
Table 2

Refinement statistics.

Resolution (Å)	20.–2.5
Number of reflections ($\|F\| > 2.0\sigma$)	19,063
Number of nonhydrogen atoms	4580
Number of solvent atoms	284
R_{value}* (%)	22.3
R_{free}* (%) (10% free set)	31.6
Rmsd	
in bond lengths (Å)	0.01
in bond angles (°)	1.6

*$R_{value} = \Sigma|F(obs)-F(calc)|/\Sigma F(obs)$.

Figure 2



Crystal structure of HCV RNA-dependent RNA polymerase (RdRp). (a) Stereoview of a cartoon model of HCV RdRp. The α fingers, β fingers, palm, and thumb domains are colored green, cyan, red, and violet, respectively. (b) Diagram of HCV RNA-dependent RNA polymerase. The α helices and β sheets are indicated sequentially in alphabetical and numerical order. The amino acid range of each secondary structure is listed in the right panel. (c) Stereoview of the α carbon trace of HCV RdRp.

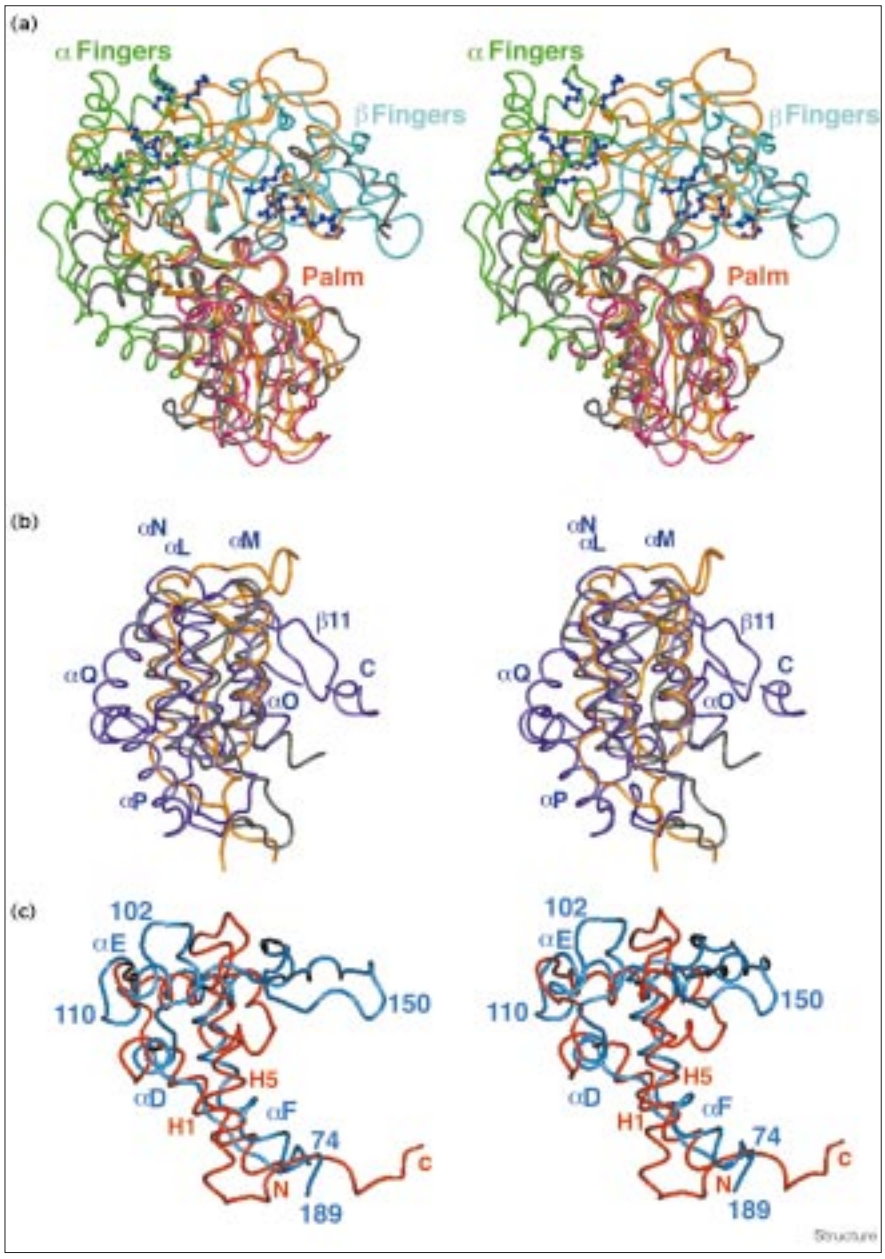
RdRp [13] is superimposable on HCV RdRp with respect to the regions corresponding to the αF, αH and αI helices of the base of the α fingers and the thumb-linking region of the β fingers, despite the major part of the fingers being undefined (Figure 3a).

Palm domain

The palm is the catalytic domain and contains a folding motif that is highly conserved among polymerases [13,20–24] and other nucleotide-utilizing enzymes, including adenylyl cyclase [26] (Figure 2). The palm is composed of a three-stranded antiparallel β sheet (β6, β7, and β3), a small helix following β3, and two supporting α helices (αJ and αK), with an additional αG helix at the thumb interface. On the basis of the crystal structure of poliovirus RdRp by Hansen *et al.* [13], motifs A–E have been assigned to the HCV RdRp palm structure to enable comparison with other polymerases. The A, B, and C motifs are highly conserved in RNA-dependent polymerases [13], and correspond to regions β3 (residues 213–228), αJ (residues 282–302) and β2/β3 (residues 310–324) in HCV RdRp, respectively.

These polymerases, including HCV RdRp, require divalent metal cations Mg^{2+} or Mn^{2+} for the ligation of rNTP [7,8]. The essential residues for metal binding [9] are Asp220 and the carbonyl of the Thr221 peptide backbone on β3 in motif A, as well as Asp318 and Asp319 on β7 in motif C (Figures 4a,c). Motif A has been proposed to be a substrate-selectivity region; especially, Asp225, of which the corresponding residue in poliovirus RdRp is thought to be primary substrate discriminator for rNTP over deoxyribonucleoside triphosphate (dNTP) [13]. Asp225 and Asp352 also contribute to the electrostatically negative surface in the vicinity of the three essential aspartates (Figure 4b). The negative surface constitutes the floor of the putative rNTP-entry path; its roof is positive, as described above.

Figure 3



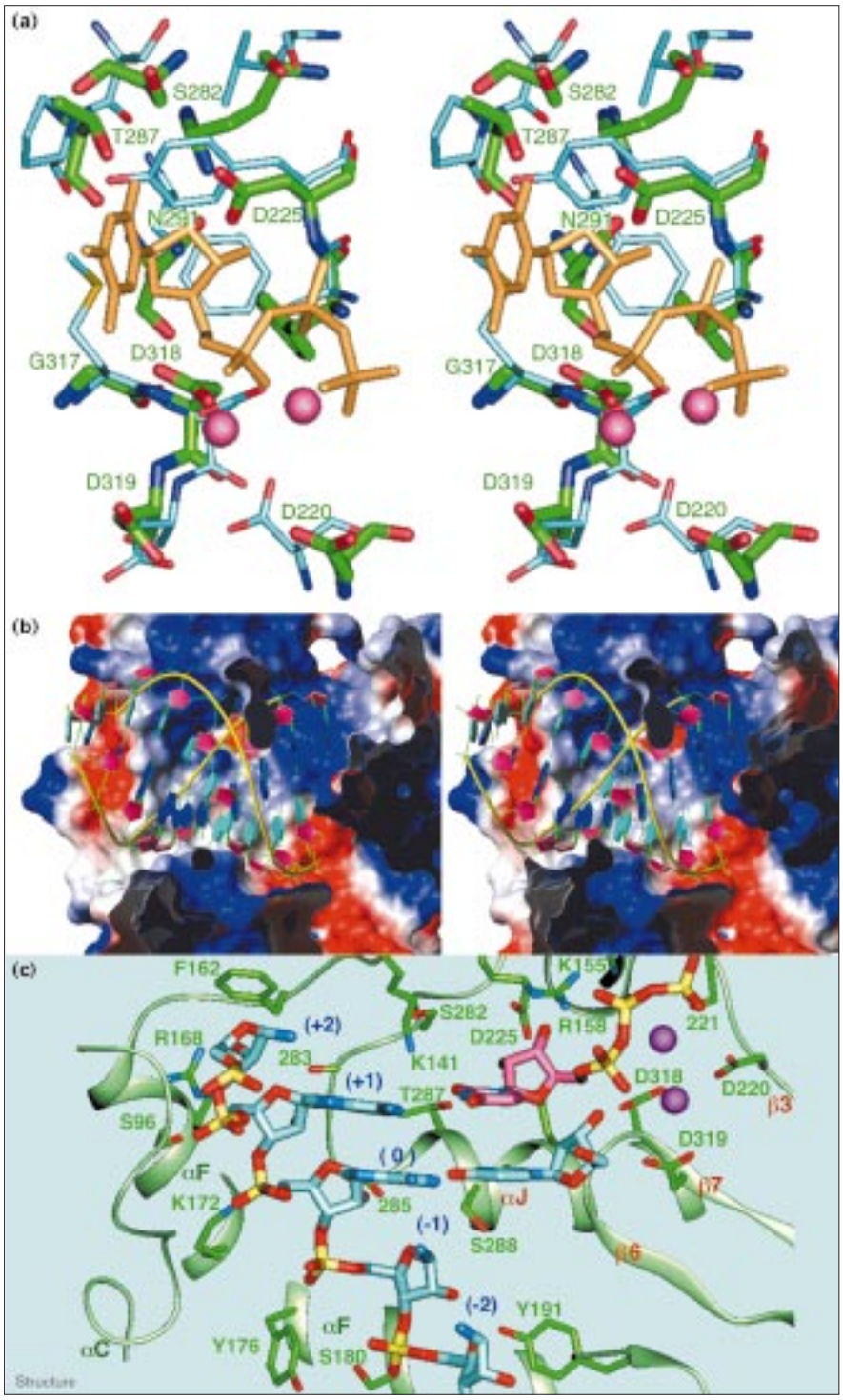
Structure comparison of HCV RdRp with poliovirus RdRp and HIV RT and with the Mrf-2 DNA-binding domain. (a) Stereoview of superimposed wire models of the α- and β-fingers subdomains and palm domain of HCV RdRp and the corresponding portions of poliovirus RdRp [13] and the ternary complex of HIV RT [24]. The α fingers, β fingers, and palm of HCV RdRp are shown in green, cyan, and pink, respectively. Poliovirus RdRp is shown in silver and HIV RT is shown in gold. The positively charged residues mentioned in the text are represented by blue balls and sticks. In the β-fingers subdomain, the basic residues are Lys51, Arg48, Arg158, Lys155 and Lys141 from the outside. In the α fingers, the residues are Lys98, Arg168, Lys172, Lys90, Arg109 and Lys106 from the top right in a clockwise direction. (b) Stereoview of superimposed wire models of the thumb domains of HCV RdRp (violet), poliovirus RdRp (silver) [13] and HIV RT (gold) [24]. (c) Stereoview of superimposed wire models of the α-fingers region, including its connecting β1–β2 loop in the β-fingers subdomain (residues 74–189), of HCV RdRp (blue) and the Mrf-2 DNA-binding domain (red) [25].

Thumb domain

The thumb domain of HCV RdRp is predominantly α-helical, being composed of seven α helices and a two-stranded antiparallel β sheet. This well-constructed thumb is another notable feature of HCV RdRp when it is compared with various polymerases (Figure 3b). Only β9, αL, αM, and αN correspond with and are superimposable on both HIV RT [24] and poliovirus RdRp [13], despite αL corresponding to an extended strand and an upside-down αN corresponding to a helix in HIV RT [24] (Figure 3b). Three α helices (αL, αM, and αN) form the core of the thumb. The other α helices (αO, αP, αQ and αR) following β11 surround and support the core helices (Figure 3b). Structural comparison suggests that the αM helix interacts with the minor groove of the template–primer duplex [24] (Figure 4b).

The extraordinarily long β sheet of β10 and β11 protrudes down into the putative RNA-duplex-binding cleft from the top of the thumb in HCV RdRp, although all other known polymerases including poliovirus RdRp have a short loop structure connecting two helices [13,20–24]. The β sheet might have a role in maintaining the stability of binding to the RNA duplex. The long β sheet might act

Figure 4



The selectivity of substrate and template–primer duplex in HCV RdRp. (a) Stick models of the substrate-binding sites of HCV RdRp and the ternary complex of HIV RT with deoxyribothymidine triphosphate (dTTP) and two Mg^{2+} ions bound [24]. Carbon atoms of HCV RdRp and HIV RT are shown in green and cyan, respectively, while nitrogen is in blue and oxygen is in red. The dTTP molecule and Mg^{2+} ions are gold and pink, respectively. (b) Electrostatic surface of HCV RdRp complexed with an A-form-RNA duplex. The model was based on the superimposed structures of HCV RdRp and the HIV RT complexed with the DNA template–primer and dTTP [24] (see the Materials and methods section; Figures 3a,b). Electrostatically positive and negative areas are shown in blue and red, respectively. The phosphodiester backbones are shown in yellow, while the ribose of the RNA duplex is pink and the bases of the template and primer are blue and cyan, respectively. This figure was prepared using MOLMOL [45]. (c) Close-up view of the cartoon model of HCV RdRp with the A-form-RNA duplex and the substrate uridine triphosphate (UTP) with two Mn^{2+} ions (violet) shown in (b). The model in (c) was rotated about 90° in anticlockwise direction relative to (b). Carbon atoms of the RNA duplex are shown in cyan, the substrate is in pink and RdRp is in green, while nitrogen and oxygen atoms are in blue and red. The number with and without residue type indicates the sidechain of the residue and the carbonyl of the backbone in HCV RdRp, respectively. The template RNA residue numbers from zero for the priming template are in parentheses (blue), and the bases at (0) and (+1) are presented. The labels for secondary structure are the same as in Figure 2. The UTP was modeled based on the HCV RdRp complex structure with UTP and Mn^{2+}.

as a barrier against the binding and release of the working RNA duplex, as the sheet fills the RNA-binding cleft (Figure 2). In another functional possibility, the β sheet might be responsible for the sequence-specific recognition for initiation from the 3′(X) sequence of HCV genomic RNA [10,12], as in the case of the specificity loop of T7 RNA polymerase [23].

The definitive C-terminal segment includes residues Leu547–Ser556, which follow a short disordered segment,

and is located between the palm domain and the tip of the β sheet. The C-terminal region between Thr532 and Arg570 might be a regulator of HCV RdRp activity before the membrane-anchor region of the C terminus (residues 571–591), as the structure of the region seems to be auto-inhibitory. In this region there are two casein kinase-2 (CK2) phosphorylation motifs at Ser543 and Ser556 [27]. These sites might coincide with phosphorylation sites for the phosphoprotein in cells [14], as the phosphorylation sites are associated with high temperature factors in the solvent region. Phosphorylation of these sites should shift the equilibrium to expose the region in the solvent by increasing its hydrophilicity, and this would make HCV RdRp more active. This postulate is supported by the report that further truncated mutants of the C-terminal region were more active than the similar construct used in this study [16].

Substrate selectivity

HCV RdRp strictly discriminates between RNA and DNA for the template as well as between rNTP and dNTP for the substrate [7,8,15]. Although other classes of polymerase can sometimes utilize both types of template [28,29], efficient elongation of the primer by HCV RdRp has only been observed with the combination of an RNA template and rNTP substrate. Unlike in other polymerases, primer elongation proceeds without frequent abortive reactions [7–11,28,29]. HCV RdRp tolerates more variation in the nucleotides of the primer than those of the template and substrate, but the incorporation rate is significantly reduced by using a DNA primer [8,15]. The structural basis of this selectivity is discussed below.

The primary substrate-discriminating residue was proposed to be Asp238 of poliovirus RdRp, and its corresponding residue is conserved among (+)ssRNA viral RdRps based on the structural alignment and mutation studies on the corresponding Phe155 residue of Moloney murine leukemia virus (MMLV) RT by Hansen *et al.* [13,30]. Superimposition of the HCV RdRp active site and the ternary complex of the HIV RT active site shows that the corresponding residue Asp225 of HCV RdRP is coincident with Tyr115 of HIV RT (Figure 4a), indicating that the hydrogen bond between Asp225 and the 2′-OH of rNTP can be formed as proposed [13]. In a site-directed mutagenesis study of MMLV RT, Gao *et al.* demonstrated that the corresponding aromatic residue Phe155 is involved in the discrimination of dNTP from rNTP, partly by steric hindrance of the 2′-hydroxyl of rNTP [30]. Furthermore, HCV RdRp possesses a shallow hydrophilic pocket with bound water molecules around the corresponding residue, whereas the corresponding area is a hydrophobic plane in RTs [23,30] (Figure 4a). The pocket is composed of residues that are highly conserved among (+)ssRNA RdRps (Asp225 in motif A and Ser282, Thr287, and Asn291 in motif B). Although the residues corresponding to Ser282 and Thr287 of HCV RdRp are also hydrophilic in both RTs [24,30], the bulky hydrophobic sidechains of Tyr115 and Met184 cover them in HIV RT [13] (Figure 4a). The hydrophilic pocket of HCV RdRp should better accommodate the more polar and bulkier ribose of rNTPs without dehydration. Coupled with the enthalpic gain from forming a hydrogen bond with Asp225, this should contribute to substrate selectivity.

Template–primer-binding selectivity

The α fingers of HCV RdRp represent the most prominent difference when compared with polymerases of other classes [20–24] and may be a structural discriminator that recognizes the A-form duplex of the template and primer during replication. Superimposition of HCV RdRp and the ternary complex of HIV RT [24] (Figures 3a,b) showed that the A-form portion of the DNA duplex of the HIV RT complex can also be accommodated in the HCV RdRp cleft without major collision, except for the dangling β10/β11 sheet and the C-terminal segment. On the basis of the superimposition the A-form RNA duplex was docked into the cleft of HCV RdRp after simply removing the β sheet and the C-terminal segment for further inspection of RNA discrimination, although some minor collisions remained. According to the model of the HCV RdRp complex with RNA duplex, the phosphodiester backbone of the RNA template lies in the vicinity of the electrostatically positive concave surface on the α fingers, and the overhanging structure of αC and the following loop might well hold the backbone in position (Figures 4b,c). At the interface of the α-fingers subdomain and the palm domain, a long shallow trench composed of the αF helix and the N-terminal edge of the αJ helix fits the phosphodiester backbone and ribose moieties as the guide of the template RNA (Figures 4b,c). Arg168 and Lys172 of the αF helix are at the fingers side of the trench and coincide with Arg78 and Lys82 in the HIV RT complex, which are next to the phosphodiester backbone of the template [24]. These positively charged sidechains should be good acceptors for the negative charge of the template backbone. The bent αF helix located next to the major groove of RNA duplex should interact directly with the phosphodiester backbone of the RNA template (Figures 4b,c). On the other side of the guide trench, the carbonyl oxygen atoms of the peptide backbone of Gly283 and Leu285 are located within hydrogen-bonding distance of the 2′-OH of the template nucleotides for substrate rNTP (+1) and the priming nucleotide (0), respectively (Figure 4c). The 2′-OH of the next template nucleotide (–1) can form a hydrogen bond with the OH of Ser288 as well as the phenolic OH of Tyr191. Surprisingly, Gly283, Leu285, and Ser288 are involved in substrate selectivity as well, and the motif-B region is highly conserved among RNA-dependent polymerases including RTs [13, 24, 30]. The results of the electrophoretic mobility-shift experiment of HCV RdRp fragments with the 3′ end of HCV genomic

Figure 5

The structural and multiple sequence alignment of RNA-dependent polymerases. (**a**) Structural sequence alignment of the fingers and palm domains of HCV RdRp, poliovirus RdRp [13], and HIV RT [24]. The sequence numbers of HCV RdRp are colored cyan (α fingers), green (β fingers), and red (palms). The residues of the α helices, β sheets, and loop regions are shown in pink, green and black, respectively. Undefined residues of poliovirus RdRp are shown as underlined lower-case letters. The number at the top line shows the digit of the HCV RdRp sequence number colored by the domain, as in Figure 2. (**b**) Multisequence alignment of (+)ssRNA virus RdRps from deng virus (strain den2p), bovine diarrhoea virus (bdvdn), polio virus (pol2l), tobacco mosaic virus (tmvto) and human rhino virus (hrv14) relative to HCV (hcv1b) using PSI-BLAST [33]. The basic residues are shown in blue.

RNA was also compatible with the model, as the fragment containing the α-fingers region (residues 1–194) and residues 196–298 containing motif B were shown to specifically bind the RNA [31].

The RNA primer backbone comes close to the other side of the α fingers in the model of HCV RdRp complexed with RNA duplex (Figure 4b). On the side a line of basic residues (Lys90, Lys106, and Arg109) make the surface electrostatically positive, and this surface faces the phosphodiester backbone of the RNA primer as it forms part of the A-form duplex. Additionally, the change in specificity for B-form to A-form DNA that in DNApol I or HIV RT is thought to be mediated by closing of the fingers domain during template–primer binding [20,24] is not possible in RdRp, as the unified connection of the fingers and thumb domains would prohibit such large domain movement as is described above. Accordingly, the α fingers of HCV RdRp might be the first structural motif to hold the primer strand directly.

The α fingers might represent a common structure that discriminates RNA duplex from DNA strictly in RdRps from (+)ssRNA viruses [32], as the defined part of the corresponding α fingers has the same folding motif in poliovirus RdRp [13] (Figures 3a,5a). When multisequence alignment of RdRps was performed using PSI-BLAST [33], the basic residues corresponding to Lys106 and Arg109 in the α fingers were significantly conserved in several (+)ssRNA viruses, including deng, bovine diarrhoea, polio, tobacco mosaic and human rhino viruses. Also, the sequences corresponding to β1 and β2 in the β-fingers domain were conserved among these RdRps (Figure 5b). Thus, the sequence alignment data also suggested that the α fingers represent a common structural discriminator, which distinguishes RNA duplex from DNA in (+)ssRNA virus RdRps.

Biological implications

The three-dimensional structure of hepatitis C virus RNA-dependent RNA polymerase (HCV RdRp) reveals the characteristic features of an RdRp, but the well-constructed fingers and thumb domains are unique among known polymerases [13,20–24]. This feature is consistent with the run-off replication of the (+)ssRNA genome without frequent abortive polymerization [9–12], because it is presumed to allow secure gripping of the template–primer RNA by the domains during elongation of the primer. Another unique feature of (+)ssRNA viral RdRps, the tight connection of the β-fingers subdomain and thumb domain, should reduce domain movement. It is not plausible that the open–close motion of the fingers domain occurs during template–primer binding as it does with HIV RT and Taq DNApol I [20,24]. This is a rational structure for a (+)ssRNA virus because RdRp replicates from the 3′ end of the single-stranded genome RNA without the need for a primer [7–12].

The catalytic aspartates (Asp220, Asp318 and Asp319) of the active site have the same geometry as those of other polymerases. Also, the putative binding site of ribose in rNTP, which is composed of residues conserved among (+)ssRNA viruses, including Asp225, Ser282, Thr287, and Asn291 [32], is a shallow hydrophilic cavity that has a preference for rNTP as the substrate. Asp225 is correctly positioned to form a hydrogen bond with the 2′-OH of rNTP, as proposed previously [13].

The fingers domain is composed of two subdomains (α and β fingers). The α fingers are proposed to be involved in the discrimination of RNA from DNA on the basis of studies of the docking of RNA and DNA duplex onto HCV RdRp. The folding of the α fingers is similar to that of the Mrf-2 DNA-binding domain [25], and a fragment containing the α-fingers region showed specific 3′-end HCV-RNA-binding affinity [31]. Study of the model of HCV RdRp complexed with RNA indicated that motif B might be another RNA-template discriminator as well as having a substrate-selectivity function; motif B is broadly conserved among RNA-dependent polymerases [13]. The glycine residue most conserved among motif-B regions of these polymerases is crucial for the structural integrity of the proteins. The fragment containing motif B also showed RNA-binding capabilities [31].

The β-fingers subdomain has a similar structure to the fingers domain of HIV RT [24], an RNA-dependent DNA polymerase. The structural similarity of the β-fingers subdomain and the fingers domain suggests an evolutionary relationship between HCV RdRps and HIV RT.

The well-constructed structure of the thumb domain is unique to HCV RdRp, and its function remains to be elucidated. The buried C-terminal region in the putative template–primer-duplex-binding cleft might have an auto-inhibitory function that modulates HCV RdRp activity *in vivo*, and the long β strand might exert a certain regulatory effect: for example, on initiation recognition during replication.

Materials and methods

Expression and purification of the soluble form of RdRp will be described elsewhere. The soluble form of RdRp (residues 1–570), with an additional histidine tag from HCV strain 1b(BK), and deletion of a 21 amino acid sequence from the hydrophobic C-terminal region, was constructed and expressed in *E. coli* BL21(DE3) (Novagen) using the T7 system [34]. SeMet protein was prepared using *Escherichia coli* B834 (DE3) (Novagen) in a modified medium [34,35] (D Irikura and Y Urade, personal communication). Proteins were purified using Ni-NTA agarose (Qiagen), Mono-S, and Sephacryl S-200 (Pharmacia) chromatography. The enzymes were shown to be monodispersive in a low ionic buffer by the dynamic light-scatter method (Protein Solutions). Purified RdRp catalyzes the elongation of nucleotides using poly[A]/oligo[U] or HCV(1b) 3′(X). Purified native and SeMet RdRps were, respectively, *Mr* = 64,160 and 64,718 by LC/ES-MS. The N-terminal sequence was verified to be SMSYTWTG by Edman degradation, indicating that the N-terminal Met residue was hydrolyzed during the expression and purification procedure.

Purified samples were concentrated to 10 mg/ml by ultrafiltration in a solution containing 5 mM dithiothreitol. Crystals of crystallographic quality were grown by the hanging-drop vapor-diffusion method, at 22.5°C using a reservoir solution containing 21–28%(w/v) polyethylene (PEG) 4000, 0.2–0.35 M ammonium acetate, 0.1 M sodium acetate, and 0.02 M TES (pH 6.0–7.5). Uranium and osmate derivatives were found to be useful for MIRAS phasing. After dipping in a reservoir solution containing 20% (w/v) trehalose as a cryoprotectant, diffraction data were collected at 100K using in-house and synchrotron facilities (KEK-PF BL6B [36] and SPring-8 BL45XU [37]). The image data were processed using DENZO [38] (HKL Inc.) and SCALA program of CCP4 suite [39]. The statistics of the diffraction data are summarized in Table 1.

The original heavy-atom sites of the uranium and osmate derivatives were determined by Shelxs [40], while MLPHARE [39] and SHARP [41] were used for heavy-atom sites refinement and MIRAS phase calculation. Subsequent density modification and phase extension up to 2.5 Å was applied using DM [39], and the Fourier map using the DM modified phase was of high enough quality to allow atomic model building with O [42] (Figure 1). In the *f*′ different Fourier map of SeMet data between λ = 1.0400 Å and λ = 0.9797 Å, 11 outstanding peaks (> 3σ) out of 12 SeMet residues were found and used as the guide and confirmation for model building, as the native and SeMet crystals were isomorphous. Iterative refinement and model rebuilding were subsequently performed using X-PLOR 98.1 [18] and O [42]. There were no residues in the disallowed region of the Ramachandran plot by PROCHECK [19]. A summary of the refinement statistics is presented in Table 2.

After a one-day soaking in the solution containing UTP (10 mM) and Mn^{2+} (5 mM), the diffraction data were collected using a Quantum 4 charged-couple device (CCD) detector at KEK-PF BL6A, and processed using Dps [43] and Mosfilm [44]. The R_{merge} was 12.2% up to 2.6 Å. The current model includes the triphosphate moiety of UTP and two Mn^{2+} ions (R_{vale} = 24.5% and R_{free} = 35.5% using 12,634 [> 2σ] reflections of 20–2.7 Å). The docking model of HCV RdRp with A-form RNA duplex was constructed using the complex structure with Insight II and Quanta (MSI) based on the superimposed structure of the ternary complex HIV RT structure [24].

Accession numbers

The atomic coordinates of HCV RdRp have been deposited in RCSB Protein Data Bank with the accession code 1quv.

Acknowledgements

We are grateful for the encouragement and support of T Miyamoto, R Oku, and T Matsumoto, as well as for the technical assistance of E Inagaki, E Nagao, K Okuda, and H Kuno at the Institute. We also thank N Sakabe, N Watanabe, M Suzuki and C Neilsen at KEK-PF and T Kumasaka at SPring-8 for assistance with data collection at these SR facilities, as well as D Irikura and Y Urade at OBI for assistance with SeMet-protein expression. HA and MM are members of the TARA Sakabe project. Finally, we are thankful for the constructive comments of a referee.

References

1. WHO (1999). Global surveillance and control of hepatitis C. *J. Viral Hepatitis* 6, 35-47.
2. Bartenschlager, R. (1997). Candidate targets for hepatitis C virus-specific antiviral therapy. *Intervirology* 40, 378-393.
3. Love, R.A., *et al.,* & Hostomska, Z. (1996). The crystal structure of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site. *Cell* 18, 331-342.
4. Kim, J.L., *et al.,* & Thomson, J.A. (1996). Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. *Cell* 18, 343-355.
5. Yao, N., *et al.*, & Weber, P.C. (1997). Structure of the hepatitis C virus RNA helicase domain. *Nat. Struct. Biol.* 6, 463-467.
6. Kim, J.L., *et al.*, & Caron, P.R. (1998). Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide. *Structure* 6, 89-100.
7. Behrens, S.E., Tomei, L. & De Francesco, R. (1996). Identification and properties of the RNA-dependent RNA polymerase of hepatitis C virus. *EMBO J.* 15, 12-22.
8. De Francesco, R., *et al.*, & Jiricny, J. (1996). RNA-dependent RNA polymerase of hepatitis C virus. *Methods Enzymol.* 275, 58-67.
9. Lohmann, V., Körner, F., Herian, U. & Bartenschlager, R. (1997). Biochemical properties of hepatitis C virus NS5B RNA-dependent RNA polymerase and identification of amino acid sequence motifs essential for enzymatic activity. *J. Virol.* 71, 8416-8428.
10. Lohmann, V., Roos, A., Körner, F., Koch, J.O. & Bartenschlager, R. (1998). Biochemical and kinetic analyses of NS5B RNA-dependent RNA polymerase of the hepatitis C virus. *Virology* 249, 108-118.

11. Lohmann, V., Overton, H. & Bartenschlager, R. (1999). Selective stimulation of hepatitis C virus and pestivirus NS5B RNA polymerase activity by GTP. *J. Biol. Chem.* 274, 10807-10815.
12. Oh, J-W., Ito, T. & Lai, M.C. (1999). A recombinant hepatitis C virus RNA-dependent RNA polymerase capable of copying the full-length viral RNA. *J. Virol.* 73, 7694-7702.
13. Hansen, J.L.H., Long, A.M. & Schultz, S.C. (1997). Structure of the RNA-dependent RNA polymerase of poliovirus. *Structure* 5, 1109-1122.
14. Hwang, S.B., Park, K-J., Kim, Y-S., Sung, Y.C. & Lai, M.M.C. (1997). Hepatitis C virus NS5B protein is a membrane-associated phosphoprotein with a predominantly perinuclear localization. *Virology* 227, 439-446.
15. Yamashita, T., *et al.,* & Murakami, 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.
16. Ferrari, E., *et al*., & Hong, Z. (1999). Characterization of soluble hepatitis C virus RNA-dependent RNA polymerase in *Escherichia coli*. *J. Virol.* 73, 1649-1654.
17. Matthews, B.W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* 33, 491-497.
18. Brünger, A.T. (1992). X-PLOR version 3.1: a system for X-ray crystallography and NMR. Yale University Press, New Haven, CT.
19. Laskowski, R.J., Macarthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26, 283-290.
20. Li, Y., Korolev, S. & Waksman, G. (1998). Crystal structure of open and closed form of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I. *EMBO J.* 17, 7514-7525.
21. Wang, J., Sattar, A.K.M., Wang, C.C., Karam, J.D., Konigsberg, W.H. & Steitz, T.A. (1997). Crystal structure of α pol, a family replication DNA polymerase from bacteriophage RB69. *Cell* 89, 1087-1099.
22. Sawaya, M.R., Prasad, R., Wilson, S.H., Kraut, J. & Pelletier, H. (1997). Crystal structure of human DNA polymerase β complexed with gapped and nicked DNA. *Biochemistry* 36, 11205-11215.
23. Cheetham, G.M.T., Jeruzalmi, D. & Steitz, T.A. (1999). Structure basis for initiation of transcription from an RNA polymerase-promoter complex. *Nature* 399, 80-83.
24. Huang, H., Chopra, R., Verdine, G.L. & Harrison, S.C. (1998). Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* 282, 1669-1675.
25. Yuan, Y-C., Whitson, R.H., Liu, Q., Itakura, K. & Chen, Y. (1998). A novel DNA-binding motif shares structural and repair nucleases and polymerases. *Nat. Struct. Biol.* 5, 959-964.
26. Tesmer, J.J.G., Sunahara, R.K., Johnson, R.A., Gossellin, G., Gilman, A.G. & Sprang, S.R. (1999). Two-metal-ion catalysis in adenylyl cyclase. *Science* 285, 756-760.
27. Maggio, F., Marin, O. & Pinna, L.A. (1994). Substrate specificity of protein kinase CK2. *Cell Biol. Res.* 40, 401-409.
28. Arnaud-Barbe, N., C-Sauvion, V., Oriol, G., Mandrand, B. & Mallet, F. (1998). Transcription of RNA templates by T7 RNA polymerase. *Nucleic Acids Res.* 26, 3550-3554.
29. Harris, D., Kaushik, N., Pandey, P.K., Yadav, P.N.S. & Pandey, V.N. (1998). Functional analysis of amino acid residues constituting the dNTP binding pocket of HIV-1 reverse transcriptase. *J. Biol. Chem.* 273, 33624-33634.
30. Gao, G., *et al*., & Goff, S.P. (1997). Conferring RNA polymerase activity to a DNA polymerase: a single residue in reverse transcriptase controls substrate selection. *Proc. Natl Acad. Sci. USA* 94, 407-411.
31. Cheng, J-C., Chang, M-F. & Chang, S.C. (1999). Specific interaction between the hepatitis C virus NS5B RNA polymerase and the 3′ end of the viral RNA. *J. Virol.* 73, 7044-7049.
32. Poch, O., Sauvaget, I., Delarue. M. & Tordo, N. (1989). Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J.* 8, 3867-3874.
33. Altschul, S.F., *et al*., & Lipman, D. (1997). Gapped BLAST and SPI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389-3402.
34. Tabor, S. & Richardson, C.C. (1985). A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc. Natl Acad. Sci. USA* 82, 1074-1078.
35. Doublié S. & Carter, C.W., Jr. (1992). Preparation of selenomethionyl protein crystals. In *Crystallization of Nucleic Acids and Proteins: A Practical Approach*. (Ducruix, A. & Giegé, R., eds), pp. 311-317, Oxford University Press, Oxford, UK.
36. Sakabe, N. (1991). X-ray diffraction data collection system for modern protein crystallography with a Weissenberg camera and an imaging plate using synchrotron radiation. *Nucl. Instrum. Methods A* 303, 448-463.
37. Yamamoto, M., Kumasaka, T., Fujisawa, T. & Ueki, T. (1998). Trichromatic concept at SPring-8 RIKEN beamline I. *J. Synchrotron Rad.* 5, 222-225.
38. Otwinowski, Z. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276, 307-326.
39. Collaborative Computational Project, Number 4. (1994). The CCP4: A suite of programs for protein crystallography. *Acta Crystallogr. D* 50, 760-763.
40. Sheldrick, G.M. (1990). Phase annealing in SHELX-90; direct methods for larger structures. *Acta Crystallogr. A* 46, 467-473.
41. Fortelle, E.L. & Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* 276, 472-494.
42. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* 47, 110-119.
43. Steller, I., Bolotovsky, R. & Rossmann, M. (1997). An algorithm for automatic indexing of oscillation images using Fourier analysis. *J. Appl. Crystallogr.* 30, 1036-1040.
44. Leslie, A.G.W. (1993). Data collection and processing. In *Proceedings of the CCP4 study weekend.* (Sawyer, L., Isaacs, N. & Bailey, S., eds), pp 44-51, SERC Daresbury Laboratory, England
45. Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structure. *J. Mol. Graphics* 14, 51-55.