Molecular views of viral polyprotein processing revealed by the crystal structure of the hepatitis C virus bifunctional protease–helicase

Nanhua Yao, Paul Reichert, S Shane Taremi, Winifred W Prosise and Patricia C Weber*

Background: Hepatitis C virus (HCV) currently infects approximately 3% of the world's population. HCV RNA is translated into a polyprotein that during maturation is cleaved into functional components. One component, nonstructural protein 3 (NS3), is a 631-residue bifunctional enzyme with protease and helicase activities. The NS3 serine protease processes the HCV polyprotein by both *cis* and *trans* mechanisms. The structural aspects of *cis* processing, the autoproteolysis step whereby the protease releases itself from the polyprotein, have not been characterized. The structural basis for inclusion of protease and helicase activities in a single polypeptide is also unknown.

Results: We report here the 2.5 Å resolution structure of an engineered molecule containing the complete NS3 sequence and the protease activation domain of nonstructural protein 4A (NS4A) in a single polypeptide chain (single chain or scNS3–NS4A). In the molecule, the helicase and protease domains are segregated and connected by a single strand. The helicase nucleoside triphosphate and RNA interaction sites are exposed to solvent. The protease active site of scNS3–NS4A is occupied by the NS3 C terminus, which is part of the helicase domain. Thus, the intramolecular complex shows one product of NS3-mediated cleavage at the NS3–NS4A junction of the HCV polyprotein bound at the protease active site.

Conclusions: The scNS3–NS4A structure provides the first atomic view of polyprotein *cis* processing. Both local and global structural rearrangements follow the *cis* cleavage reaction, and large segments of the polyprotein can be folded prior to proteolytic processing. That the product complex of the *cis* cleavage reaction exists in a stable molecular conformation suggests autoinhibition and substrate-induced activation mechanisms for regulation of NS3 protease activity.

Address: Structural Chemistry Department, Schering-Plough Research Institute, 2015 Galloping Hill Road, K15-3-3855, Kenilworth, NJ 07033-0539, USA.

*Corresponding author.
E-mail: patricia.weber@spcorp.com

Key words: bifunctional enzyme, HCV NS3, polyprotein processing, protease, RNA helicase

Received: 29 July 1999
Revisions requested: 17 August 1999
Revisions received: 23 August 1999
Accepted: 27 August 1999

Published: 27 October 1999

Structure November 1999, 7:1353–1363

0969-2126/99/$ – see front matter

Introduction

Hepatitis C virus (HCV), a (+)-strand RNA virus of the *Flaviviridae* family, is the major etiological agent of post-transfusion non-A, non-B hepatitis [1]. In infected cells, the HCV RNA is directly translated into a continuous 3011-residue polypeptide chain [2–4]. The polyprotein is subsequently cleaved to yield envelope and core proteins, which assemble into new virus particles, and enzymes essential for viral replication [5–7]. At least ten proteins are encoded by the HCV genome and significant progress has been made in dissecting their biological functions in the viral life-cycle (for recent reviews, see [8–10]).

Both host cell and virally encoded proteases are required for maturation of the HCV polyprotein. Initial cleavages catalyzed by host proteases liberate individual envelope and capsid proteins, located at the polyprotein N terminus, and the portion of the HCV polyprotein that contains the viral enzymes, or nonstructural (NS) proteins. This polyprotein segment contains the free N terminus of nonstructural protein 2 (NS2) — an integral membrane protein with proteolytic activity [11,12]. NS2, either alone or in conjunction with NS3 [13,14], then cleaves the NS2–NS3 scissile bond in what is termed *cis* processing, because cleavage involves an intramolecular autoproteolysis step. The NS2-mediated autoproteolysis generates the NS3 N terminus, and releases a 1984-residue polyprotein that incorporates a bifunctional 631-residue enzyme (NS3) possessing both serine protease and RNA helicase activities. The NS3 protease is responsible for processing the remaining polyprotein [7,15–19].

Completion of polyprotein maturation is initiated by autocatalytic cleavage at the NS3–NS4A junction. The reaction is catalyzed by the NS3 serine protease and occurs as a *cis* processing step. Subsequent NS3-mediated cleavages

of the HCV polyprotein are thought to involve the recognition of polyprotein cleavage junctions by an NS3 molecule of another polypeptide and are termed *trans* processing events. In these reactions, NS3 liberates an NS3 cofactor (NS4A), two proteins of unassigned function (NS4B and NS5A), and an RNA-dependent RNA polymerase (NS5B) [20].

Many questions concerning the structural events that occur during polyprotein processing are unanswered. The *cis* cleavage reaction by which NS3 releases itself from the HCV polyprotein requires the polyprotein segment containing the NS3–NS4A junction to bind at the NS3 protease active site. The C terminus of the NS3 helicase domain and the N terminus of NS4A reside on either side of the scissile bond. Previously, only structures of the isolated helicase domain [21] and the protease domain (both free and in complex with a peptide containing the region of NS4A known to activate the NS3 protease [22–24]) were available. Modeling of the *cis* cleavage reaction was complicated because the relative orientations of the protease and helicase in the native bifunctional NS3 molecule were unknown. In addition, structures of the helicase domain showed a well-defined helix at the C terminus, which implied that either some degree of helix unfolding occurs on formation of a substrate-recognition complex typical of chymotrypsin-like serine proteases, or that the HCV NS3 protease forms a novel substrate-recognition complex. Clearly structural studies of full-length NS3 are needed to reveal structural aspects of the *cis* and *trans* cleavage reactions.

Studies of this molecule were also undertaken to extend the structural biology of HCV proteins towards the completeness achieved for human immunodeficiency virus (HIV) proteins [25,26]. The NS3–NS4A complex contains 685 residues, comprising nearly 23% of the HCV genome and 35% of the nonstructural region. A molecular view of single-chain (sc) NS3–NS4A would aid understanding of the structural basis for the observed enhancement of helicase activity in NS3 relative to the isolated helicase domain. In addition, several engineered forms of the NS3–NS4A complex, which are expressed at high levels and are more soluble and stable *in vitro* than the native enzyme, have been reported [27–29]. While the engineered molecules exhibit protease and helicase activities similar to those of the native complex, crystallographic analyses were undertaken to validate their use in structure-based design of protease and helicase inhibitors.

Results

Overall structure

The structure of scNS3–NS4A was solved by molecular replacement using the protease [22,24] and helicase [21] domain structures as probe molecules. The structure was refined to 2.5 Å resolution with a final R factor of 20% (R_{free} 26%). The cross-validated coordinate error is estimated to be 0.22 Å. Data collection and refinement statistics are given in Table 1. In the electron-density maps, the entire 647-residue polypeptide of both molecules in the crystallographic asymmetric unit was well-defined with few exceptions. The molecules were refined without symmetry constraints and superimpose with a Cα coordinate root mean square deviation (rmsd) of 0.6 Å. A Cα trace of scNS3–NS4A is shown in Figure 1. The molecule is globular in shape with approximate dimensions of 65 Å × 65 Å × 75 Å.

Table 1

Data collection and refinement statistics.

Data collection	
Resolution range (Å)	40.0–2.5
R_{sym}*	0.073
Unique reflections (F ≥ 0)	50,087
Theoretically possible reflections (%)	99.7
Refinement	
R factor†	0.20
R_{free} (1003 reflections)	0.26
Rms deviation from ideal bond distances (Å)	0.013
Rms deviation from ideal angle (°)	1.67
Protein nonhydrogen atoms	9614
Zinc ions	2
Phosphate ions	2
Ordered water molecules	262

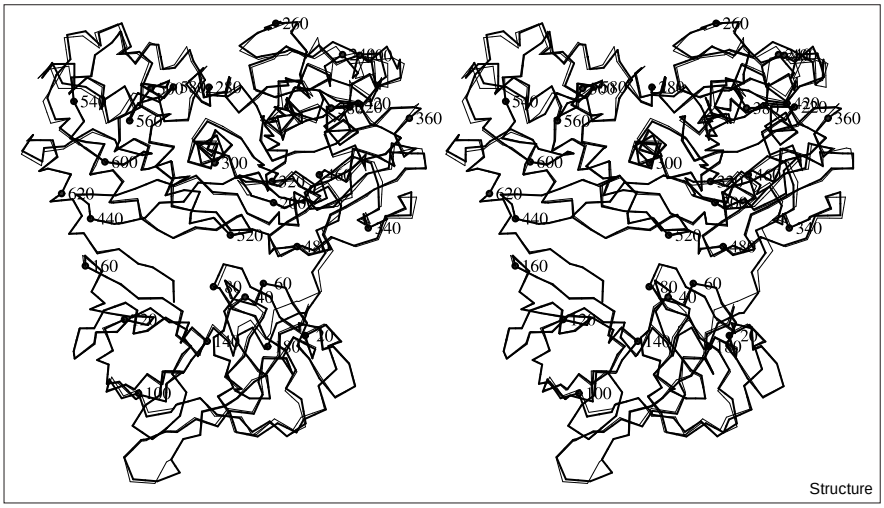
*R_{sym} calculated for 601,477 reflections where F > 0. †R factor calculated using 48,937 reflections for which F > 0.

scNS3–NS4A is composed of six subdomains which occur in sequence along the polypeptide chain (Figure 2). The protease domain exhibits the dual β-barrel fold that is common among members of the chymotrypsin serine protease family. The helicase domain contains two structurally related β-α-β subdomains and a third subdomain of seven helices and three short β strands. The latter domain is usually referred to as the helicase α-helical subdomain. The 13-residue protease activation domain of NS4A contributes one strand to the N-terminal protease β barrel and is considered to be the sixth subdomain. The scNS3–NS4A molecule is colored by subdomain in the schematic representation in Figure 2a and the amino acid sequence is given in Figure 2c.

Differences in subdomain structure in the scNS3–NS4A molecule and in the structures of the isolated protease and helicase domains were assessed in several ways. Inspection of the molecule revealed that the subdomain folds are similar. Overall preservation of structure is also apparent when the subdomains from the various structures are superposed. As detailed in Table 2, the subdomains exhibit rmsd values in Cα positions averaging ~0.5 Å. Only the NS3 C terminus dramatically changes fold in the two structures, and this change results in

Figure 1

Stereoview of scNS3–NS4A represented as a Cα trace. The two crystallographically independent molecules are shown superposed and are represented by thin and thick lines, respectively. Every twentieth residue is labeled.



slightly higher deviations on superposition of the helicase α-helical subdomains. Comparison of the relative B values (Figure 3) also shows a similar distribution of structural flexibility for scNS3–NS4A and the isolated domains.

Figure 2

The structure of scNS3–NS4A. (a) Ribbon diagram of scNS3–NS4A, an engineered molecule that consists of 631 NS3 residues, 14 NS4A residues known to activate the NS3 protease, and three linker residues. In scNS3–NS4A the NS4A sequence is linked to the NS3 N terminus. In the hepatitis C virus, NS3 and 54-residue NS4A exist as a tightly associated noncovalent complex. scNS3–NS4A has six subdomains: two β barrels in the protease domain (magenta and red); two β-α-β subdomains (green and yellow) and one α-helical subdomain in the helicase (blue) and the NS4A-derived sequence (cyan). Residues of the protease catalytic triad (His57, Asp99 and Ser139) are shown in ball-and-stick representation, and the protease structural zinc ion is shown as a white sphere. The red spheres represent a phosphate molecule located at the NTP-binding site of the helicase. This view clearly shows the interaction between the C terminus (blue) and protease active-site region. (b) The electrostatic surface potential of NS3 viewed in the same molecular orientation as (a). Residues contributing to the region of excess positive charge (dark blue surface) near the center of the protease are labeled. Regions of excess negative charge (red surface) are a recurrent and poorly understood feature of helicase structure. (c) The amino acid sequences of NS3 and NS4A in one-letter code, colored by subdomain as in (a). Junctions in the HCV polyprotein that are processed by NS3 are aligned below with the NS3–NS4A autoprocessing site. Yellow bars denote the conserved P1 and P6 residues.

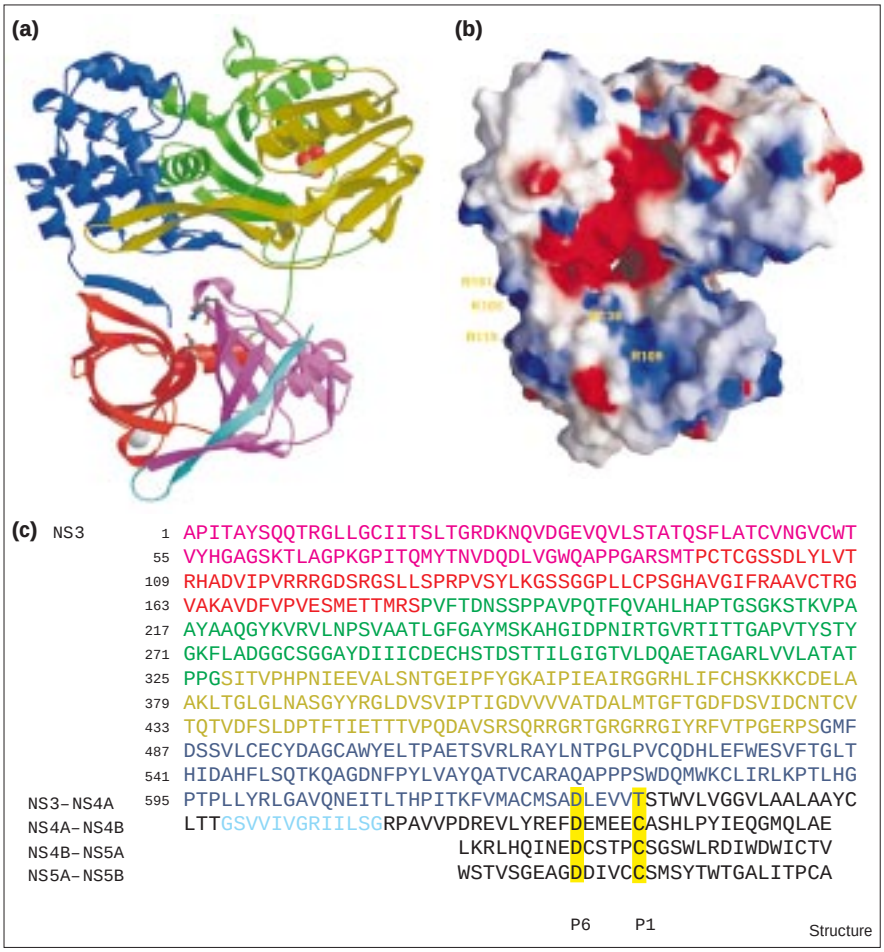


Table 2

Comparison of subdomain structure.

Domain	Protease		Helicase			NS4A
Subdomain	1	2	3	4	5	6
Rmsd (Å)	0.5	0.5	0.5	0.5	0.6	0.3
Number of Cα atoms	89	80	141	130	139	11

Subdomains of scNS3–NS4A were superimposed on those of the isolated protease–NS4A [22] and helicase [21] structures.

However, the B value comparison does reveal that several loops in the helicase and the NS3 C terminus are more ordered in scNS3–NS4A.

The protease and helicase catalytic centers are segregated in the bifunctional enzyme (Figure 2a). The nucleoside triphosphate (NTP) and the single-stranded RNA-binding sites, as well as the deep surface grooves that define the helicase active-site region, are oriented away from the interdomain interface and are exposed to solvent. In contrast, the NS3 protease active site is oriented towards the molecular interior. The P-side (as defined by Schechter and Berger [30]) of the substrate-recognition site and the active-site region are occupied by the molecule's own C terminus. The NS3 C terminus is the P-side product of the *cis* cleavage reaction at the NS3–NS4A junction and, following proteolysis, it carries the C-terminal carboxylate. This is in contrast to the P′-side (or primed side) segment of the substrate, which following *cis* processing at the NS3–NS4A junction contains the N-terminal amino group of NS4A. With the P-side product bound at the protease active site, the scNS3–NS4A structure provides the first molecular view of autoproteolytic polyprotein processing.

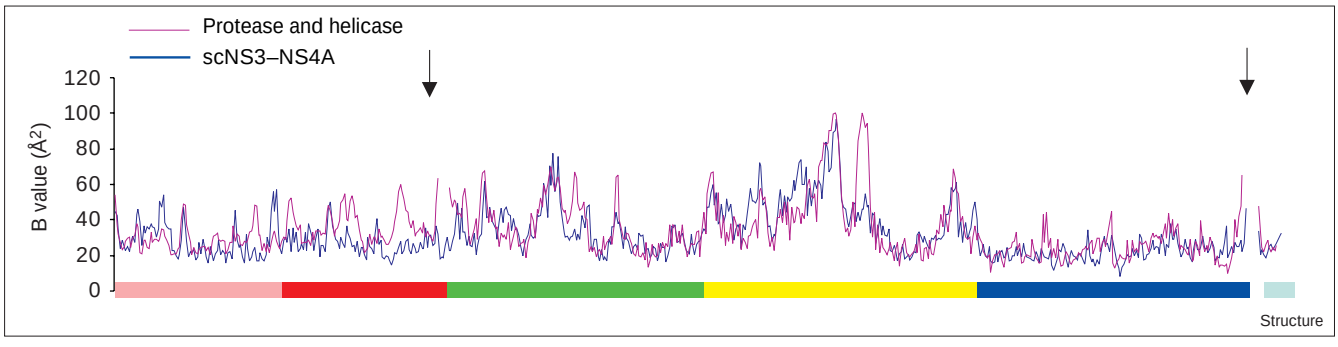
Interactions between the protease and helicase domains

The protease and helicase domains are covalently connected by a short solvent-exposed strand incorporating polypeptide residues Pro182, Val183, Phe184, Thr185 and Asp186. The connecting strand appears flexible and adopts a slightly different conformation in each independent copy of the scNS3–NS4A molecule that comprises the crystallographic asymmetric unit (Figure 1). The principal surface interaction between the domains buries ~900 Å^2 at the domain interface and involves residues Tyr56, Gly60, Ser61, Val78 and Asp79 from the protease domain and Gly327, Val329, Pro482, Ser483, Gly484, Met485, Pro520, Gly521 and Val524 from the helicase domain.

Extensive interactions between the protease and helicase domains occur around the protease active site. Electron density for this region of the molecule (Figure 4a) is clear and allows the precise definition of the intramolecular interactions. The NS3 C-terminal sequence comprising residues 626–631 forms an extended antiparallel β strand along the edge of the protease β barrel, as well as numerous additional interactions between the C-terminal carboxylate and protease catalytic residues. The latter include hydrogen-bond interactions with the His57 imidazole and Ser139 hydroxyl of the catalytic triad, and with the backbone amides of residues 137 and 139 that stabilize the oxyanion transition state. The six C-terminal residues of NS3 form 12 hydrogen bonds and create an ~500 Å^2 contact surface near the NS3 protease active site.

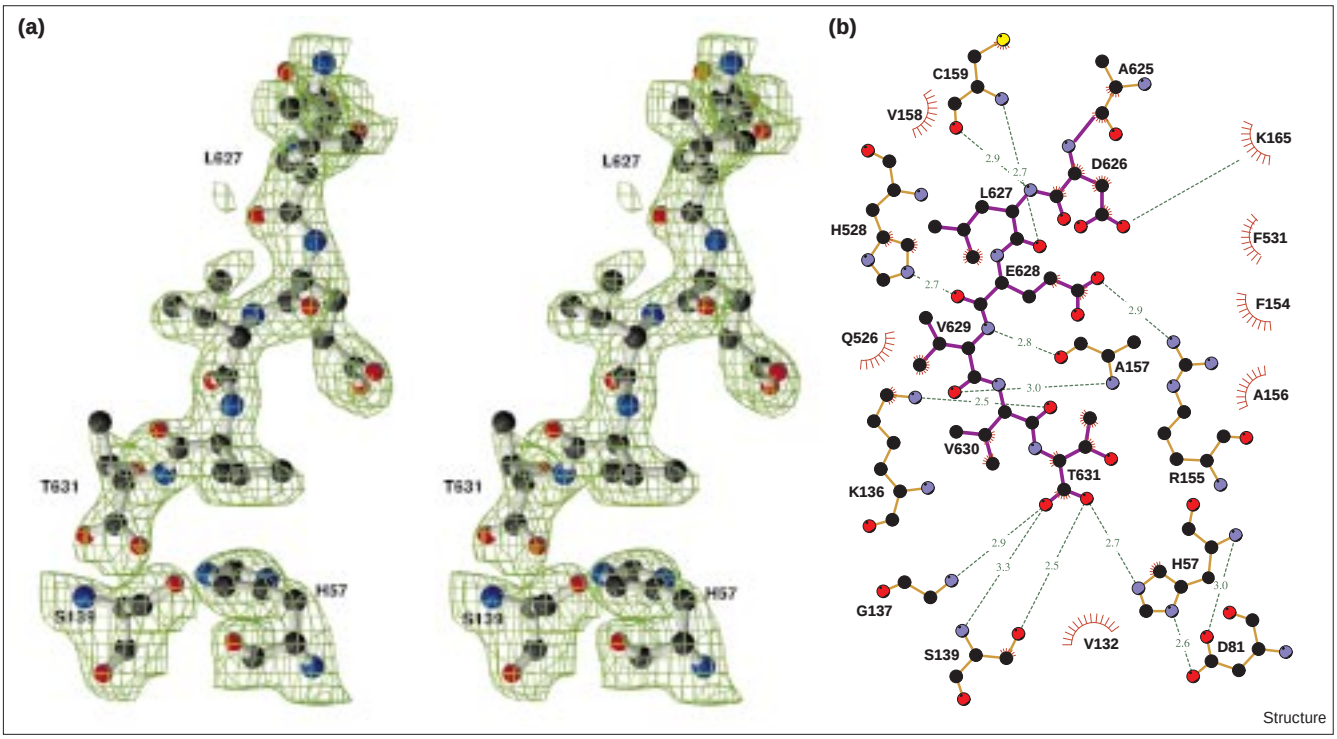
Details of the intramolecular backbone and sidechain interactions are shown in Figure 4b. The interaction patterns along the C-terminal strand and with the terminal carboxylate are similar to those observed in other protease–product peptide complexes. These include the glutamic-acid-specific protease complex with a tetrapeptide

Figure 3



Comparison of mainchain B values. The blue line shows values for scNS3–NS4A, and the magenta line shows values reported for the isolated protease [22] and helicase [21] domains. The bar on the horizontal axis represents the linear sequence and is color coded as in Figures 2a and 2c. Arrows indicate the N and C termini of the isolated domain structures. In scNS3–NS4A, the NS4A sequence (cyan) is at the N terminus of the molecule. In the HCV polyprotein, the NS4A sequence follows NS3; after processing NS4A is noncovalently associated with NS3.

Figure 4



Interactions of the NS3 C terminus at the NS3 protease active site. (a) Stereoview $|2F_o–F_c|$ electron density contoured at 1σ for residues 626–631, and His57 and Ser139 of the catalytic triad. (b) Schematic diagram of the intramolecular backbone and sidechain interactions. Hydrogen bonds are shown as dashed lines and distances are given in Å. The distance between the salt-linked atoms of Lys165 and Asp626 is 3.6 Å.

ligand [31] (PDB code 1HPG) and the Sindbis virus protease–peptide complex [32] (PDB code 2SNV). Clearly, the NS3 C terminus binds in what is evidently a product complex of NS3-mediated *cis* cleavage at the NS3–NS4A junction of the HCV polyprotein.

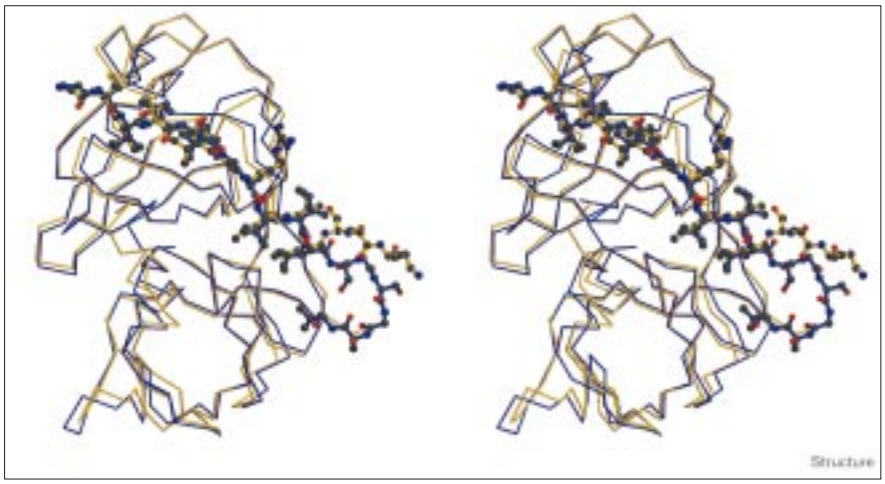
It is notable that when the proteolysis of peptides corresponding to the polyprotein processing sites is studied *in vitro*, inhibition by the P-side products is observed. For example, while K_m values of 10 μM and 3.8 μM are observed for the 12-residue peptide substrates corresponding to the NS4A–NS4B and NS5A–NS5B junctions, the P-side hexameric products exhibit K_i values of 0.6 μM and 1.4 μM, respectively [33]. These observations are consistent with the present structural results that show the P-side product of *cis* cleavage at the NS3–NS4A junction bound in the protease active site.

Engineered scNS3–NS4A

The NS3 protease is activated by the 54-residue NS4A protein. Crystallographic studies of the protease domain have shown that prior to binding NS4A, the NS3 protease is partially unfolded [23,34]. Protease activation involves completion of the N-terminal protease β barrel by NS4A, the ordering of ~30 N-terminal NS3 residues, and assembly of the NS3 serine protease catalytic triad. Only the central portion of NS4A is required for protease activation. The functional role of the NS4A C terminus is unknown, but the hydrophobic N terminus is known to localize the NS3–NS4A complex at the host cell membrane.

Using the crystal structure of the NS3 protease domain in complex with the NS4A-derived activation peptide, several groups constructed molecules incorporating both proteins into a single polypeptide chain [27–29]. In general, the single chain NS3–NS4A molecules exhibit activities comparable to the native NS3–NS4A complex. Preservation of activity in the single-chain molecules is consistent with comparison of the scNS3–NS4A structure to that of the NS3 protease domain in complex with an NS4A-derived peptide. Superpositions (Figure 5; Table 2) and B value comparisons (Figure 3) revealed no significant differences in the protease or NS4A-activation peptide structures. These results are similar to those observed for other two-chain proteases such as HIV-1 protease [35,36], pepsinogen [37] and coagulation factor VIIIa [38], where protein engineering methods proved effective for the production of soluble, stable proteases for mechanistic and biophysical studies.

Figure 5



Preservation of the NS3 protease domain and NS4A activation sequence structures in engineered scNS3–NS4A. Stereoview superposition of the protease portion of scNS3–NS4A (blue) on the NS3 protease domain–NS4A-derived peptide complex (yellow, PDB code 1A1R [37]). Residues of NS4A, the linker Ser-Gly-Ser sequence, and NS3 Ile3 are shown in ball-and-stick representation.

Discussion

The C terminus of scNS3–NS4A is the P-side product of the NS3-mediated *cis* cleavage at the NS3–NS4A junction of the HCV polyprotein. In the structure, the P-side product is bound at the protease active site. Thus, the crystallographic study reported here provides the first molecular snapshot of polyprotein processing by autolysis. This unique structural result advances our understanding of HCV maturation and viral polyprotein processing in general.

Several features of the NS3 molecular mechanism can be extended from this work. First, it is apparent that during the *cis* and *trans* cleavage reactions, both local and global conformational changes occur in the NS3 structure. Second, the completeness of the structure allows construction of a model of NS3 in complex with the NS3–NS4A junction prior to the *cis* cleavage reaction. The structure and model suggest that the individual proteins can be substantially folded in the polyprotein. Finally, the results afford a structural basis for the observed enhancement of helicase activity in the bifunctional enzyme relative to the isolated domain.

The *cis* cleavage reaction

The structure of the polyprotein substrate complex prior to *cis* cleavage at the NS3–NS4A junction can be readily modeled from the product complex structure reported here. This requires the addition of residues 1–20 of NS4A. In the substrate complex, these residues are situated between the C terminus of NS3 bound at the protease active site and residue 21 of the NS4A segment that completes the protease β barrel. A 20-residue sequence can easily span the ~20 Å distance between crystallographically defined residues, and many favorable backbone conformations can be modeled. In fact, the distance is sufficiently short that a small α-helical segment can be accommodated. Thus the model suggests that a number of hydrophobic residues, potentially able to form a helical membrane anchor, can adopt a defined helical secondary structure prior to proteolytic processing at the NS3–NS4A junction (Figure 6a).

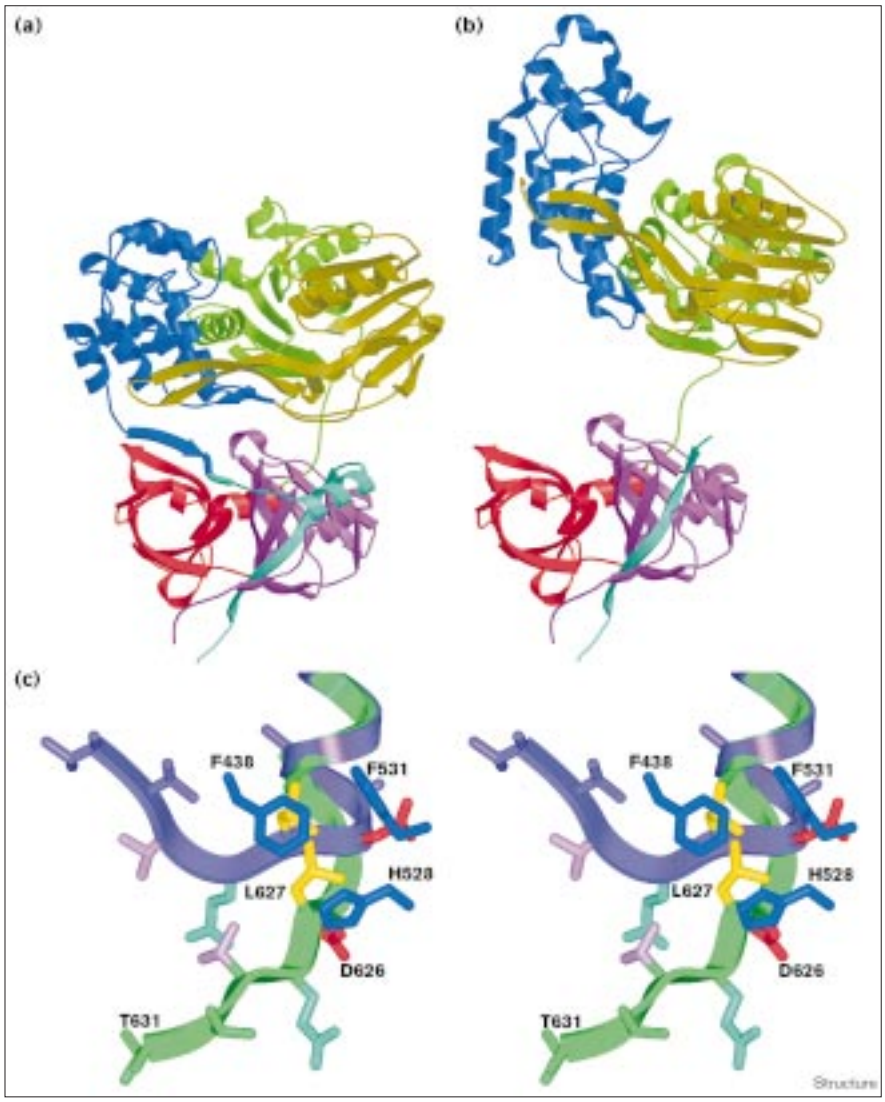
The *trans* cleavage reaction

The binding and catalytic cleavage of successive polyprotein recognition sequences requires the displacement of the product NS3 C terminus from the protease active site. In the structure of the isolated helicase domain [21,39,40], the C terminus forms an α helix ending at residue 627 followed by an extended tetrapeptide sequence. In the structure of bifunctional NS3 reported here, the helix is shorter (Figure 6c) and the C terminus is more extended in order to form the extensive β-sheet interactions seen in the product complex and presumably required for formation of the substrate complex.

On the basis of the scNS3–NS4A structure, exposure of the NS3 protease site to polyprotein substrates involves at least two types of conformational change. First, when the C terminus of NS3 adopts the conformation observed in the structures of the isolated helicase domain, the protease active site becomes more exposed. However, this local change leaves the substrate-recognition subsites partially blocked by the helicase. Additional global conformational changes appear necessary to accommodate a polyprotein substrate. More complete exposure of the substrate-recognition subsites is achieved by relative rotations of the protease and helicase domains (Figure 6b). Rigid-body rotations are readily accomplished by minor changes in backbone torsional angles within the interdomain strand.

The protease activities towards polypeptide substrates are similar for the native NS3–NS4A and scNS3–NS4A enzymes, and for the isolated protease domain [19,41]. This suggests that relatively little energy is required to disrupt

Figure 6



Structural changes in NS3 during *cis* and *trans* processing events. Models of (a) the NS3–NS4A complex prior to *cis* cleavage and (b) the NS3 conformation capable of accommodating HCV polyprotein processing sites in *trans*. Color coding corresponds to that in Figure 2. (c) Stereoview superposition of the NS3 C terminus (green) and the isolated helicase domain (blue). Sidechains with markedly different locations in the two structures have the same color (Asp626, red; Leu627, yellow; Glu628, cyan; and Val629, pink). Val630 and Thr631 are well-defined only in the scNS3–NS4A crystal structure, and for completeness are modeled here. Structural changes in the backbone conformation are accommodated by preservation of an important sidechain–sidechain contact involving a relatively nondirectional interaction between the Leu627 sidechain and an aromatic cluster containing Phe438, Phe531 and His528.

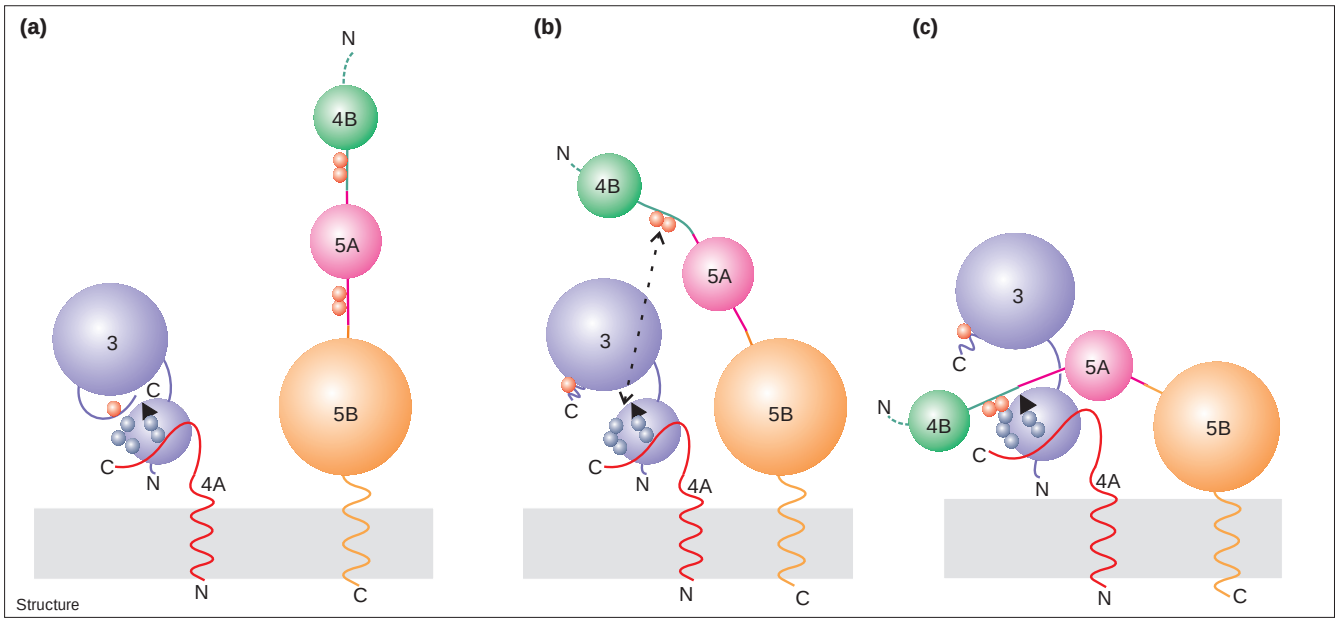
the *cis* product complex, possibly because the NS3 C terminus can readily adopt the alternative low-energy α-helical conformation seen in the isolated helicase structure. Some structural alterations in the protease domain may occur during this process and account for differences in P1 specificity for the *cis* and *trans* cleavage reactions (Figure 2c).

The polyprotein substrate itself may promote the structural changes required for the *trans* cleavage reaction. All cleavage sites in the HCV polyprotein contain a conserved acidic residue at P6 and several contain additional acidic residues (Figure 2c). In the product complex, the P6 acidic sidechain of Asp626 forms a weak, solvent-exposed salt linkage with the amino group of Lys165. Although it is unclear whether this interaction is energetically important, it is possible that the conserved P6 acidic group serves another more generalized function. For example, the acidic P6 residue may interact with the positively charged region near the protease active site to electrostatically guide protease–substrate complex formation (Figure 7). Several basic residues, Arg109, Arg119, Lys136, Arg161 and Lys165, are conserved in this region (Figure 2b). Similar electrostatic recognition has been observed for the substrates of proteases such as thrombin [42].

Polyprotein processing

The scNS3–NS4A structure suggests a plausible mechanism for polyprotein processing and membrane localization of HCV enzymes. This process is outlined in Figure 8. Figure 8a shows the 1984-residue nonstructural polyprotein anchored to the host cell membrane through a C-terminal helix. Figure 8b shows a self-associated state of the polyprotein where the NS4A-activating sequence has complexed with the protease domain and the NS3–NS4A cleavage

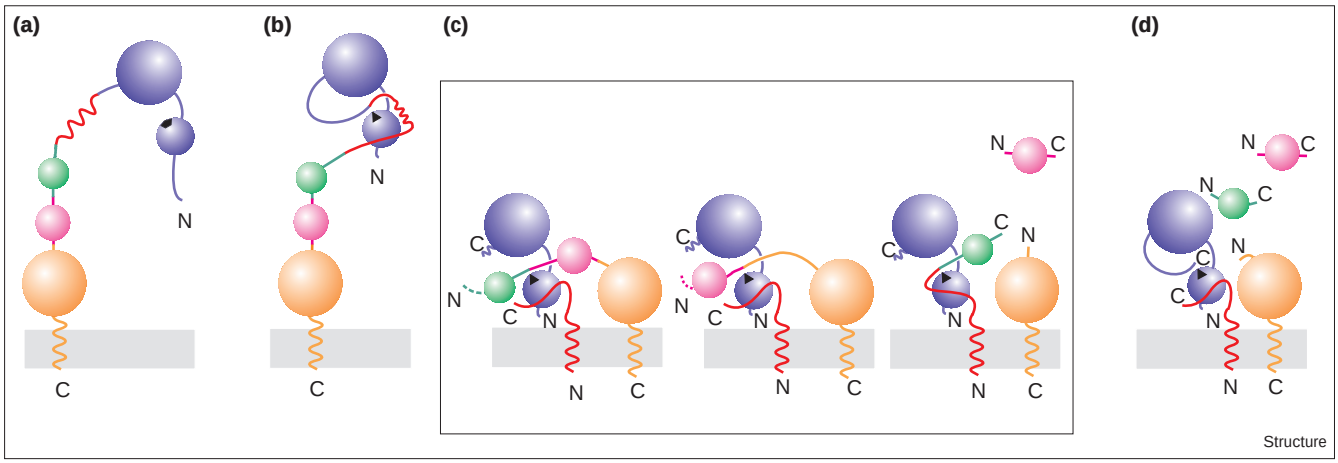
Figure 7



Activation of NS3–NS4A. (a) Schematic of the NS3–NS4A complex following *cis* cleavage at the NS3–NS4A junction, and a segment of the nonstructural region of the HCV polyprotein. Both fragments are attached to the host cell membrane via hydrophobic helices. Characters 3, 4A, 4B, 5A and 5B denote the nonstructural proteins NS3, NS4A, NS4B, NS5A and NS5B, respectively. The N and C termini are labeled. Small red and blue spheres indicate negatively charged residues on the polyprotein substrate, and positively charged residues on the NS3 protease domain, respectively. A black triangle indicates the NS3 protease active site. (b) Electrostatic recognition between NS3–NS4A and the polyprotein substrate, indicated by a dashed line. (c) Cleavage at the NS5A–NS5B junction. Local and global structural changes in NS3 to accommodate the polyprotein substrate are depicted in (b) and (c).

sequence has bound to the protease active site. Crystallographic studies of the protease domain have shown that prior to binding the activating cofactor NS4A, the NS3 protease is partially unfolded [23,34]. Protease activation

Figure 8



HCV polyprotein processing in the nonstructural region. Nonstructural proteins NS3, NS4A, NS4B, NS5A and NS5B are colored purple, red, green, pink and orange, respectively. (a) Attachment of the 1984-residue polyprotein to the membrane. (b) NS4A activation and folding of the NS3 N terminus. (c) Subsequent cleavage reactions. To highlight the fact that the sequence of cleavage reactions has not been firmly established, the N terminus of the polyprotein substrate is dotted and the schematic diagrams are enclosed in a box. (d) The release of NS4B and NS5A and formation of the replication complex core.

involves completion of the protease β barrel by NS4A and ordering of the protease domain N terminus.

Information about the sequential order of polyprotein processing events within the HCV polyprotein encompassing NS3 through NS5B suggests that *cis* cleavage at the NS3–NS4A junction occurs very rapidly [20]. Interestingly, the NS5A and NS5B proteins appeared before NS4B and indicate that subsequent cleavages may not be sequentially ordered along the polyprotein; the latter processing events occur in *trans* (Figure 8c). Modeling studies suggest that *cis* cleavage at the NS4A–NS4B junction can be accommodated even if NS4A serves both as the NS3 activator and carries the NS4A–NS4B junction (Figure 8c). Polyprotein processing finally results in the liberation of NS4B, NS5A and NS5B; NS5B along with NS3 constitutes the core of the HCV membrane-associated replication complex, (Figure 8d) [43]. In the replication complex, when the proteolytic role of NS3 is presumably complete, the NS3 structure could revert to the conformation observed in the crystal structure, where a compact globular shape is stabilized by interactions between the C terminus and the protease active-site region.

Bifunctional protease–helicase

HCV is a member of the *Flavivirus* family, all members of which possess an NS3 analog with dual protease–helicase functions [44]. Many studies have demonstrated that the catalytic efficiency of the protease measured using *in vitro* model systems is similar in the isolated domains and the full-length NS3 bifunctional enzyme [19,41]. In contrast, helicase activity is enhanced in the full-length NS3 enzyme relative to the isolated helicase domain [29,41]. Although the overall helicase fold is preserved in the two structures, both β-α-β domains are more completely folded in the bifunctional enzyme (Figure 3). The electron density is clear for the entire scNS3–NS4A molecule, whereas in several helicase domain structures residues at the N terminus are disordered [21,39,40]. Stabilization of the helicase fold in the full-length enzyme may contribute to more effective utilization of the energy arising from NTP hydrolysis. Additionally, it is known that the NS3 protease domain contributes binding sites for the helicase RNA substrate [41,45,46]. As shown in Figure 2b, several positively charged regions suitable for electrostatic binding of RNA feature prominently on the protease domain surface.

In the native bifunctional enzyme, the protease domain apparently acts as a localization factor to spatially direct helicase activity. Association of NS3–NS4A with the host membrane colocalizes the protease and helicase at the membrane surface as part of a larger replication complex with the HCV RNA-dependent RNA polymerase [43] (Figure 8d). A similar localization mechanism for helicase regulation is employed by the spliceosome, where at least seven discrete RNA unwinding activities are involved in intron processing [47]. The conserved helicase domains reside within larger, multidomain splicing factors which form specific intermolecular interactions to appropriately localize the various helicases.

Biological implications

Hepatitis C virus (HCV) is the major etiological agent of post-transfusion non-A, non-B hepatitis. Chronic HCV infections cause liver cirrhosis and, in a significant number of cases, progress to hepatocellular carcinoma [48]. HCV currently infects approximately 3% of the world's population and has been termed the silent epidemic because of its widespread incidence and the long, 10–15 year, latency period between the initial detection of virus in patients and the onset of clinical symptoms [49]. In infected cells, the HCV RNA is translated into a continuous polypeptide chain that is subsequently cleaved to produce envelope proteins, core proteins and nonstructural (NS) proteins. A more complete understanding of viral polyprotein processing may aid the discovery of new therapeutic agents to control or eliminate HCV infections.

We report here the crystal structure of the HCV nonstructural protein 3 (NS3) — a bifunctional protease–helicase. The structure reveals a compact globular molecule stabilized by interactions between the C terminus and the protease active-site region. Given that HCV polyprotein processing involves NS3-mediated autoproteolysis at the NS3 C terminus, this structure represents the complex between the protease and P-side product of the *cis* cleavage reaction, and thus provides the first structural view of *cis* polyprotein processing. The structure, along with previously determined structures of the isolated protease and helicase domains, indicates the extent to which local and global structural rearrangements are required for subsequent NS3-mediated polyprotein processing.

Materials and methods

Protein expression and purification

A recombinant single-chain construct in which the NS3-activating sequence of NS4A (residues 21–32) precedes the N terminus of NS3 (residues 3–631) was cloned and expressed in *Escherichia coli* as previously described [28,29]. Throughout the paper, this molecule is referred to as scNS3–NS4A.

The cell pellet from a 10 L fermentor was resuspended in 600 ml of lysis buffer containing 25 mM HEPES pH 7.5, 10% (v/v) glycerol, 0.3 M NaCl, 0.1% n-octyl-β-D-octyl glucoside, 2 mM β-mercaptoethanol (βME), 4 mM $MgCl_2$ and 19,000 units/L DNAse. Twenty tablets of EDTA-free complete protease inhibitor cocktail (Boehringer Mannheim) were added per liter of lysis buffer. The mixture was treated for 2 min with a cell homogenizer (Omni Mixer ES) and cells were subsequently disrupted by two passes through a Microfluidizer (Model M-110F, Microfluidics, MA) operated at 10,000 psi The lysate was clarified by centrifugation at 85,000 × g for 60 min. The supernatant was applied at 120 cm/h to a 10 ml Ni-NTA (Qiagen) column in the presence of 20 mM imidazole using a gradifrac system (Pharmacia). The column was washed with 15–20 column volumes of the lysis buffer. The bound scNS3–NS4A was eluted with ten column volumes of lysis buffer supplemented with 250 mM imidazole.

The pooled fractions were further purified by size-exclusion chromatography using three sephacryl-100 size-exclusion columns (26 × 60 cm, Pharmacia) in series (flow rate 0.5 ml/min). Columns were pre-equilibrated in buffer containing 25 HEPES pH 7.5, 10% (v/v) glycerol, 0.3M NaCl and 10 mM βME. Fractions containing greater than 95% pure recombinant scNS3–NS4A, as judged by sodium dodecyl sulfate (SDS–PAGE), were pooled. The pooled fractions were diluted with an equal volume of buffer (25 mM HEPES pH 7.5, 10% (v/v) glycerol and 10 mM βME) and applied to a Q-Sepharose column equilibrated in 25 mM HEPES pH 7.5, 10% (v/v) glycerol, 0.15 M NaCl and 10 mM βME. The flow-through containing the scNS3–NS4A was collected and the final NaCl concentration adjusted to 1.0 M. The purified protein was frozen and stored at –80°C.

Crystallization

Purified scNS3–NS4A was concentrated to 0.07 mM by centrifugal filtration. After centrifugation at 100,000 × g for 30 min, crystallization experiments were conducted using a batch method with microseeding. Crystals suitable for structure determination were grown from a droplet containing 1 μl of protein mixed with 4 μl of a solution containing 0.08 M sodium phosphate pH 6.4, 16% (w/v) PEG 6000, 32 mM di-potassium hydrogen phosphate, 8% (w/v) 2-methyl-2,4-pentanediol (MPD), 5 mM HEPES, 0.2 M sodium chloride, 2% (v/v) glycerol and 2 mM βME. The solution was also used to surround the crystallization chamber containing the protein mixture placed on a microbridge. Crystals of rectangular prism habit (20 × 20 × 50 μ) appeared within 1–2 days. Prior to data collection, crystals were transferred for 2 min to an artificial mother liquor solution supplemented with 20% (v/v) glycerol and flash frozen using either a nitrogen gas stream or liquid propane.

Structure determination and analysis

scNS3–NS4A crystals are orthorhombic (space group $P2_12_12_1$; cell dimensions a = 91 Å, b = 111 Å, c = 141 Å) and contain two molecules per asymmetric unit (V_M = 2.5 Å^3 Dalton^{-1}). Using one crystal and a 1.000 Å wavelength, diffraction data were collected on a MAR CCD detector at beamline X17 of the Advanced Photon Source (Argonne, IL). A total of 270 frames were collected, each with 0.5° oscillation range. Data were integrated, scaled and reduced with the programs DENZO and SCALEPACK [50].

The scNS3–NS4A structure was determined by molecular replacement methods as coded in X-PLOR [51] using the previously determined structures of the protease domain [22,24] (PDB code 1A1R) and helicase domain [21] (PDB code 1HE1) as independent search models. Searches using the helicase revealed two solutions, whereas the protease yielded only a single solution. Rigid-body refinement procedures applied using these solutions produced a globular scNS3–NS4A molecule, which was used as the search model in a second molecular replacement cycle. This yielded two distinct solutions related by a noncrystallographic twofold symmetry axis nearly parallel to the b unit cell axis. The structure was further refined using simulated-annealing and positional and B-factor refinement (X-PLOR 3.1), while gradually extending the resolution. Both search models were derived from the HCV strain 1a, and the appropriate amino acid changes to model the 1B strain were made after the refinement resolution was beyond 2.8 Å. The R_{free} [52] was closely monitored throughout the refinement. In the final model, continuous electron density is observed for all residues of the protein with few exceptions. Data collection, the final refinement and model statistics are listed in Table 1.

Graphics

Figures 1, 2a, 4a, 5 and 6a were generated with Molscript [53] and Raster3D [54], Figure 2b with GRASP [55], and Figure 4b with LIGPLOT [56].

Accession numbers

The atomic coordinates have been deposited in the Brookhaven Protein Data Bank with accession code 1CU1.

References

1. Choo, Q.L., Kuo, G., Weiner, A.J., Overby, L.R., Bradley, D.W. & Houghton, M. (1989). Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* 244, 359-362.
2. Kato, N., *et al.,* & 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.
3. Takamizawa, A., *et al.,* & Okayama, H. (1991). Structure and organization of the hepatitis C virus genome isolated from human carriers. *J. Virol*. 65, 1105-1113.
4. Grakoui, A., Wychowski, C., Lin, C., Feinstone, S.M. & Rice, C.M. (1993). Expression and identification of hepatitis C virus polyprotein cleavage products. *J. Virol*. 67, 1385-1395.
5. Selby, M.J., Glazer, E., Masiarz, F. & Houghton, M. (1994). Complex processing and protein:protein interactions in the E2:NS2 region of HCV. *Virology* 204, 114-122.
6. Lin, C., Lindenbach, B.D., Pragai, B.M., McCourt, D.W. & Rice, C.M. (1994). Processing in the hepatitis C virus E2-NS2 region: identification of p7 and two distinct E2-specific products with different C termini. *J. Virol*. 68, 5063-5073.
7. Bartenschlager, R., Ahlborn-Laake, L., Mous, J. & Jacobsen, H. (1993). Nonstructural protein 3 of the hepatitis C virus encodes a serine-type proteinase required for cleavage at the NS3/4 and NS4/5 junctions. *J. Virol*. 67, 3835-3844.
8. Houghton, M. (1996). Hepatitis C viruses. In *Fields Virology.* (Fields, B.N., Knipe, D.M. & Howley, P.M., eds), pp. 1035-1058, Lippincott-Raven Publishers, Philadelphia, USA.
9. Blight, K.J., Kolykhalov, A.A., Reed, K.E., Agapov, E.V. & Rice, C.M. (1998). Molecular virology of hepatitis C virus: an update with respect to potential antiviral targets. In *Therapies for Viral Hepatitis.* (Schinazi, R.F., Sommadossi, J.-P. & Thomas, H.C., eds), pp. 207-217, International Medical Press, London, UK.
10. Neddermann, P., Tomei, L., Steinkuhler, C., Gallinari, P., Tramontano, A. & De Francesco, R. (1997). The nonstructural proteins of the hepatitis C virus: structure and functions. *J. Biol. Chem*. 378, 469-476.
11. Grakoui, A., McCourt, D.W., Wychowski, C., Feinstone, S.M. & Rice, C.M. (1993). A second hepatitis C virus-encoded proteinase. *Proc. Natl Acad. Sci. USA* 90, 10583-10587.
12. Hijikata, M., *et al.*, & 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.
13. Santolini, E., Pacini, L., Fipaldini, C., Migliaccio, G. & Monica, N. (1995). The NS2 protein of hepatitis C virus is a transmembrane polypeptide. *J. Virol*. 69, 7461-7471.
14. Wu, Z., Yao, N., Le, H.V. & Weber, P.C. (1998). Mechanism of autoproteolysis at the NS2–NS3 junction of the hepatitis C virus polyprotein. *Trends Biochem. Sci*. 23, 92-94.
15. Grakoui, A., McCourt, D.W., Wychowski, C., Feinstone, S.M. & 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.
16. Tomei, L., Failla, C., Santolini, E., De Francesco, R. & La Monica, N. (1993). NS3 is a serine protease required for processing of hepatitis C virus polyprotein. *J. Virol*. 67, 4017-4026.
17. Manabe, S., *et al.,* & Okayama, H. (1994). Production of nonstructural proteins of hepatitis C virus requires a putative viral protease encoded by NS3. *Virology* 198, 636-644.
18. De Francesco, R., Urbani, A., Nardi, M.C., Tomei, L., Steinkuhler, C. & Tramontano, A. (1996). A zinc binding site in viral serine proteinases. *Biochemistry* 35, 13282-13287.
19. Sali, D.L., *et al.,* & Weber, P.C. (1998). Serine protease of hepatitis C virus expressed in insect cells as the NS3/4A complex. *Biochemistry* 37, 3392-3401.
20. Bartenschlager, R., Ahlborn-Laake, L., Mous, J. & Jacobsen, H. (1994). Kinetic and structural analyses of hepatitis C virus polyprotein processing. *J. Virol*. 68, 5045-5055.
21. Yao, N., *et al.,* & Weber, P.C. (1997). Structure of the hepatitis C virus RNA helicase domain. *Nat. Struct. Biol*. 4, 463-467.
22. 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* 87, 343-355. Erratum published in *Cell* (1997) 89, 159.
23. 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* 87, 331-342.
24. Yan, Y., *et al.,* & Chen, Z. (1998). Complex of NS3 protease and NS4A peptide of BK strain hepatitis C virus: a 2.2 Å resolution structure in a hexagonal crystal form. *Protein Sci*. 7, 837-847.

25. Turner, B.G. & Summers, M.F. (1999). Structural biology of HIV. *J. Mol. Biol*. 285, 1-32.
26. Wlodawer, A. & Vondrasek, J. (1998). Inhibitors of HIV-1 protease: a major success of structure-assisted drug design. *Annu. Rev. Biophys. Biomol. Struct.* 27, 249-284.
27. Pasquo, A., *et al.,* & De Francesco, R. (1998). Rational design and functional expression of a constitutively active single-chain NS4A-NS3 proteinase. *Fold. Des*. 3, 433-441.
28. Taremi, S.S., *et al.,* & Malcolm, B.A. (1998). Construction, expression, and characterization of a novel fully activated recombinant single-chain hepatitis C virus protease. *Protein Sci*. 7, 2143-2149.
29. Howe, A., *et al.,* & Lau, J. (1999). A novel recombinant single-chain hepatitis C virus NS3-NS4A protein with improved helicase activity. *Protein Sci*. 8, 1-10.
30. Schechter, I. & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun*. 27, 157-162.
31. Nienaber, V.L., Breddam, K. & Birktoft, J.J. (1993). A glutamic acid specific serine protease utilizes a novel histidine triad in substrate binding. *Biochemistry* 32, 11469-11475.
32. Tong, L., Wengler, G. & Rossmann, M.G. (1993). Refined structure of Sindbis virus core protein and comparison with other chymotrypsin-like serine proteinase structures. *J. Mol. Biol*. 230, 228-247.
33. Llinas-Brunet, M., *et al.,* & Lamarre, D. (1998). Studies on the C-terminal of hexapeptide inhibitors of the hepatitis C virus serine protease. *Bioorg. Med. Chem. Lett*. 8, 2719-2724.
34. Love, R.A., *et al.,* & Hostomska, Z. (1998). The conformation of hepatitis C virus NS3 proteinase with and without NS4A: a structural basis for the activation of the enzyme by its cofactor. *Clin. Diagn. Virol.* 10, 151-156.
35. McPhee, F., Good, A.C., Kuntz, I.D. & Craik, C.S. (1996). Engineering human immunodeficiency virus 1 protease heterodimers as macromolecular inhibitors of viral maturation. *Proc. Natl Acad. Sci. USA* 93, 11477-11481.
36. Tang, J. & Lin, X. (1994). Engineering aspartic proteases to probe structure and function relationships. *Curr. Opin. Biotechnol*. 5, 422-427.
37. Lin, X., Koelsch, G., Loy, J.A. & Tang, J. (1995). Rearranging the domains of pepsinogen. *Protein Sci*. 4, 159-166.
38. Pipe, S.W. & Kaufman, R.J. (1997). Characterization of a genetically engineered inactivation-resistant coagulation factor VIIIa. *Proc. Natl Acad. Sci. USA* 94, 11851-11856.
39. Kim, J.L., *et al.,* & Caron, P.R. (1998). Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. *Structure* 6, 89-100.
40. Cho, H., *et al.,* & Oh, B. (1998). Crystal structure of RNA helicase from genotype 1b hepatitis C virus. A feasible mechanism of unwinding duplex RNA. *J. Biol. Chem*. 273, 15045-15052.
41. Gallinari, P., *et al.,* & De Francesco, R. (1998). Multiple enzymatic activities associated with recombinant NS3 protein of hepatitis C virus. *J. Virol.* 72, 6758-6769.
42. Bode, W., Turk, D. & Karshikov, A. (1992). The refined 1.9 Å X-ray crystal structure of D-Phe-Pro-Arg chloromethylketone-inhibited human α-thrombin: structure analysis, overall structure, electrostatic properties, detailed active-site geometry, and structure-function relationships. *Protein Sci*. 1, 426-471.
43. Ishido, S., Fujita, T. & Hotta, H. (1998). Complex formation of NS5B with NS3 and NS4A proteins of hepatitis C virus. *Biochem. Biophys. Res. Commun*. 244, 35-40.
44. Rice, C.M. (1996). Flaviviridae: the viruses and their replication. In *Fields Virology.* (Fields, B.N., Knipe, D.M. & Howley, P.M., eds), pp. 931-959. Lippincott-Raven Publishers, Philadelphia, USA.
45. Kumar, P.K., *et al.,* & Nishikawa, S. (1997). Isolation of RNA aptamers specific to the NS3 protein of hepatitis C virus from a pool of completely random RNA. *Virology* 237, 270-282.
46. Urvil, P.T., Kakiuchi, N., Zhou, D.M., Shimotohno, K., Kumar, P.K. & Nishikawa, S. (1997). Selection of RNA aptamers that bind specifically to the NS3 protease of hepatitis C virus. *Eur. J. Biochem.* 248, 130-138.
47. Wang, Y. & Guthrie, C. (1998). PRP16, a DEAH-box RNA helicase, is recruited to the spliceosome primarily via its nonconserved N-terminal domain. *RNA* 4, 1216-1229.
48. Hoofnagle, J.H. (1997). Hepatitis C: the clinical spectrum of disease. *Hepatology* 26, 15S-20S.
49. World Health Organization (1997). Hepatitis C: global prevalence. *Weekly Epidemiol. Record* 72, 341-344.
50. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol*. 276, 307-326.
51. Brünger, A.T. (1992*). X-PLOR version 3.1. A System for X-Ray Crystallography and NMR.* Yale University Press, New Haven, CT.
52. Kleywegt, G.J. & Brünger, A.T. (1996). Checking your imagination: applications of the free R value. *Structure* 4, 897-904.
53. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 24, 946-950.
54. Merritt, E.A. & Bacon, D.J. (1997). Raster3D photorealistic molecular graphics. *Methods Enzymol*. 277, 505-524.
55. Nicholls, A., Sharp, K. & Honig, B. (1993). GRASP: graphical representation and analysis of surface properties*. Biophys. J.* 64, 166-170.
56. Wallace, A.C., Laskowski, R.A. & Thornton, J.M. (1995). LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions. *Protein Eng*. 8, 127-134.

Because *Structure with Folding & Design* operates a 'Continuous Publication System' for Research Papers, this paper has been published on the internet before being printed (accessed from http://biomednet.com/cbiology/str). For further information, see the explanation on the contents page.