Cell, Vol. 87, 331–342, October 18, 1996, Copyright ©1996 by Cell Press

The Crystal Structure of Hepatitis C Virus NS3 Proteinase Reveals a Trypsin-like Fold and a Structural Zinc Binding Site

Robert A. Love,* Hans E. Parge,*
John A. Wickersham,* Zdenek Hostomsky,*
Noriyuki Habuka,†‡ Ellen W. Moomaw,*
Tsuyoshi Adachi,†‡ and Zuzana Hostomska*
*Agouron Pharmaceuticals, Inc.
3565 General Atomics Court
San Diego, California 92121
†Japan Tobacco Inc.
Central Pharmaceutical Research Institute
1-1, Murasaki-cho, Takatsuki
Osaka 569
Japan
‡Center for Tsukuba Advanced Research Alliance
University of Tsukuba
Tsukuba 305
Japan

Summary

During replication of hepatitis C virus (HCV), the final steps of polyprotein processing are performed by a viral proteinase located in the N-terminal one-third of nonstructural protein 3. The structure of NS3 proteinase from HCV BK strain was determined by X-ray crystallography at 2.4 Å resolution. NS3P folds as a trypsin-like proteinase with two β barrels and a catalytic triad of His-57, Asp-81, Ser-139. The structure has a substrate-binding site consistent with the cleavage specificity of the enzyme. Novel features include a structural zinc-binding site and a long N-terminus that interacts with neighboring molecules by binding to a hydrophobic surface patch.

Introduction

Hepatitis C virus (HCV) is the major etiologic agent of human parenterally and community-acquired non-A, non-B hepatitis (Choo et al, 1989). Chronic HCV infection is a global disease, and the number of carriers is estimated to be about 300 million. Chronic infection may lead to the development of chronic hepatitis, liver cirrhosis, and hepatocellular carcinoma (reviewed by Houghton, 1996). In Europe and Japan, the disease is more prevalent than either hepatitis B virus or human immunodeficiency virus infections. Protective vaccination is not available for HCV, and current treatments with interferon are successful only in a limited number of patients. Therefore, considerable attention has been focused in recent years on understanding HCV replication and obtaining structural information about essential HCV proteins.

The HCV virion has a positive-strand RNA genome that was cloned in 1989 (Choo et al., 1989). It is composed of about 9,400 nucleotides and contains a single large open reading frame encoding a polyprotein of 3010–3033 amino acid residues (Kato et al., 1990; Choo et al., 1991; Takamizawa et al., 1991). The genetic organization of HCV (reviewed by van Doorn, 1994) is similar to that of flavi- and pestiviruses, and it was classified as a separate genus of the family Flaviviridae. Sequence analysis of HCV isolates reveals that HCV exists in many distinct variants. A total of six major genotypes and at least 11 subtypes have been recognized (Simmonds et al., 1994).

The nonstructural (NS) proteins involved in replication of the HCV genome are released by the action of two proteinases: NS2-3 and NS3. NS2-3 proteinase is a zinc-dependent enzymatic activity that performs a single proteolytic cut to release the N-terminus of NS3 (Grakoui et al., 1993, Hijikata et al., 1993). The action of NS3 proteinase (NS3P), which resides in the N-terminal one-third of the NS3 protein, then yields all remaining nonstructural proteins: NS4A, NS4B, NS5A, and NS5B. The C-terminal two-thirds of the NS3 protein contain a helicase. While the functional relationship of these two domains is unknown, the separately expressed proteinase and helicase domains of NS3 exhibit their respective activities in vitro (Suzich et al., 1993; D'Souza et al., 1995; Steinkühler et al. 1996).

The N-terminal domain of NS3 has been found to contain the catalytic motif of a trypsin-like serine proteinase (Miller and Purcell, 1990). The positions of amino acids His-57, Asp-81, and Ser-139 (numbering from the start of NS3) are strictly conserved among all HCV–derived sequences; their relative order and spacing in the sequence correspond to the catalytic triad of the trypsin family. However, the NS3P exhibits several other features that are highly unusual for a trypsin-like proteinase: it is covalently attached to a helicase possessing NTPase activity, it requires a protein cofactor (NS4A), and displays sensitivity to divalent metal ions. Using in vitro transcription/translation systems, the NS4A protein (located immediately downstream of NS3 on the polyprotein) was shown to be required for cleavage at the NS3/NS4A, NS4A/NS4B, and NS4B/NS5A sites. The NS3/NS4A cleavage occurs rapidly and in *cis* (intramolecular event), while the others occur in trans (intermolecular events). NS4A also accelerates the rate of cleavage at the NS5A–NS5B junction (Bartenschlager et al., 1994; Failla et al., 1994; Lin and Rice, 1995; Koch et al., 1996). While NS4A may have several functions, it has been suggested that NS3 and NS4A proteins form a complex that is important for modulation of proteolytic activity (Hijikata et al., 1993; Lin et al., 1994).

The proper understanding of these unique features of the NS3P as a molecular target can be achieved only in the context of a high resolution atomic structure, which in turn would also accelerate design and development of new drugs to treat HCV infection. However, crystallization of NS3P has proven difficult owing to its poor solubility and tendency to aggregate. To circumvent this problem, we examined a series of NS3P domains of variable sizes, derived from different HCV strains. Here, we report the three-dimensional structure of the NS3P domain from the strain BK, which belongs to the HCV genotype 2b. HCV BK NS3P shows significant amino acid sequence identity with representative HCV genotypes: 89% with HCV H (genotype 1a), 96% with

HCV J (1b), 71% with HCV J6 (2a), 71% with HCV J8 (2b), and 78% with HCV 3A (3a). It shares 33% sequence identity with the NS3 proteinase from the recently discovered hepatitis G virus (Linnen et al., 1996). Crystals of HCV BK NS3P were obtained using a 189 amino acid recombinant fragment purified from Escherichia coli. This structure represents the first view of a processing enzyme from the flavivirus family and opens new possibilities for design of drugs targeting HCV replication.

Results and Discussion

Overall Structural Features

HCV NS3 proteinase (NS3P) is folded into two six-stranded β barrels, similar to those of trypsin-like serine proteinases (Figures 1A and 2A–2B). However, apart from the three-catalytic triad residues (His-57, Asp-81, Ser-139) of NS3P and the sequence of GXSGG at Ser-139, it shares virtually no sequence similarity with proteinases possessing a trypsin-like fold (Figure 1B). In our discussions, features of NS3P are described while it is compared with other classes of trypsin-like proteinases. These classes include cellular enzymes such as pancreatic proteinases (e.g., elastase; Meyer et al., 1988) and bacterial proteinases (e.g., α-lytic proteinase; Fujinaga et al., 1985), all having serine as the active-site nucleophile. Also included are viral proteinases with a nucleophile of either serine (e.g., Sindbis virus core protein; Choi et al., 1991) or cysteine (e.g., rhinoviral 3C proteinase; Matthews et al., 1994). These examples were chosen because they contain an uncharged S1 specificity pocket, which for elastase and α-lytic enzymes is also relatively small. NS3P is expected to have a small uncharged specificity site because the consensus amino acid at substrate P1 is either cysteine or threonine (Grakoui et al., 1993).

The number of amino acids at the NS3P N-terminus, before the first β barrel, is significantly larger than in most proteinases in their active states. These initial 30 residues extend away from the protein and form several β strands that interact with neighboring molecules (Figure 2). The possible significance of this interaction is discussed later. The existence of secondary structure in the N-terminus is reminiscent of the short β strand found in many cellular proteinases (e.g., residues 19–22 of elastase) but differs from the case of the α helix found in the picornaviral 3C proteinases.

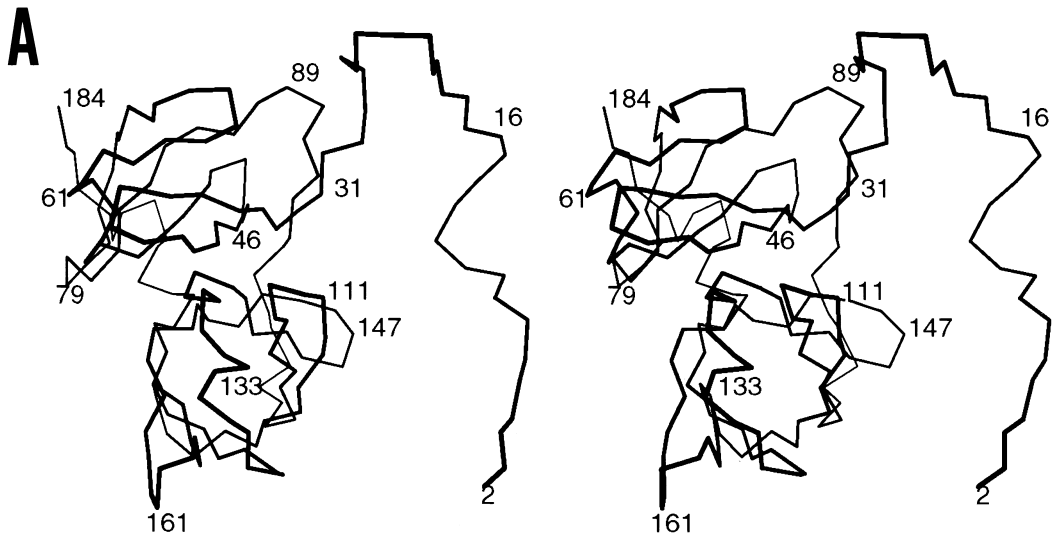
NS3P displays a spatial arrangement of strands within its β barrels similar to other trypsin-like proteinases; however, the loops connecting these strands are relatively short (Figures 2 and 3). In this regard, NS3P parallels the economical use of amino acids found in Sindbis core protein (Figure 3B), or possibly even a viral 2A proteinase (the presumed trypsin-like cysteine proteinase upstream of 3C in picornaviruses and enteroviruses, but with unknown structure). The total number of residues comprising the two β barrel motifs (from the first β barrel strand to the end of the last β barrel strand) is about 140 for NS3P, only slightly larger than 133 in Sindbis, and similar to the predicted number of about 140 for 2A proteinases (Bazan and Fletterick, 1988); this number exceeds 170 residues in the cellular proteinases. As a consequence of this economy, several loops common in the cellular proteinases are absent in NS3P, such as the calcium-binding loop (βD1 to βE1), the autolysis loop (βA2 to βB2), and the so-called methionine loop (βB2 to βC2). In the bacterial and picornaviral proteinases, this latter loop (but in pancreatic proteinases, the βE1 to βF1 loop) is positioned such that it can interact with residues on the P side of the substrate (Figure 3B); the absence of a corresponding loop in NS3P is consistent with its apparent lack of substrate recognition over P2 to P5 (Bartenschlager et al., 1995; Grakoui et al., 1993).

After the final strand βF2, there is one turn of α helix (171–174) that closely matches the first turn of the C-terminal helix found in cellular proteinases. Following this turn, the polypeptide is disordered in two of three monomers in our asymmetric unit. In the monomer discussed here, the chain turns back toward βE1 to form a short antiparallel β-interaction between βE0 (181–182) and βE1b (75–76; see Figure 2); this interaction may be the result of crystal packing only. Questions of flexibility and conformation at the C-terminus of NS3P are relevant because this crystal structure represents an enzyme which in vivo is permanently attached (via its C-terminus) to a helicase, the two entities together defining the NS3 protein. The exact range of residues within NS3 that corresponds to the folded helicase is unknown, as is the length of the polypeptide linker between NS3P and the helicase.

Zinc-Binding Site

In cell-free transcription/translation experiments, processing of the HCV NS polyprotein by NS3P is stimulated by the addition of Zn^{2+}. Inductive coupled-plasma mass spectroscopy of purified recombinant NS3P reveals an equimolar ratio of NS3P and zinc (Z. H., unpublished data). Thus, the existence of a functionally relevant zinc in HCV NS3P was expected. From our NS3P homology models, the close proximity of conserved cysteines 97, 99, 145, and histidine 149 suggested a zinc-chelation site, because these residues are frequently members of structural zinc sites (Vallee and Auld, 1990; Schwabe and Klug, 1994), and because disulfide linkages are unlikely in an intracellular proteinase. The crystal structure confirms this idea; these three cysteines together provide a partial tetrahedral geometry around the zinc ion (see Figure 2B), with cysteine sulfur to zinc distances of 2.0–2.5 Å. In two of three monomers in the asymmetric unit of the crystal, the fourth member of the tetrahedral coordination of the zinc is a water molecule; this water is within hydrogen-bonding distance of the His-149 side chain. In the third monomer, His-149-Nδ is 4.0 Å away from the zinc and thus does not play a direct chelation role, but the imidazole isolates the zinc from solution and is positioned to coordinate the metal readily (Figure 2B). Point mutations of Cys-97, -99, -145, and His-149 (to alanine) show that removal of any one has a negative impact on NS3P processing (Hijikata et al, 1993; Stempniak et al., submitted). Thus, His-149 may be an integral part of zinc coordination at least during the initial folding of NS3P.

The zinc site serves to anchor the turn at βD2–βE2 (containing Cys-145 and His-149) to the interbarrel loop



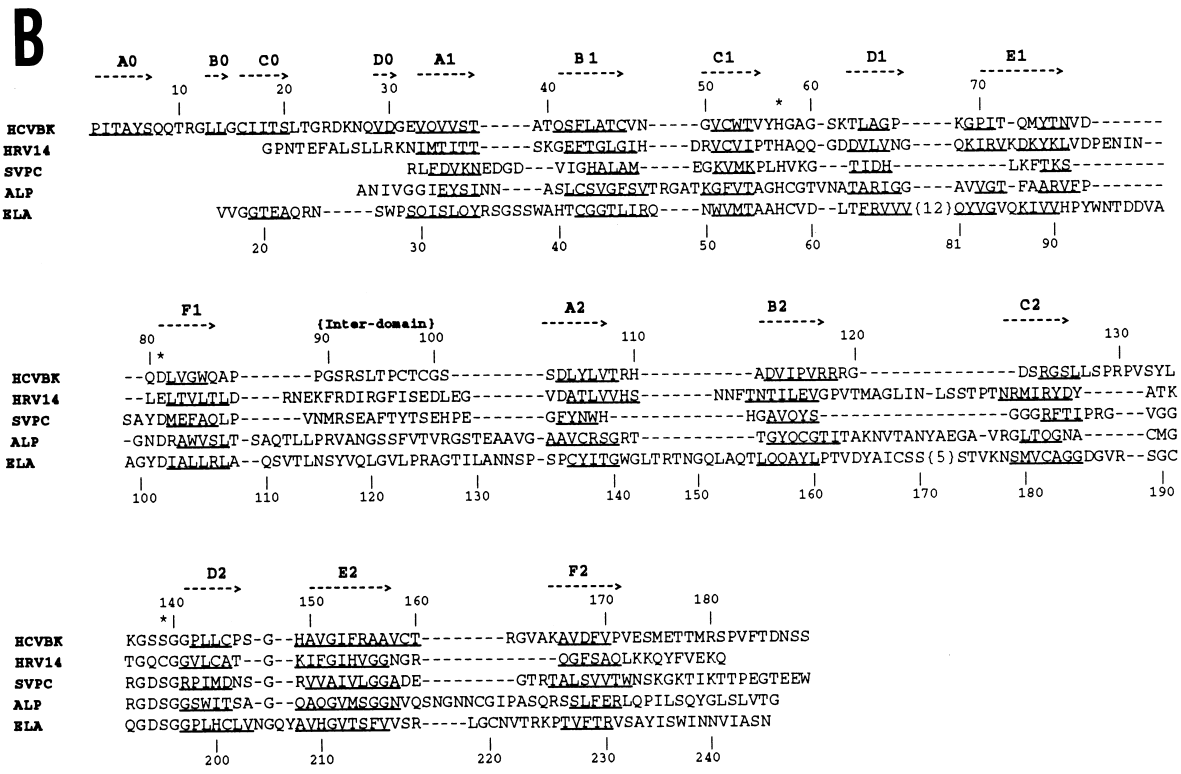


Figure 1. Trace of HCV NS3P Structure and Its Sequence Alignment with Other Proteases

(A) Side-by-side stereo view of the Cα trace for HCV NS3P (residues 2–184), drawn using Xfit (McRee, 1992). This is monomer 1 of the asymmetric unit. The trace is labeled at intervals for clarity. The N-terminal 30 residues (before the first β barrel) are extended away from the core of the protein.

(B) Alignment of HCV BK NS3P sequence with other proteinases based on superposition of crystal structures (as in Figure 3B). Residues of β strands are underlined. The general location of each β strand among all sequences is indicated by a dashed arrow; labeling of these strands follows the common convention, except for the extra N-terminal strands of HCV, which are labeled A0–D0. Catalytic triad residues are marked by an asterisk. Excess residues not shown are indicated in brackets. Abbreviations for proteins are: SVCP, Sindbis virus core protein; ALP, α-lytic proteinase; ELA, porcine elastase; HRV14, human rhinovirus type 14 3C proteinase; HCVBK, hepatitis C virus, BK strain NS3 proteinase. Amino acid numbering is for either HCVBK (top) or ELA (bottom).

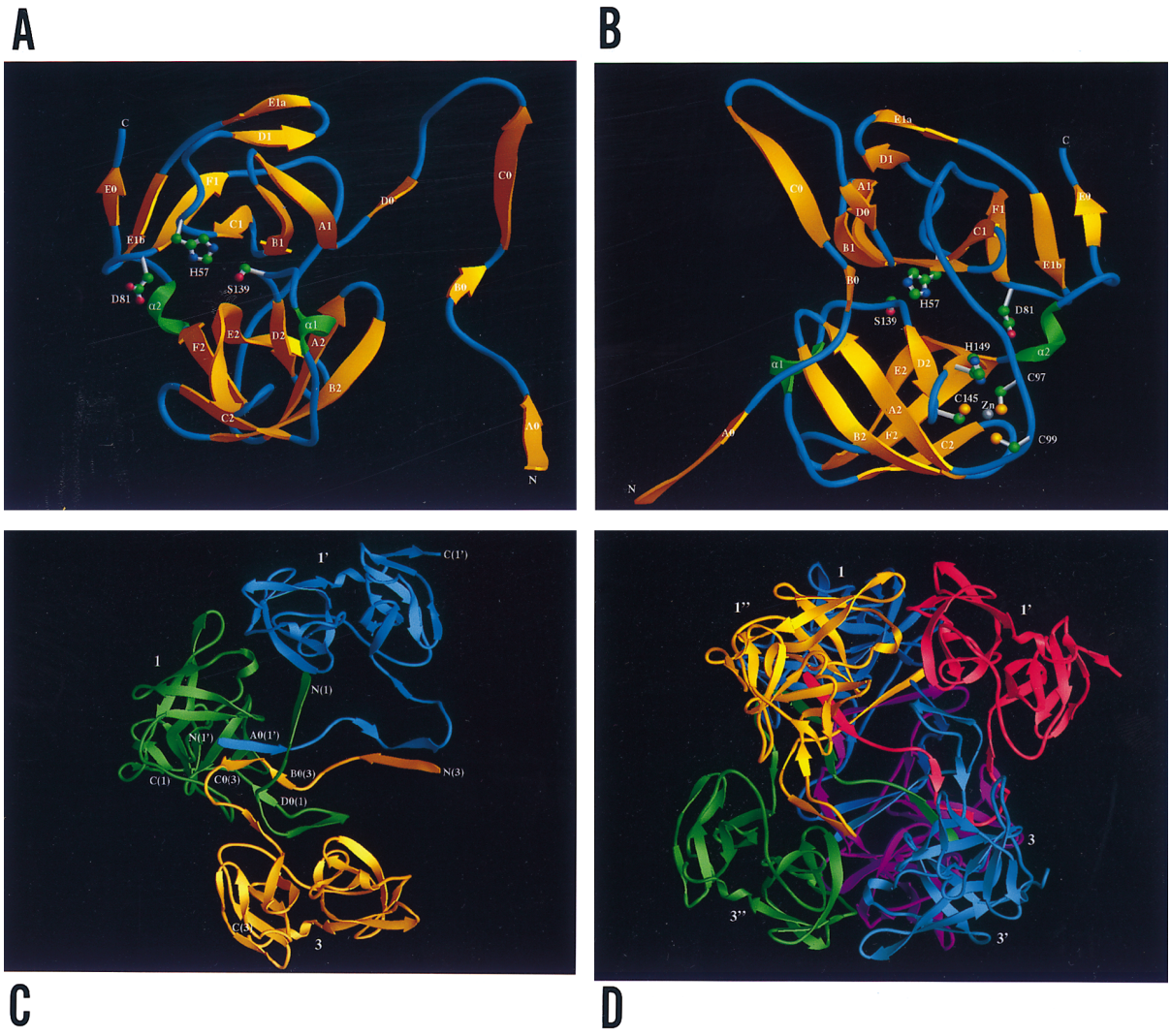


Figure 2. Overall Folding of HCV NS3P and Assembly of Molecules in the Crystal, Emphasizing N-Terminal Configuration and Exchange Displayed by Ribbons (Carson, 1991).

(A) View into the active site of monomer 1 in the asymmetric unit with catalytic triad residues labeled by the single amino acid code. Secondary structural elements are color-coded yellow for β strand, green for α helices, and blue for coil. β strands in the trypsin-like barrels (A1–F2) are labelled in the standard convention. Catalytic triad residues are shown in sphere and cylinder representation, with spheres color-coded green for carbon, blue for nitrogen, and red for oxygen.

(B) Trace of monomer 1 color-coded as in (A), rotated relative to (A) for viewing of the structural zinc-binding site. The zinc (gray sphere) is tetrahedrally coordinated by the sulfurs (yellow spheres) of cysteines 97, 99, and 145. Histidine 149, though not within bonding distance, is seen poised to complete the tetrahedral coordination (see text).

(C) Architecture of the N-terminal strand exchange. The N-terminal residues of monomer 1 (in green) extend away from the molecule permitting two antiparallel β strands, A0 of monomer 1', in blue (generated by the crystallographic 3-fold from monomer 1), and C0 of monomer 3 (in yellow), to lie on its surface. This interaction involves nonpolar side chains from the blue and yellow β strands being buried against a hydrophobic patch on the surface of the green molecule (see text). A short parallel β sheet interaction is also seen, involving D0 and B0 strands of monomers 1 and 3, respectively.

(D) Hexamer generated from monomers 1 and 3 using the crystallographic (3-fold) and noncrystallographic (2-fold) symmetry elements. In this view, the crystallographic 3-fold axis is vertical. The crystallographic trimer of monomer 1 (blue, red, and yellow) associates with the crystallographic trimer of monomer 3 (purple, light blue, and green) using N-terminal strand exchange (see text). Each molecule within the hexamer accepts two antiparallel β strands from neighboring molecules; e.g., the green molecule accepts the purple and yellow strands. The noncrystallographic 2-fold symmetry mates are red and green, yellow and purple, and blue and light blue.

(containing Cys-97 and Cys-99). Such a location for zinc, remote from the active site, implies a structural rather than a catalytic role for the metal; its effects on polyprotein processing are probably linked to accurate NS3P folding or post-folding stability of the enzyme or both. However, perturbations at the zinc site could conceivably affect the active-site conformation, because the two are linked directly through strands βD2 and βE2.

During our derivative searches (see Experimental Procedures), we found that many metals and heavy atom

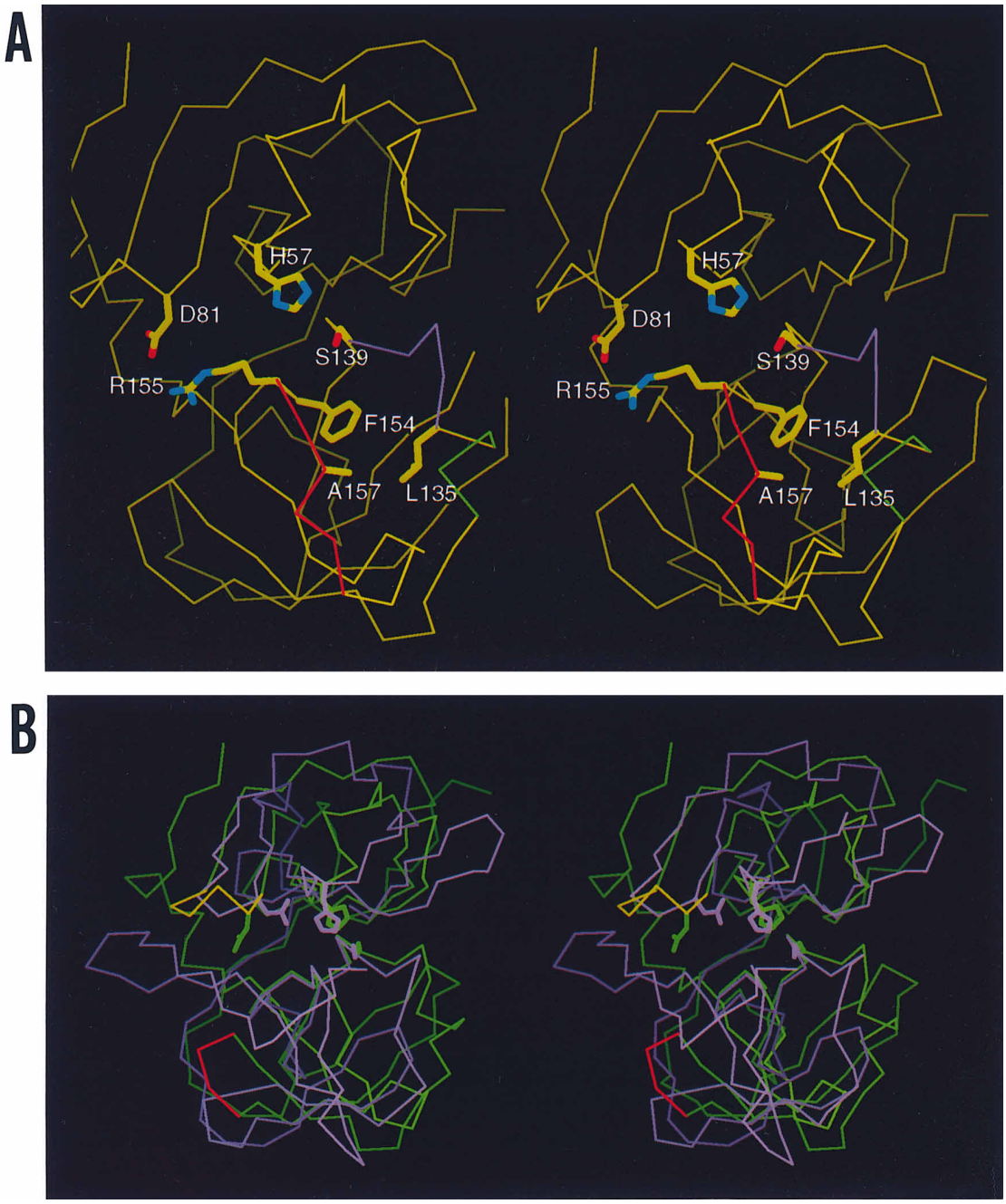


Figure 3. Active Site of NS3P and Comparison with Other Trypsin-like Serine Proteinases

(A) Side-by-side stereo view of the active site and surroundings of HCV NS3P. A Cα trace is given for clarity, with purple coloring the oxyanion loop, green for the short α helix preceeding the active-site serine, and red for the section of βE2 that interacts with substrate. Side chains, whose atoms are colored yellow (carbon), blue (nitrogen), or red (oxygen), are labeled by the single-letter amino acid code. Shown here is the active-site triad (H57, D81, S139) and residues defining the S1 specificity pocket (L135, F154, A157). Overall, this S1 pocket is relatively small and nonpolar.

(B) Superposition of HCV NS3P (green) and Sindbis core protein SCP (purple). This superposition is based on structural alignment of β structure within the barrels, which gives a root-mean-square difference on Cα's of 1.5 Å. The catalytic triad residues are shown for each molecule (thicker lines) to illustrate their similar spatial arrangement, as well as the torsional deviation of Asp-81 in NS3P. The βE2–βC2 loop (red) and βE1–βF1 loop (yellow), often determinants of P2–P5 substrate specificity in pancreatic/bacterial proteinases, are unusually short in NS3P as well as SCP. The extended N-terminus of NS3P has been omitted for clarity.

compounds bind at this zinc site in the crystal. For the case of mercury binding, we observed a disruption of the structure in this region, essentially through a disordering of the polypeptide around Cys-97–99. Other experiments showed that the zinc ion is lost when mercury binds to NS3P (presumably at cysteine) and that this binding inhibits the enzyme (Z. Hostomska, unpublished data). Copper has also been found as a potent inhibitor of NS3P (Han et al., 1995). Copper binding in our crystal occurs primarily at the zinc site, but substitution of zinc with copper is not yet confirmed. In the cases above, inhibition may result from an altered active site that in turn arises from destabilization of the zinc site during metal binding.

The zinc-coordinating amino acids of NS3P are not present in other members of Flaviviridae family such as yellow fever virus or bovine diarrhea virus but are found in the recently discovered GB viruses (Simons et al., 1995) and hepatitis G virus (Linnen et al., 1996), which are more closely related to HCV. The zinc site of NS3P may play a role analogous to the disulfide common among cellular proteinases (136–201 in elastase), which also anchors the βD2–βE2 turn to the interbarrel loop. In picornaviral 3C proteinases, which possess neither disulfides nor a structural zinc, stability in this region may be provided by the N-terminal α helix, which packs against the βD2–βE2 turn and connecting loop simultaneously.

Alignment of picornaviral 2A proteinase sequences (Yu and Lloyd, 1992; Bazan and Fletterick, 1988) has revealed two conserved motifs, CXC within the interbarrel loop and CXH at the end of βD2, which are similar in residue type and sequence position to the chelation motifs in NS3P (CXC and CXXXH). Several biochemical and biophysical studies of 2A proteinases have shown that zinc is an integral tightly bound component of the structures, required for correct folding and stability (but not involved in catalysis) and probably chelated by residues of the two motifs (Yu and Lloyd, 1992; Sommergruber et al., 1994; Voss et al., 1995). Thus, in the picornaviral 2A proteinases there is probably a zinc site with general location, coordination geometry, and stabilizing function that parallels the NS3P case.

Active Site

In the crystal structure of NS3P, the catalytic triad residues (His-57, Asp-81, Ser-139), the "oxyanion-stabilizing loop" (135–139), and strand βE2 forming one side of the specificity pocket, together have the same relative spatial positions as in other trypsin-like proteinases (Figure 3). Mutation experiments have identified the presumed NS3P triad residues as essential to proteolytic activity (Hijikata et al., 1993).

The imidazole of His-57, expected to extract a proton from nucleophile Ser-139 during the enzymatic reaction, is oriented toward Ser-139 but is not close enough to form the hydrogen bond often observed in proteinase structures. The stretch of residues from 57 to 63 has greater mobility than most loops in the structure, as indicated by higher temperature factors; however, the conformation we have built here is consistent with our experimental density maps and resembles the helical turn found in other proteinases. Flexibility in this region may be related to a lack of structural anchoring compared with cellular proteinases, where a conserved disulfide links Cys-58 to Cys-42 on strand βB1.

The side chain of Asp-81, expected to provide charge stabilization for His-57 after deprotonation of Ser-139, is oriented away from His-57 in the crystal structure of NS3P. However, the Ser-139-Cα to Asp-81-Cα distance of 11.4 Å is close to that in rhinovirus 3C proteinase (11.4 Å) and also Sindbis core protein (10.9 Å), while only slightly longer than that in cellular proteinases (approximately 10 Å). Asp-81 forms an ion pair with Arg-155 (on βE2), the latter being a conserved amino acid among HCV sequences. Arg-155 in NS3P corresponds to conserved Ser-214 of other proteinases, where the hydroxyl of the serine hydrogen-bonds to the carboxylate of triad member Asp-102. Interestingly, the rotation of Asp-81 away from His-57 observed in NS3P, and the Asp-81–Arg-155 interaction, are paralleled in the crystal structure of 3C proteinase from hepatitis A (Allaire et al., 1994), where the aspartate points away from histidine while interacting with a lysine on strand βF2. In the hepatitis A case, however, it was suggested that the aspartic acid is not crucial for histidine charge stabilization because the cysteine nucleophile is easier to deprotonate than serine. In NS3P, only minor readjustments of amino acids 80–82 would position the Asp-81 carboxylate closer to the His-57 side chain, as probably required for catalysis. We propose that a classic catalytic triad configuration for His-57/Asp-81/Ser-139 will exist during substrate cleavage by NS3P in vivo and in vitro and that the positional deviation of Asp-81 we observe is a consequence of an apo-enzyme structure or the crystallization conditions or both.

The sequence around Ser-139 in NS3P (Gly-Ser-Ser-Gly-Gly) follows the conserved GXSGG motif seen in trypsin-like serine (and cysteine) proteinases, and the polypeptide backbone conformation of the oxyanion-stabilizing loop (135–139) is nearly identical to that in other proteinases (see Figures 3B and 6). Alignment of all backbone atoms of NS3P residues 135–139 with the corresponding atoms of elastase (191–195) gives a root-mean-square difference of 0.71 Å. The βC2–βD2 loop, always variable among proteinases, is unique in NS3P because it contains one turn of α helix (residues 131–134) just prior to the oxyanion loop. This helix could add conformational stability at the binding site, perhaps analogous to the conserved disulfide found in many proteinases (involving Cys-191) that anchors the oxyanion loop to the βE2–βF2 turn.

The cleavage performed by NS3P is between polyprotein substrate residues Cys/Ser (at 4B/5A and 5A/5B sites), Cys/Ala (at 4A/4B), or Thr/Ser (at 3/4A). Our structure shows a specificity pocket that is shallow and nonpolar, formed primarily by the side chains of invariant residues Phe-154, Ala-157, and Leu-135 (Figure 3A). Homology modeling had predicted this type of S1 pocket, involving at least Phe-154 (Pizzi et al., 1994; see also "Crystallographic Phasing Strategies," below). Phe-154 corresponds to position 213 of the cellular proteinases, where it is typically a small hydrophobic amino acid. Picornavirus 3C proteinases have histidine at 213; thus, they show a closer steric resemblance to NS3P at this

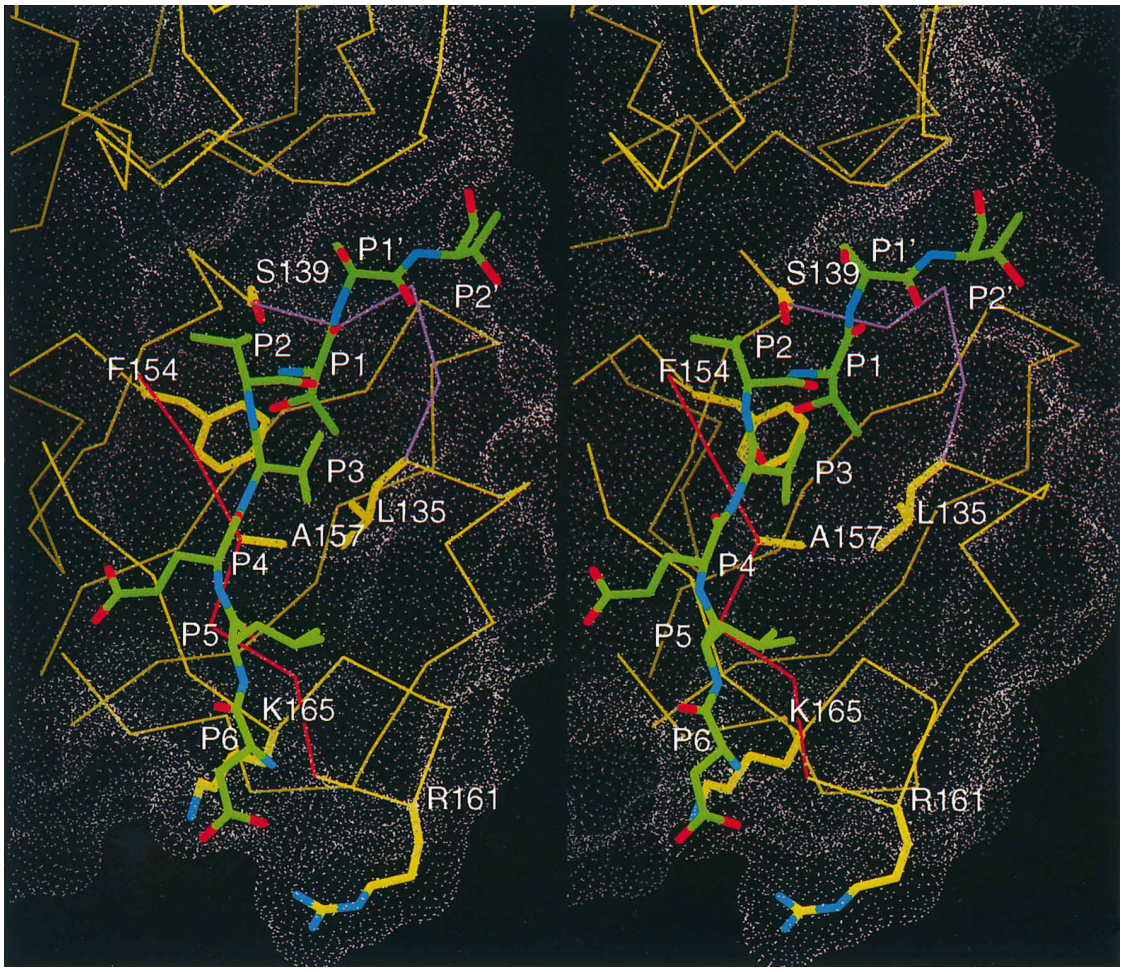


Figure 4. Model of a Bound Polypeptide Substrate Containing the NS3/4A Junction

Side-by-side stereo view of the NS3/4A substrate (color-coded: green, C; red, O; blue, N) modeled into the active site of HCV NS3P, which is covered by a molecular surface in white dots (generated by the MS program [Connolly, 1983] using a 1.6 Å probe radius). The Cα trace of NS3P is shown in yellow, with the oxyanion-stabilizing loop in purple and the extended strand βE2 in red. NS3P side chains (color-coded: yellow, C; red, O; blue, N) are shown for residues comprising the S1 specificity pocket (L135, F154, A157), the potential P6 recognition elements (R161, K165), and active-site serine (S139). The substrate is labeled according to the standard P/P' convention, while the HCV residues are labeled in the single-letter amino acid code.

position. The Leu-135 side chain forms the bulk of the remaining S1 pocket and is located approximately at position 190 of cellular proteinases, often a small polar amino acid and specificity determinant. Ala-157 and Val-167 of NS3P correspond to positions 216 and 226, respectively, of cellular proteinases, which are usually glycine to accommodate long P1 side chains; elastase, which recognizes a small nonpolar P1 side chain and has Val-216/Thr-226, most closely parallels NS3P at these positions.

Substrate Modeling in NS3P

Hydrolysis of the NS3/4A peptide bond is the first (in *cis*) cleavage event by NS3P, involving a threonine at the P1 position of the substrate; for subsequent cleavage events, the P1 residue is cysteine. A model of the Michaelis complex between a polypeptide substrate containing the NS3/4A cleavage site and NS3P was constructed (Figure 4), based on the known binding mode of polypeptide inhibitors of trypsin-like enzymes (Read and James, 1986). This model suggests the possibility of a favorable interaction between the OH (or SH) of the P1 side chain and the electron-rich π clouds on the aromatic ring of nearby Phe-154. A sulfhydryl–aromatic interaction was earlier proposed as a possible NS3P substrate recognition mechanism (Pizzi et al., 1994). Interaction between a threonine hydroxyl and a tyrosine aromatic ring has been observed (Ji et al., 1992; Liu et al., 1993). Numerous observations and calculations support the notion that hydrogen bonding between proton donors and aromatic rings plays a significant role in protein stability (Burley and Petsko, 1986; Levitt and Perutz, 1988).

Strand βE2 of NS3P, expected to form β sheet antiparallel hydrogen bonds with the substrate over P2–P3, is more extended than in most proteinases. Its length most closely resembles the corresponding β strand in Sindbis core protein. We hypothesize that the NS3P enzyme–

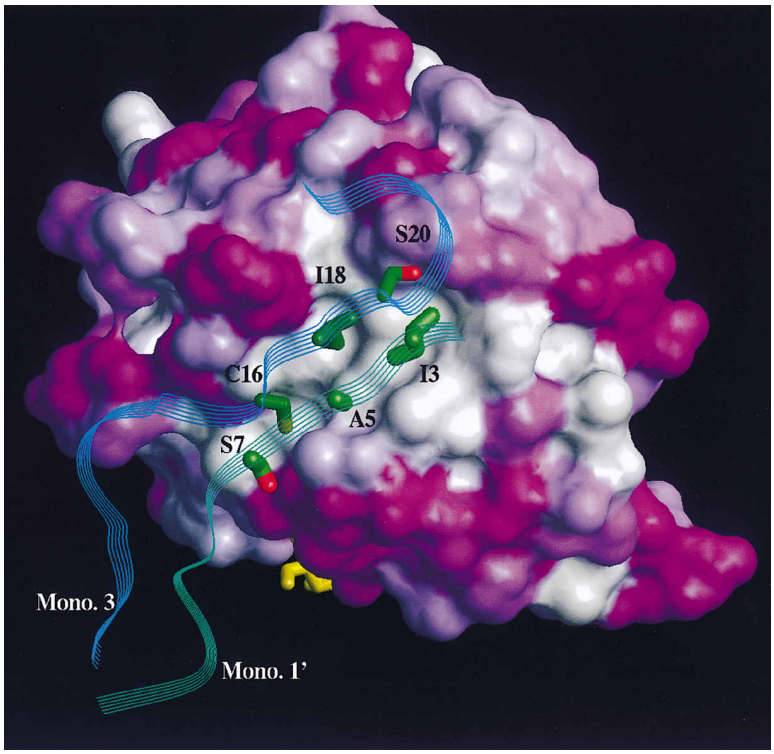


Figure 5. Solvent-Accessible Surface of NS3P Monomer 1, Colored by Hydrophobicity, Such That Nonpolar Residues Are White, Charged Residues Are Deep Magenta, and Polar Residues Are Medium Shades of Magenta

The central white region is the hydrophobic patch (approximately 400 Å^2) discussed in the text. In the crystal, there are two N-terminal strands from neighboring molecules bound to this patch, namely βC0 from monomer 3 (blue ribbon) and βA0 from monomer 1' (green ribbon). Strand βC0 buries Cys-16 and Ile-18, while strand βA0 buries Ile-3 and Ala-5, into the center of the hydrophobic surface patch. Residues Ser-20 of βC0 and Ser-7 of βA0 lie at the edge of the interface and hydrogen-bond to monomer 1. Also shown in yellow (at bottom horizon of surface) is the P3′ residue of a modeled substrate polypeptide (see Figure 4), included here as a directional reference to the active site.

substrate β-interaction at P2–P3 continues for another three residues to P6 (Figure 4), for two reasons. First, the consensus acidic residue at substrate P6 could interact with the Arg-161 and Lys-165 of NS3P, which are relatively isolated on the extended βE2–βF2 turn. Secondly, a continuous P2–P6 main-chain β-interaction might compensate for the apparent lack of P2–P5 side chain to enzyme interactions, which results from the relative shortness of the βB2–βC2 and βE1–βF1 loops in NS3P.

N-Terminal Configuration

A quite unexpected feature of the crystal structure of NS3P is the configuration of its long N-terminus (the first 30 amino acids), which extends away from the protein and contains β strands that interact with neighboring molecules (see Figure 2). Monomers 1 and 3 of the asymmetric unit ultimately assemble into a hexamer via a combination of crystallographic symmetry (a trimer is generated from either monomer) and noncrystallographic symmetry (association of these two trimers). The order of this assembly is unknown. A similar hexamer is formed from six copies of monomer 2 by (32) crystallographic symmetry operators. However, weaker electron density at the N-termini in this hexamer precludes the building of polypeptide chains between copies of monomer 2.

One consequence of the strand exchange is that a hydrophobic patch on the surface of each molecule is covered by two relatively nonpolar β strands from N-termini of a different neighbor (Figures 2C and 5). Specifically, monomer 3 contributes βC0 (residues 16–20; sequence CIITS) to the patch on monomer 1, with Cys-16 and Ile-18 buried and Ser-20 partially buried at the interface boundary. Antiparallel to this strand, and also lying against the patch, is βA0 (residues 2–7; sequence PITAYS) from a crystallographically related copy of monomer 1, which buries Ile-3 and Ala-5 into the patch while partially burying Ser-7. An additional association, apart from the patch, involves a short two-stranded parallel β-interaction (βB0: βD0 between monomers 1 and 3) that occurs in between trimers, where two traversing N-termini approach one another (Figure 2C).

The "hydrophobic patch" is a long narrow (approximately 20 Å × 8 Å) shallow valley with a continuous nonpolar molecular surface area of about 400 Å^2, located mainly on the second β barrel, and composed of residues primarily from strands βA2, βB2, and βD2 but also two residues from βA1 (Figure 5). Specifically, the patch is formed by the side chains of conserved residues Val-33, Val-35, Leu-44, Leu-94, Val-107, Leu-127, Ala-111, Val-113, Pro-115, Pro-142, Leu-144, Tyr-105 (ring portion), and Arg-109 (alkyl chain portion).

Since hydrophobic interactions are often key components of molecular recognition, the association of the two β strands with the hydrophobic patch on NS3P may mimic one or more molecular interactions that occur during HCV polyprotein processing, rather than being an artifact of crystallization. The patch could qualify as a functionally important binding site because it is defined by residues from strand βD2, which contains the nucleophile, and by the antiparallel strands βA2/βB2, which together directly support and hydrogen-bond to the oxyanion loop. It is known that the NS4A domain (immediately downstream of NS3) binds to NS3P during processing, and this facilitates the cleavages by NS3P. Furthermore, the N-terminus of NS3P itself participates in the cofactor modulation of NS4A (Koch et al., 1996).

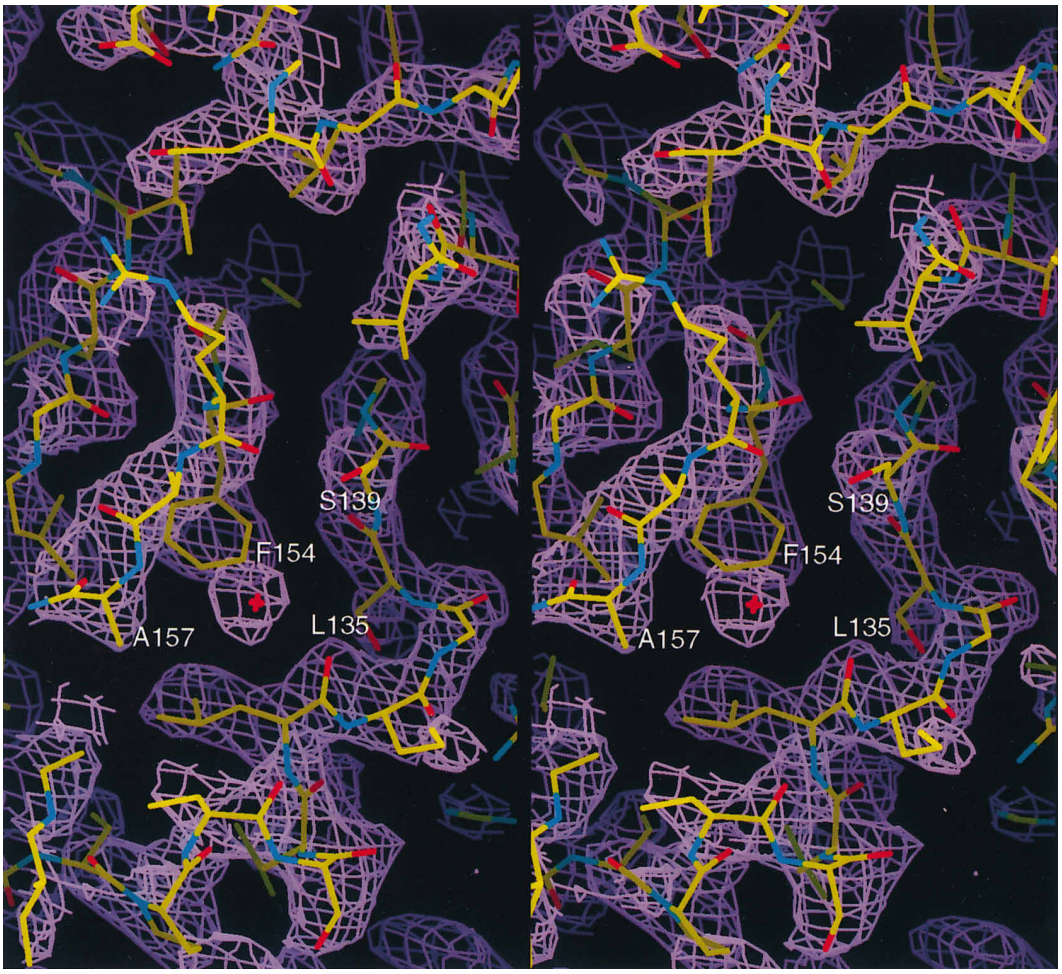


Figure 6. Section of the 2.7 Å Experimental Electron Density Map

Side-by-side stereo view of NS3P (color-coded: yellow, C; red, O; blue, N) in the vicinity of the active site, superimposed on the 2.7 Å experimental ISIRAS map (purple mesh), contoured at 1.5 σ. A clear definition of main chain, side chain, and, frequently, carbonyl direction exists throughout the majority of the map.

Therefore, one possibility is that this hydrophobic patch forms part of a site through which the N-terminus of NS3P or the cofactor NS4A or both modulate the activity of NS3P.

Experimental Procedures

Protein Expression and Purification

HCV BK NS3 proteinase (residues 1–189) was expressed in E. coli as a soluble protein. The isolation of a pure protein involved three chromatographic steps: Fast Flow SP Sepharose; FPLC Mono-S, and Sepharose S200. The final preparation of HCV BK 1–189 migrated as a single entity on SDS–polyacrylamide gel electrophoresis in reducing and nonreducing buffer system, with an apparent molecular mass of 20,000 kDa. The specific activity of the pure sample of HCV BK 1–189 was about 50 nmol/min/mg using a 5A/5B 15 residue synthetic peptide substrate (C. Lewis, unpublished data). Automated Edman degradation gave an N-terminal sequence PITAYS; thus, the initiation methionine and alanine residues from the original sequence are missing. Details of protein purification and enzyme characterization will be published elsewhere.

Protein Crystallization and X-Ray Data Collection

Crystals were grown at 4°C by the hanging-drop vapor diffusion method, using plastic Linbro tissue culture plates. Aliquots (5 μl) of 10 mg/ml of protein HCV BK 1–189 in 50 mM sodium acetate, 10 mM dithiothreitol, 350 mM NaCl (pH 6.0) were mixed with 5 μl of reservoir containing 5% PEG 400, 3.5 M NaCl, 150 mM Tris–HCl (pH 8.5). Crystals appeared after 2–3 weeks, and they grew as hexagonal rods with typical dimensions 0.1 $\times$ 0.1 $\times$ 0.6 mm.

The crystals belong to space group R32 with hexagonal cell parameters of a = b = 133 Å, c = 223 Å. There are three molecules per asymmetric unit, which gives a calculated solvent content of 62%. The extreme sensitivity of heavy atom–soaked crystals to X-rays, plus the eventual use of synchrotron radiation, prompted a uniform application of cryo-cooling techniques (−170°C). In-house data were obtained from a 30 cm MAR imaging plate and processed with DENZO and SCALEPACK (Otwinowski and Minor, 1993), and complete data to 3.5 Å (derivatives) or 3.0 Å (native) were routinely obtained. Synchrotron sources at Photon Factory, Japan, and European Synchrotron Radiation Facility, France, were ultimately used to obtain higher resolution native and derivative data (see Table 1).

Crystallographic Phasing Strategies

During heavy-atom screening in house, most trials led to isomorphous difference Patterson maps containing peaks at similar positions; these peaks were also seen in several mercury and gold anomalous difference Pattersons. Compounds with a wide variety of binding mechanisms led to the same peaks. Nonisomorphism was a significant problem, and data from short soaks frequently

Table 1. Crystallographic Data Statistics

Parameters	$HgCl_2$/Native (1)	BMDA[a]	Native (2)
	Diffraction data		
Heavy atom (mM)	0.05	6.0	
Soaking time (hrs)	25.0	41.0	
Synchrotron facility[b]	Photon Factory	Photon Factory	ESRF
Data collection device[c]	Fuji ip	Fuji ip	Mar ip
Wavelength (Å)	1.0000	1.0064	0.995
Maximum resolution (Å)	2.7	2.5	2.4
Data completeness (%)	93.8	93.9	97.8
I/σI at resolution limit	3.8	4.4	3.9
Observations (No.)	93,997	135,383	96,285
Unique reflections	19,945	24,998	29,437
Rsym[d] (%)	7.5	7.7	7.9
Riso[e] (%)		16.5	
	SIRAS statistics to 2.7 Å		
R_{cullis}[f]		0.56	
R_{cullis}[g] anomalous		0.32	
FOM[h]		0.69	
solvent flattened FOM		0.87	
rms f/Eiso		2.34	
rms f′′/Eano		2.86	
Sites (No.)		9	

[a] 3,6-bis(mercurimethyl) dioxane acetate.
[b] Beamline 18B, Photon Factory Tsukuba City, Japan; Beamline 4, ESRF European synchrotron radiation facility in Grenoble, France.
[c] Fuji ip: data measured using a Weissenberg camera in oscillation mode, Mar ip: Mar research image plate system. Data from both systems were processed with DENZO and SCALEPACK (Otwinowski and Minor, 1996).
[d] R_{sym} is the unweighted R factor on I between symmetry mates.
[e] $R_{iso} = \Sigma(|F_{der}|-|F_{nat}|)/\Sigma|F_{nat}|$ where F_{der} and F_{nat} are derivative and native structure factor amplitudes, respectively.
[f] $R_{cullis} = \Sigma||F_{der} +/- F_{nat}|-F_{H(calc)}||/\Sigma|F_{der} +/- F_{nat}|$ for all centric reflections.
[g] R_{cullis} anomalous was calculated for the top 25 percent largest Bijvoet differences.
[h] FOM is the figure of merit.

had to serve as "native" relative to longer soaks as "derivative." A total of three major sites per asymmetric unit was eventually deduced from Patterson maps using HASSP (Terwilliger et al., 1987). Later, it was found that each site lay near a cluster of cysteines that formed a structural zinc site and that this region was disrupted by heavy-atom binding. The best in-house phases were obtained by a combination of isomorphous and anomalous data from $HgCl_2$ soaks. Heavy-atom site refinement using PHASES (Furey and Swaminathan, 1990) at 5 Å gave an overall figure of merit, phasing power isomorphous/anomalous, and centric R factor of 0.75, 2.5/3.0, and 0.45, respectively. This was used to initiate solvent flattening (Wang, 1985) and to check the handedness of the sites. Phasing information beyond 5 Å was poor, but the 5 Å electron density maps showed well-defined and separated molecular envelopes, indicating three molecules per asymmetric unit. The heavy-atom positions were used to determine noncrystallographic operators, so that 3-fold averaging combined with phase extension to 3.5 Å could be initiated using PHASES. While such maps could not be used for tracing, they did show features that suggested a trypsin-like structure, namely two globular domains, one of which resembled a thick-walled barrel. This permitted a positioning of our NS3P homology model into the unit cell. This model was based on rhinovirus 3C proteinase and generated with LOOK (Molecular Applications Group, Stanford, CA). When higher resolution maps became available (see below), it was found that this model had been correctly positioned and could serve as a guide for building the structure.

Data collected at Photon Factory from two mercury-soaked crystals, using 3,6-bis(mercurimethyl) dioxane (BMDA) and $HgCl_2$, provided the higher resolution phases necessary to produce traceable maps. The isomorphous signal from differences between BMDA and $HgCl_2$ ($HgCl_2$ being more native-like), combined with the anomalous signal from the BMDA data, led to the best results (Table 1). For BMDA, a total of six more sites per asymmetric unit was found by difference Fourier methods; these were spatially close to the first set of three. Phasing statistics indicated usable experimental phases to 2.7 Å. Density maps calculated to 3.0 Å and 2.7 Å revealed a nearly complete trace for the protein, with abundant side-chain information, and confirmed that NS3P folds with a double–β barrel trypsin-like motif.

Model Building and Refinement

Based on 3 Å and 2.7 Å solvent-flattened electron density maps, a model was built for one molecule of the asymmetric unit into the region with best connectivity, using FRODO (Jones, 1978) and XFIT (McRee, 1992). This model was then copied to the other two positions via the previously refined noncrystallographic symmetry (NCS) operators and adjusted locally where necessary. At this point, each monomer consisted of a trypsin-like core with two disconnected parallel elongated strands lying upon one surface, which represented in some manner the N-terminus (approximately 29 residues). Side-chain density in these strands clearly defined a unique sequence fit and direction, which was inconsistent with the strands belonging to the same (nearest) molecule. For example, density for an approximately eight–amino acid β strand ending close to residue 30 (the start of the first barrel) could not be fit with the sequence 21–29 as expected; instead, it was fit clearly by 14–21 (most notably at Cys-16) and running in the opposite direction. This situation, along with nearly continuous density extending between molecules (before 14 and after 21), led to the N-terminal strand-exchange structure of NS3P.

Initial crystallographic refinement with XPLOR (Brünger, 1992b) and manual readjustment of the model were done at 2.7 Å using the $HgCl_2$ data (this being more native-like), with NCS restraints applied to the three monomers of the trimer. Later, native data to 2.4 Å resolution from European Synchrotron Radiation Facility was used, and the three molecules were refined with only weak NCS restraints, the weight chosen to optimize the free R value (Brünger, 1992a). XPLOR treatment included simulated annealing ("slow-cooling" from progressively lower temperatures as the model improved), positional refinement, and then restrained individual temperature

factor refinement. Rebuilding was done from annealed-omit maps and (2Fo-Fc) maps, with continual reference to the high quality experimental maps. With the European Synchrotron Radiation Facility native data, it was possible to construct a zinc site, based on 6 σ density in (Fo-Fc) maps, between two neighboring surface loops where cysteines 97, 99, and 145 were known to be located. The zinc site was eventually refined with only nonbonded parameters (no explicit chelation bonds) and remained stable as three sulfhydryls surrounding the zinc, with sulfur–zinc distances of 2.0–2.5 Å. Verification of the N-terminal trace and strand exchange was made for two molecules out of three in the asymmetric unit (our monomers "1" and "3"). For monomer "1," the only disordered residues are the last five (out of 188); this is the molecule chosen for the Discussion. In the other two molecules, density beyond residue 177 is poorly defined. Using only residues in the β barrel motifs of each molecule, the root-mean-square diffence in Cα positions between monomers ranges from 0.5 Å for monomers 1–3 to 0.7 Å for monomers 1–2.

A total of 125 water molecules and seven chloride ions were eventually included during the refinement process of the trimer (5058 protein atoms). The chlorides were chosen to replace those waters whose B factors had refined very low and which had locations near electrostatically unpaired basic side chains. The R factor is 22.5% over 8–2.4 Å (European Synchrotron Radiation Facility data; 25,000 reflections with I > 2.0 σ[I]), and the free R factor is 32% (2,000 reflections). Bond and angle deviations are 0.02 Å and 2.1°, respectively, as determined by XPLOR using the Engh and Huber (1991) parameters. PROCHECK (Laskowski et al., 1993) analysis of ϕ/ψ angles indicates no residues in disallowed regions and two in "generously allowed" regions. The average temperature factor is 17 Å^2 for main-chain atoms and 20 Å^2 for side-chain atoms.

Acknowledgments

Correspondence should be addressed to Z. Hostomska. We wish to thank the following individuals: Drs. D. Knighton and C. Kissinger for help with data collection and analysis; M. Stempniak, B. Nodes, N. Tajima, S. Rahmati, B. Aust, and G. Hudson for excellent technical assistance; Drs. S. Reich, E. Villafranca, P. Dragovich, and C. Lewis for valuable discussions; Dr. D. Matthews for critical reading of the manuscript; and Drs. H. Tsuge, M. Miyano, H. Ago, N. Watanabe, M. Suzuki, N. Sakabe, and E. Inagaki for their assistance with data collection at Photon Factory. This study was in part supported by the Sakabe project of Tsukuba Advanced Research Alliance.

Received September 13, 1996; revised September 24, 1996.

References

Allaire, M., Chernala, M.M., Malcolm, B.A., and James, M.N.G. (1994). Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases. Nature *369*, 72–76.

Bartenschlager, R., Ahlborn-Laake, L., Mous, J., and Jacobsen, H. (1994). Kinetic and structural analyses of hepatitis C virus polyprotein processing. J. Virol. *68*, 5045–5055.

Bartenschlager, R., Ahlborn-Laake, L., Yasargil, K., Mous, J., and Jacobsen, H. (1995). Substrate determinants for cleavage in *cis* and in *trans* by the hepatitis C virus NS3 proteinase. J. Virol. *69*, 198–205.

Bazan, J.F., and Fletterick, R.J. (1988). Viral cysteine proteases are homologous to the trypsin-like family of serine proteases: structural and functional implications. Proc. Natl. Acad. Sci. USA *85*, 7872–7876.

Brünger, A. (1992a). Free R-value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature *355*, 472–475.

Brünger, A. (1992b). XPLOR Version 3.1: A System for X-Ray Crystallography and NMR. (New Haven, Connecticut: Yale University).

Burley, S.K., and Petsko, G.A. (1986). Amino–aromatic interactions in proteins. FEBS Lett. *203*, 139–143.

Carson, M. (1991). Ribbons 2.0. J. Appl. Cryst. *24*, 958–961.

Choi, H.-K., Tong, L., Minor, W., Dumas, P., Boege, U., Rossman, M.G., and Wengler, G. (1991). Structure of Sindbis virus coat protein reveals a chymotrypsin-like serine proteinase and the organization of the virion. Nature *354*, 37–43.

Choo, Q.L., Kuo, G., Weiner, A.J., Overby, L.R., Bradley, D.W., and Houghton, M. (1989). Isolation of a cDNA clone derived from blood-borne non-A, non-B viral hepatitis. Science *244*, 359–362.

Choo, Q.L., Richman, K.H., Han, J.H., Berger, K., Lee, C., Dong, C., Gallegos, C., Coit, D., Medina-Selby, R., Barr, P.J., Weiner, A.J., Bradley, D.W., Kuo, G., and Houghton, M. (1991). Genetic organization and diversity of the hepatitis C virus. Proc. Natl. Acad. Sci. USA *88*, 2451–2455.

Connolly, M.L. (1983). Solvent-accessible surfaces of proteins and nucleic acids. Science *221*, 709–713.

D'Souza, E.D.A., Grace, K., Sangar, D.V., Rowlands, D.J., and Clarke, B.E. (1995). *In vitro* cleavage of hepatitis C virus polyprotein substrates by purified recombinant NS3 protease. J. Gen. Virol. *76*, 1729–1736.

Engh, R.A., and Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. Acta Cryst. *A47*, 392–400.

Failla, C., Tomei, L., and De Francesco, R. (1994). Both NS3 and NS4A are required for proteolytic processing of hepatitis C virus nonstructural proteins. J. Virol. *68*, 3753–3760.

Fujinaga, M., Delbaere, T.J., Brayer, G.D., and James, M.N.G. (1985). Refined structure of α-lytic protease at 1.7 Å resolution: analysis of hydrogen bonding and solvent structure. J. Mol. Biol. *183*, 479–502.

Furey, W., and Swaminathan, S. (1990). "PHASES"—a program package for the processing and analysis of diffraction data from macromolecules. ACA Meeting Summaries *73* (New York: American Crystallographic Association).

Grakoui, A., McCourt, D.W., Wychowski, C., Feinstone, S.M., and Rice, C.M. (1993). Characterization of the hepatitis C virus–encoded serine proteinase: determination of proteinase-dependent polyprotein cleavage sites. J. Virol. *67*, 2832–2843.

Han, D.S., Hahm, B., Rho, H.-M., and Jang, S.K. (1995). Identification of the protease domain in NS3 of hepatitis C virus. J. Gen. Virol. *76*, 985–993.

Hijikata, M., Mizushima, H., Akagi, T., Mori, S., Kakiuchi, N., Kato, N., Tanaka, T., Kimura, K., and Shimotohno, K. (1993). Two distinct proteinase activities required for the processing of a putative nonstructural precursor protein of hepatitis C virus. J. Virol. *67*, 4665–4675.

Houghton, M. (1996). Hepatitis C viruses. In Fields Virology, Third Edition, B.N. Fields, D.M. Knipe, and P.M. Howley, eds. (New York: Raven Press), pp. 1035–1058.

Ji, X., Zhang, P., Armstrong, R.N., and Gilliland, G.L. (1992). The three-dimensional structure of a glutathione S-transferase from the Mu class: structural analysis of the binary complex of isoenzyme 3-3 and glutathione at 2.2 Å resolution. Biochemistry *31*, 10169–10184.

Jones, T.A. (1978). A graphics model building and refinement system for macromolecules. J. Appl. Cryst. *11*, 268–272.

Kato, N., Hijikata, M., Ootsuyama, Y., Nakagawa, M., Ohkoshi, S., Sugimura, T., and Shimotohno, K. (1990). Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis. Proc. Natl. Acad. Sci. USA *87*, 9524–9528.

Koch, J.O., Lohman, V., Herian, U., and Bartenschlager, R. (1996). *In vitro* studies on the activation of the hepatitis C virus NS3 proteinase by the NS4A cofactor. Virology *221*, 54–66.

Laskowski, R.J., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J Appl. Cryst. *26*, 283–291.

Levitt, M., and Perutz, M.F. (1988). Aromatic rings as hydrogen bond acceptors. J. Mol. Biol. *201*, 751–754.

Lin, C. and Rice, C.R. (1995). The hepatitis C virus NS3 serine proteinase and NS4A cofactor: establishment of a cell-free trans-processing assay. Proc. Natl. Acad. Sci. USA *92*, 7622–7626.

Lin, C., Pragai, B.M., Grakoui, A., Xu, J., and Rice, C.M. (1994). Hepatitis C virus NS3 serine proteinase: trans-cleavage requirements and processing kinetics. J. Virol. *68*, 8147–8157.

Linnen, J., Wages, J., Jr., Zhang-Keck, Z.-Y., Fry, K.E., Krawczynski,

K.Z., Alter, H., Koonin, E., Gallagher, M., Alter, M., Hadziyannis, S., Karayiannis, P., Fung, K., Nakatsuji, Y., Shih, J.W.-K., Young, L., Piatak, M., Jr., Hoover, C., Fernandez, J., Chen, S., Zou, J.-C., Morris, T., Hyams, K.C., Ismay, S., Lifson, J.D., Hess, G., Foung, S.K.H., Thomas, H., Bradley, D., Margolis, H., and Kim, J.P. (1996). Molecular cloning and disease association of hepatitis G virus: a transfusion-transmissible agent. Science *271*, 505–508.

Liu, S., Ji, X., Gilliland, G.L., Stevens, W.J., and Armstrong, R.N. (1993). Second-sphere electrostatic effects in the active site of glutathione S-transferase: observation of an on-face hydrogen bond between the side chain of threonine 13 and the π-cloud of tyrosine 6 and its influence on catalysis. J. Am. Chem. Soc. *115*, 7910–7911.

Matthews, D.A., Smith, W.W., Ferre, R.A., Condon, B., Budahazi, G., Sisson, W., Villafranca, J.E., Janson, C.A., McElroy, H.E., Gribskov, C.L., and Worland, S. (1994). Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein. Cell *77*, 761–771.

McRee, D.E. (1992). XtalView: a visual protein crystallographic system for X11/Xview. J. Mol. Graph. *10*, 44–47.

Meyer, E., Cole, G., Radhakrishnan, R., and Pepp, O. (1988). Structure of native porcine pancreatic elastase at 1.65 Å resolution. Acta Cryst. *B44*, 26–38.

Miller, R.H., and Purcell, R.H. (1990). Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus supergroups. Proc. Natl. Acad. Sci. USA *87*, 2057–2061.

Otwinowski, Z., and Minor, W. (1996). Processing of X-ray diffraction data collected in oscillation mode. Meth. Enzymol. *276*, 307–326.

Pizzi, E., Tramontano, A., Tomei, L., La Monica, N., Failla, C., Sardana, M., Wood, T., and De Francesco, R. (1994). Molecular model of the specificity pocket of the hepatitis C virus protease: implications for substrate recognition. Proc. Natl. Acad. Sci. USA *91*, 888–892.

Read, R.J., and James, M.N.G. (1986). Introduction to the protein inhibitors: X-ray crystallography. In Proteinase Inhibitors, A.J. Barret and G. Salvesen, eds. (Amsterdam; New York; Oxford: Elsevier Science Publishers BV), pp. 301–336.

Schwabe, J.W.R., and Klug, A. (1994). Zinc mining for protein domains. Struct. Biol. *1*, 345–349.

Simons, J.N., Leary, T.P., Dawson, G.J., Pilot-Matias, T.J., Muerhoff, A.S., Schlauder, G.G., Desai, S.M., and Mushahwar, I.K. (1995). Isolation of novel virus-like sequences associated with human hepatitis. Nature Med. *1*, 564–569.

Simmonds, P., Smith, D.B., McOmish, F., Yap, P.L., Kolberg, J., Urdea, M.S., and Holmes, E.C. (1994). Identification of genotypes of hepatitis C virus by sequence comparisons in the core, E1 and NS5 regions. J. Gen. Virol. *75*, 1053–1061.

Sommergruber, W., Casari, G., Fessl, F., Seipelt, J., and Skern, T. (1994). The 2A proteinase of human rhinovirus is a zinc containing enzyme. Virology *204*, 815–818.

Steinkühler, C., Tomei, L., and De Francesco, R. (1996). *In vitro* activity of hepatitis C virus protease NS3 purified from recombinant baculovirus–infected SF9 cells. J. Biol. Chem. *271*, 6367–6373.

Suzich, J.A., Tamura, J.K., Palmer-Hill, F., Warrener, P., Grakoui, A., Rice, C.M., Feinstone, S.M., and Collett, M.S. (1993). Hepatitis C virus NS3 protein polynucleotide–stimulated nucleoside triphosphatase and comparison with related pestivirus and flavivirus enzymes. J. Virol. *67*, 6152–6158.

Takamizawa, A., Mori, C., Fuke, I., Manabe, S., Murakami, S., Fujita, J., Onishi, E., Andoh, T., Yoshida, I., and Okayama, H. (1991). Structure and organization of the hepatitis C virus genome isolated from human carriers. J. Virol. *65*, 1105–1113.

Terwilliger, T.C., Kim, S.-H., and Eisenberg, D. (1987). Generalized method of determining heavy-atom positions using the difference Patterson function. Acta Cryst. *A43*, 1–5.

Vallee, B.L., and Auld, D.S. (1990). Zinc coordination, function, and structure of zinc enzymes and other proteins. Biochemistry *29*, 5647–5659.

van Doorn, L.J. (1994). Review: molecular biology of the hepatitis C virus. J. Med. Virol. *43*, 345–356.

Voss, T., Meyer, R., and Sommergruber, W. (1995). Spectroscopic characterization of rhinoviral protease 2A: Zn is essential for the structural integrity. Protein Sci. *4*, 2526–2531.

Wang, B.C. (1985). Resolution of phase ambiguity in macromolecular crystallography. Meth. Enzymol. *115*, 90–112.

Yu, S.F., and Lloyd, R.E. (1992). Characterization of the roles of conserved cysteine and histidine residues in poliovirus 2A protease. Virology *186*, 725–735.