Protein Science (1998), 7:837–847. Cambridge University Press. Printed in the USA.

Complex of NS3 protease and NS4A peptide of BK strain hepatitis C virus: A 2.2 Å resolution structure in a hexagonal crystal form

YOUWEI YAN,[1] YING LI,[1] SANJEEV MUNSHI,[1] VINOD SARDANA,[1] JAMES L. COLE,[1] MOHINDER SARDANA,[1] CHRISTIAN STEINKUEHLER,[2] LICIA TOMEI,[2] RAFFAELE DE FRANCESCO,[2] LAWRENCE C. KUO,[1] AND ZHONGGUO CHEN[1]

[1]Department of Antiviral Research, Merck Research Laboratories, West Point, Pennsylvania 19486
[2]Istituto di Ricerche di Biologia Molecolare, P. Angeletti, Pomezia 0040, Rome, Italy

(RECEIVED October 20, 1997; ACCEPTED December 21, 1997)

Abstract

The crystal structure of the NS3 protease of the hepatitis C virus (BK strain) has been determined in the space group $P6_322$ to a resolution of 2.2 Å. This protease is bound with a 14-mer peptide representing the central region of the NS4A protein. There are two molecules of the $NS3_{1-180}$–$NS4A_{21'-34'}$ complex per asymmetric unit. Each displays a familiar chymotrypsin-like fold that includes two β-barrel domains and four short α-helices. The catalytic triad (Ser-139, His-57, and Asp-81) is located in the crevice between the β-barrel domains. The NS4A peptide forms an almost completely enclosed peptide surface association with the protease. In contrast to the reported H strain complex of NS3 protease–NS4A peptide in a trigonal crystal form (Kim JL et al., 1996, *Cell 87*:343–355), the N-terminus of the NS3 protease is well-ordered in both molecules in the asymmetric unit of our hexagonal crystal form. The folding of the N-terminal region of the NS3 protease is due to the formation of a three-helix bundle as a result of crystal packing. When compared with the unbound structure (Love RA et al., 1996, *Cell 87*:331–342), the binding of the NS4A peptide leads to the ordering of the N-terminal 28 residues of the NS3 protease into a β-strand and an α-helix and also causes local rearrangements important for a catalytically favorable conformation at the active site. Our analysis provides experimental support for the proposal that binding of an NS4A-mimicking peptide, which increases catalytic rates, is necessary *but* not sufficient for formation of a well-ordered, compact and, hence, highly active protease molecule.

Keywords: hepatitis C virus; NS3 protease; NS4A peptide; serine protease; X-ray structure

Hepatitis C virus (HCV) is a member of Flaviviridae and is the major cause of parenterally transmitted non-A, non-B hepatitis (Choo et al., 1989; Kuo et al. 1989; Houghton, 1996). Chronic infection with HCV in humans has been linked to the development of hepatocellular carcinoma (Saito et al., 1990; Bisceglie, 1995). The HCV genome encodes a polypeptide with the sequence C-E1-E2-NS2-NS3-NS4A-NS4B-NS5A-NS5B, which is proteolytically processed during maturation by the catalytic action of proteases (Hijikata et al., 1991; Grakoui et al., 1993; Lin et al., 1994). Cleavages of the HCV polyprotein at the junctions subsequent to NS3 in the nonstructural (NS) segment are mediated by a viral serine protease, encoded within the N-terminal portion of the NS3 protein encompassing roughly the first 180–200 amino acid residues of NS3 (Bartenschlager et al., 1993; Eckart et al., 1993; Hijikata et al., 1993a; Tomei et al., 1993). In addition, NS4A, which is 54 residues in length, is required for increasing the efficiency of cleavage of the NS3 protease (Bartenschlager et al., 1994; Failla et al., 1994). It has been shown that the central region of NS4A, residue 21 to residue 34, suffices to mimic the effects of the full-length protein (Lin et al., 1995; Tomei et al., 1996; Shimizu et al., 1996). Several reports have suggested that NS4A may be also membrane associated (Hijikata et al., 1993b; Tanji et al., 1995). Thus, the NS3 protease complexed with the NS4A cofactor is believed to play a central role in the replication and maturation of HCV.

We report here a resolution of 2.2 Å crystal structure of the protease domain of NS3 containing residues 1 to 180. This protease is derived from the BK strain of the human hepatitis C virus and is complexed with a 14-mer peptide representing the central region (amino acid residues 21 to 34) of the NS4A protein. The protease-cofactor complex is crystallized in the $P6_322$ hexagonal space group. There are two molecules of the protein complex per

Reprint requests to: Lawrence C. Kuo, Department of Antiviral Research, Merck Research Laboratories, West Point, Pennsylvania 19486.

Abbreviations: DMSO, dimethyl sulfoxide; DTT, dithiothreitol; EDTA, ethylene diamine tetraacetic acid; HCV, hepatitis C virus; HEPES, N-2-hydroxyethyl-piperazine-N'-2-ethanesulfonic acid; MIR, multiple isomorphous replacements; NS, non-structural; RMSD, root-mean-square deviation.

asymmetric unit. The overall folding of the complex and the interactions between the NS3 protease and the NS4A peptide in the hexagonal crystal form reported here are similar to *one* of the two previously reported structures found for the complex of HCV H-strain NS3 and NS4A in the R32 trigonal space group (Kim et al., 1996). The similarity between the $P6_322$ and R32 structures indicates that folding of the NS3 protease is conserved within two virus strains and two crystal forms. However, the second complex structure in the R32 crystals contains a disordered N-terminal 28 amino acids. Based on our data, we surmise that the folded conformation of the N-termini of the NS3 protease in both the $P6_322$ and R32 crystal forms needs be stabilized by crystal packing interactions that could represent its membrane-attached active conformation. The effects of NS4A on the structural integrity and catalytic activity of the NS3 protease are discussed in the context of the $P6_322$ hexagonal crystal form.

Results and discussion

Previous crystals of the protease domain of NS3 have been obtained in the trigonal R32 space group for both the unbound and NS4A-bound forms (Kim et al., 1996; Love et al., 1996). For the NS4A-free form, there are three protease molecules per asymmetric unit. For the NS4A-bound form, there are two protease molecules per asymmetric unit. The truncated proteases used in the study of the NS4A-free and -bound forms are different in sequence and length. The NS3 protease of HCV used to study the NS4A-free form is 189 residues in length, and its amino acid composition is that of the BK strain (Love et al., 1996). The length of the NS3 protease used in the study with bound NS4A peptide is 181 amino acids, and the sequence is that of the H strain (Kim et al., 1996). The crystallization conditions employed previously (Kim et al., 1996; Love et al., 1996) are similar, with NaCl (1.8–3.5 M) as the precipitant, but different from the conditions we use to crystallize our NS3 protease.[3] The availability of the $P6_322$ crystal form from this work permits a comparison of the free and bound forms of the HCV NS3 protease in different crystal packing environments. Our results offer an insight into the importance of NS4A on the folding and activity of NS3 protease.

Overall three-dimensional structure

The crystal structure of the complex of HCV $NS3_{1-180}$–$NS4A_{21'-34'}$ in the $P6_322$ space group is solved with use of the MIR method.[4] At 15–2.8 Å resolution, the refined crystallographic *R*-factor is 0.22, while the R_{free} is 0.31. At 2.2 Å resolution, the *R*-factor is 0.245 and the R_{free} is 0.352. The results indicate that high *B* factors are associated with the structure in this crystal form. Although the data are measured to a resolution of 2.2 Å, diffraction intensities at the high end of the resolution range are relatively weak. We suspect that the weak diffraction spots lower the overall accuracy of the data set and contribute to the higher than anticipated R_{free} value. Nonetheless, the electron density map indicates that the protein structure is well defined. Also, the difference between R_{free} and the crystallographic *R* value lies within the range commonly seen for published structures (Kleywegt & Brünger, 1996). The root-mean-square deviations (RMSDs) from ideal values are 0.013 Å for bond lengths and 2.31° for bond angles. Statistics of data are given in Table 1. There is no amino acid residue falling in

[3]Our protease does not crystallize in the R32 form nor in buffers containing high concentrations of NaCl as the main precipitant.

[4]For clarity, the residue numbers of the NS4A peptide will be referred to in this paper with a prime (').

Table 1. *X-ray diffraction data and refinement statistics for the $NS3_{1-180}$–$NS4A_{21'-34'}$ complex in the $P6_322$ crystal form*

	Native I (−160 °C)	Native II (25 °C)	C_9H_9Hg-O_2SNa[a]	CH_3HgCl	C_2H_5HgCl	CH_3HgI
A. Data set						
Resolution (Å)	2.2	2.8	2.8	2.8	3.2	3.2
Reflection	22,953	12,125	12,117	12,181	6,609	6,919
Completeness (%)	94.0	96.3	96.2	97.8	78.4	81.6
R_{sym} (%)	0.082	0.076	0.102	0.104	0.107	0.115
R_{iso} (%)			0.134	0.127	0.143	0.135
R_{cullis} (%)			0.56	0.69	0.59	0.69
Phasing power			1.73	1.06	1.52	1.11
Number of sites			7	5	5	5
Mean figure of merit 0.69 at 15–3.5 Å and 0.58 at 15–2.8 Å						
B. Refinement statistics						

Resolution 15–2.2 Å, $R = 0.245$, $R_{free} = 0.352$ (10% data); reflections ($F > 2\sigma$) = 21,978.
RMSD bond lengths: 0.013Å, RMSD bond angles: 2.31°.
$R_{sym} = \Sigma|I - \langle I\rangle|/\Sigma I$, where I = observed intensity, $\langle I\rangle$ = average intensity of symmetry related reflections.
$R_{iso} = \Sigma|F_{PH} - F_P|/\Sigma F_P$, where F_P = protein structure amplitude, F_{PH} = heavy-atom derivative structure factor amplitude.
$R_{cullis} = \{\Sigma||F_{PH} \pm F_P| - F_H|\}/\{\Sigma|F_{PH} \pm F_P|\}$, where F_H = heavy atom structure amplitude.
Phasing power: mean value of the heavy-atom structure factor amplitudes divided by the residual lack-of-closure error.

[a]Thimerosal.

the disallowed region of the Ramachandran plot as determined with the program PROCHECK (Laskowski et al., 1993).

The model we have employed to fit the electron density map consists of amino acid residues 3 (isoleucine) to 179 (methionine) of the NS3 protease and residues 21′ to 32′ of the NS4A, plus one zinc ion for each complex in the asymmetric unit. The amino acid sequences of the cloned NS3 protease and NS4A peptide used are given in Figure 1A. The topology of the secondary structure of the protein complex as defined in our structure is given in Figure 1B.

Electron densities are not seen for the first two and the last residues of the NS3 protease, and the C-terminal lysine tag of NS4A is also not seen. Presumably, these residues are too flexible to be detected in the X-ray structure. A total of 215 water molecules are included in our structural model. The two complexes in the asymmetric unit are essentially identical with their N-terminal 28 residues ordered and folded identically. An RMSD of 0.65 Å is found for the C_α atoms of the two complexes. A segment of the final $2F_o - F_c$ map around the active site, phased by the refined model at 2.2 Å, is shown in Figure 2.

As shown in Figure 3, the overall folding of the $NS3_{1-180}$–$NS4A_{21'-34'}$ is similar to chymotrypsin in the serine proteases family.[5] There are two domains, roughly dividing the protease in halves. Each domain is composed of a β-barrel and two short α-helices. The NS4A peptide is bound within the N-terminal domain; its presence augments an otherwise six-strand β-barrel structure to an eight-stand β-barrel structure. The proposed catalytic triad (Hijikata et al., 1993a) of His-57, Asp-81, and Ser-139 resides in a crevice between the two domains. His-57 is located at the beginning of the short helix, α1. Asp-81 is found at the start of the F1 strand, and both His-57 and Asp-81 are in the N-terminal domain. Ser-139 and the "oxyanion loop" (137–139) are situated immediately following the helix, α2, in the C-terminal domain. The catalytic triad has a geometrical arrangement as found in other serine proteases. Additionally, there is density for a metal ion in the C-terminal domain with donor ligands derived from the loop between the F1 and A2 strands and from the loop between the D2 and E2 strands.

When compared with structures of other serine proteases such as thrombin and trypsin, the overall folding of the C-terminal domain of the NS3 protease is a better match than the N-terminal domain. The RMSD of 58 conserved C_α atoms in the C-terminal domain is 1.7 and 1.5 Å, respectively, when super-positioned onto the same domain of thrombin and trypsin. The conserved helix α3 overlaps particularly well with its counterpart in thrombin and trypsin. In contrast, the RMSD of the 42 and 38 conserved C_α atoms in the N-terminal domain when super-positioned onto the same domain of thrombin and trypsin is 1.9 and 1.7 Å, respectively. This domain contains eight β-strands instead of six due to the augmentation introduced by the NS4A peptide. The N-terminal 28 residues in the unbound NS3 protease is completely unfolded (Love et al., 1996). In our structure, the N-terminal 28 residues are folded into a β-strand (A0) and an α-helix (α0) upon binding of the NS4A peptide (Figs. 3, 4). The NS4A peptide itself forms a β-strand, D1′, and is almost completely buried between two β-strands, A0 and A1, in an anti-parallel configuration. Helix α0 is juxtaposed against the N-terminus of the NS4A peptide. There is no structural equivalent

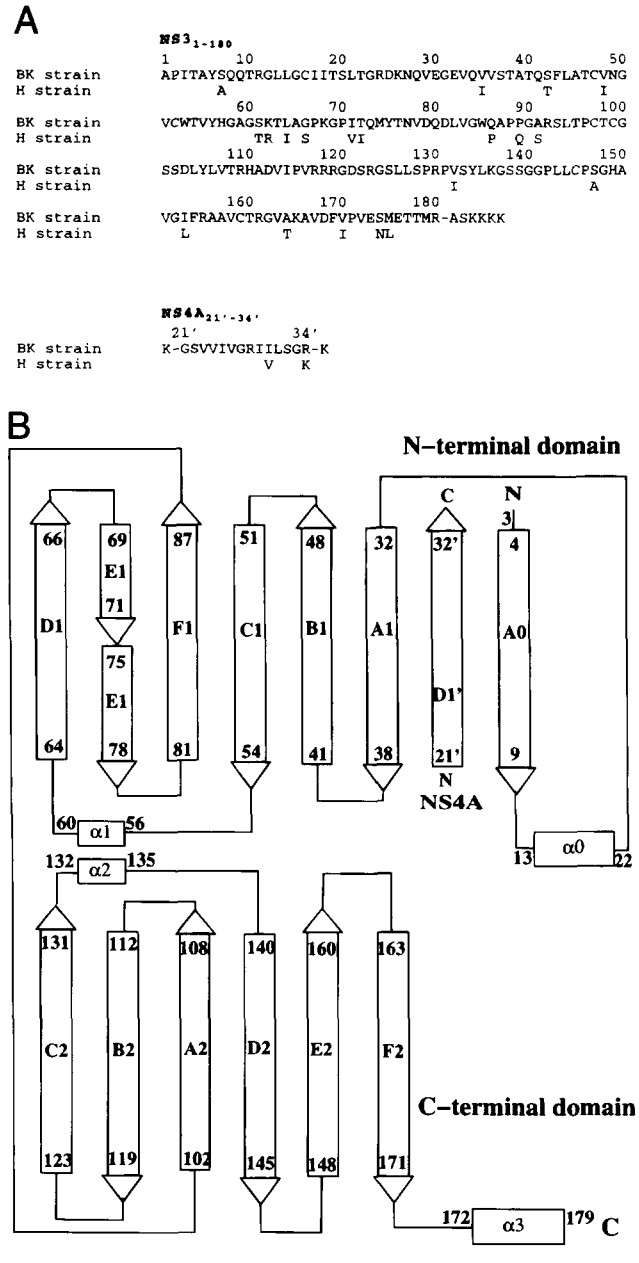

Fig. 1. A: Amino acid sequences of the NS3 protease and the NS4A peptide used in this study. These sequences correspond to residues 1026–1205 and residues 1678–1691 encoded by the open reading frame of the BK strain of HCV (Takamizawa et al., 1991). For comparison purposes, residues in these regions of the H strain that differ from those in the BK strain are also shown. **B:** Topology of the secondary structure of the $NS3_{1-180}$–$NS4A_{21'-34'}$ complex. The α-helices and β-strands are marked according to the labeling scheme employed for trypsin (Huber & Bode, 1978). Helices are shown as cylinders and β-strands as ribbons with arrow heads.

of β-strand A0 and helix α0 for proteases in the chymotrypsin family.

A previous study has revealed that a zinc ion is required for the structural integrity and activity of the NS3 protease (DeFrancesco et al., 1996). An electron density above the 4σ level is seen in the MIR map between the F1/A2 and D2/E2 loops. A zinc ion is thus assigned to it. Three cysteines and one histidine (Cys-97, Cys-99,

[5]Previously, the NS3 protease structure has been compared with chymotrypsin and sindbis virus core protein (Blow, 1976; Tong et al., 1993). Therefore, the comparisons are not repeated here.

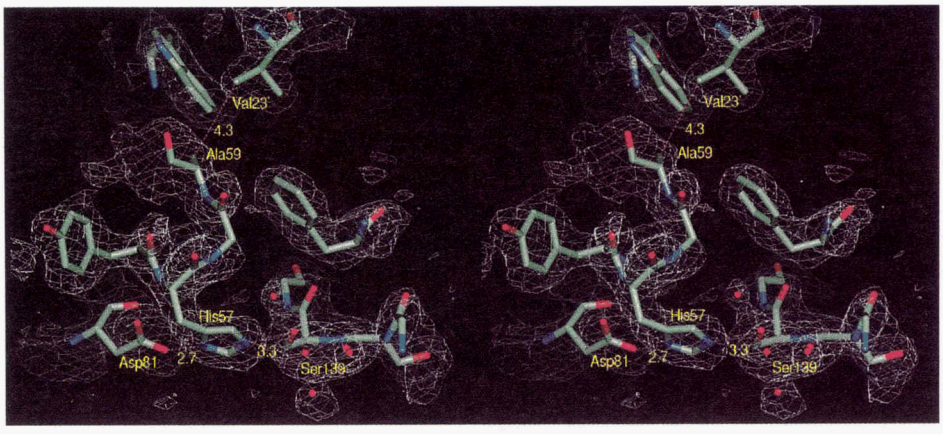

Fig. 2. Stereoview of the electron density map of the active site of the NS3 protease. The $2F_o - F_c$ map is calculated at 2.2 Å and phased with the refined model. The map is contoured at the 1σ level. Val-23′ of NS4A is also shown. It participates in hydrophobic interactions with Ala-59 of helix $\alpha 1$ (see text).

Cys-145, and His-149) provide a distorted tetrahedral scaffold for the binding of the zinc ion. A water molecule may be present between the Zn^{2+} and the imidazole ring of His-149. However, its density is not well defined in the map. Residues 97 to 101 on the F1/A2 loop are also poorly defined, as revealed by the high-temperature factors indicating that high mobility is associated with this loop as well as the zinc ion. Due to the distance (23 Å) between this metal and the active site, it is unlikely that the zinc ion has more than a structural role.

Recently, another viral protease essential for polyprotein processing has been identified to be also a serine protease in the cytomegalovirus. This protease is active only as a homodimer in catalyzing peptide hydrolysis. Each subunit of this protease contains a catalytic triad that maintains a similar geometry as seen for other serine proteases. On the other hand, each subunit is a single domain α/β protein and processes a different overall fold from other members of the serine protease family (Chen et al., 1996; Qiu et al., 1996; Shieh et al., 1996; Tong et al., 1996).

Fig. 3. Stereoview of a ribbon representation (Kraulis, 1991) of $NS3_{1-180}$–$NS4A_{21'-34'}$ as determined with MIR for the protein complex found in the $P6_322$ hexagonal crystal form of this study. The structure is folded into two distinct domains. The N-terminal domain is shown at the top and the C-terminal domain at the bottom of the figure. The NS4A peptide, represented as a yellow ribbon, penetrates through the N-terminal domain of the NS3 protease and augments an otherwise six β-strand barrel to an eight β-strand barrel (see text). The catalytic triad residues, His-57, Asp-81, and Ser-139, are found in a crevice between the two domains. The zinc ion is shown in magenta.

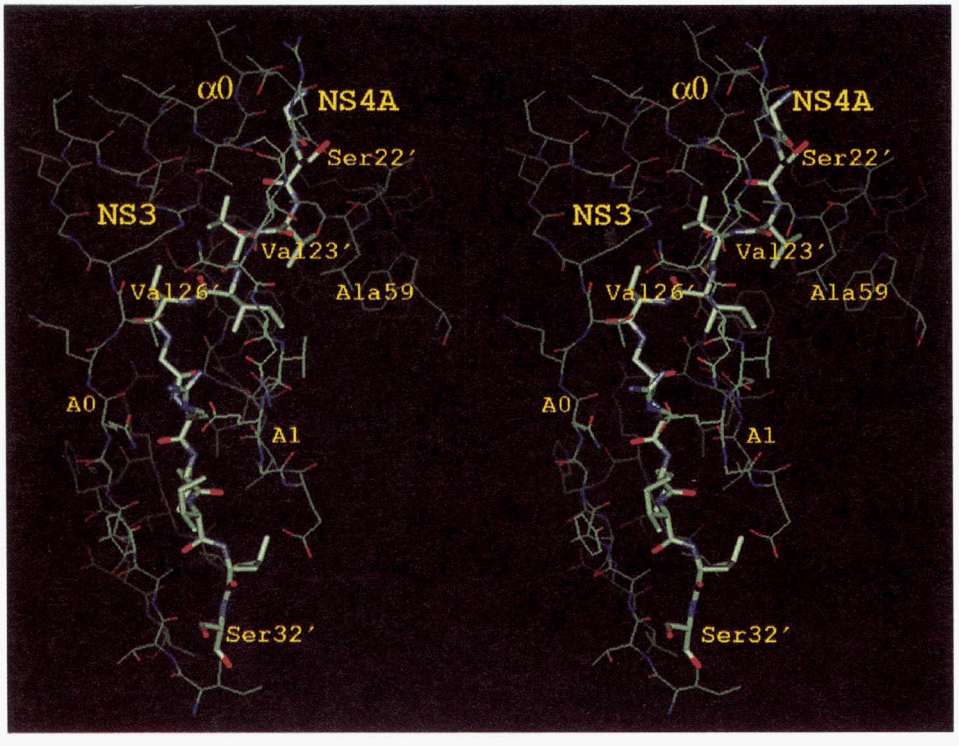

Fig. 4. Stereoview of the binding mode of the NS4A peptide to the NS3 protease. The $NS4A_{21'-34'}$ peptide (thick bonds) interacts with the N-terminus of the NS3 protease in an extended β-strand conformation. The $NS4A_{21'-34'}$ peptide shows a kink in the middle and has extensive interactions with the protease. The N-terminal portion of NS4A is buried and flanked by helices $\alpha 0$ and $\alpha 1$.

Bound conformation of NS4A

As shown in Figure 4, the $NS4A_{21'-34'}$ peptide displays a kink at Ile-25′ and Val-26′. Otherwise, the peptide is in an extended conformation. It forms main-chain hydrogen bonds with the A0 and A1 strands of the protease in an anti-parallel fashion. With the exception of Gly-21′, which is well defined in the electron density map but mostly solvent exposed, and Gly-33′ and Arg-34′, which are not defined presumably due to disorder, the NS4A peptide is essentially completely enclosed. Figure 5 shows the extent to which each residue of the NS4A peptide is buried. We note that the appearance of Figure 5 is quite different from a similar figure derived on the basis of absolute buried area of NS4A (Kim et al., 1996). In the previous report, Val-24′ and Gly-27′ show small buried areas due to their small side chain sizes, whereas our structure shows that both are essentially completely buried—100 and 98%, respectively.

The invasion into a hydrophobic channel of NS3 by the peptide begins from its N-terminal with Ser-22′, which is flanked by helices $\alpha 0$ and $\alpha 1$ (Fig. 3). The C-terminal end of the NS4A peptide emerges from the surface of the protease near the N-terminal end of A0. This association between the protease and peptide, arising from main-chain hydrogen bonding interactions between three well-folded β-strands, is consistent with the observation that the NS3 protease is more stable in solution in the presence of NS4A (Tanji et al., 1995). This augmentation of β-strands by interactions between a protein and a peptide is strikingly similar to the assembly interactions found for the VP1 protein in SV40 and polyomavirus capsids (Liddington et al., 1991; Yan et al., 1995). In the formation of a capsid, 360 VP1 protein molecules are involved. The C-terminal segment, in a β-strand, of every VP1 interacts with a β-strand in the N-terminal region of the next VP1 and is clamped in place with the help of the N-terminal segment of that VP1. A three-chain, intercalating anti-parallel β-strand interaction is formed. Both the "invading" C-terminus and the "clamping" N-terminus of the VP1 are ordered in the capsid, but otherwise disordered in the unassembled VP1. This type of cooperative, perpetuating peptide–protein surface association offers great flexibility for binding, because the

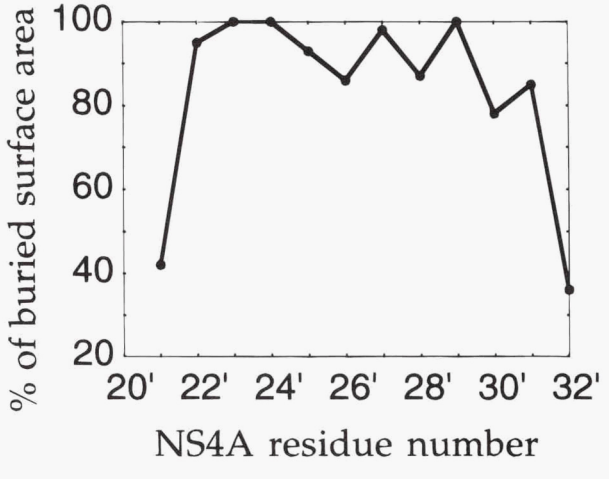

Fig. 5. Buried surface area as a percentage of total surface area of the NS4A peptide in the NS3 protease as determined by the $P6_322$ crystal structure.

direction the peptide enters and leaves the region of contact can vary (Harrison, 1996).

$NS4A_{21'-34'}$ represents the central region of the 54-amino acid NS4A protein. The binding conformation of this peptide is essentially identical to that found by Kim et al. (1996), based on a visual comparison. The NS4A peptide employed in the Kim et al. study is 19 residues in length (21′–39′), while the amino acids involved in direct association with the NS3 protease are the same 12 residues (21′–32′) as found in this work despite a difference in HCV strain and crystal form. These similarities suggest that binding to NS3 protease by the full length NS4A most likely retains the same binding mode in this region of NS4A.

Induced changes in the conformation of NS3 protease

The binding of the NS4A peptide alters the folding of the NS3 protease in several places. These conformational changes may have significant impact on the activity of the protease. The most obvious conformational change is the folding of the N-terminal 28 residues of the protease. These residues are unfolded in the unbound form (Love et al., 1996). Upon binding of the NS4A peptide, they are folded into a β-strand (A0) and an α-helix (α0), as shown in Figure 3. The same observation has been made for one of the two NS3 protease complexes in the R32 crystal form of HCV H-strain (Kim et al., 1996).

A second conformational change of NS3 is the displacement of D1 β-strand of the protease. In the free enzyme, this strand is situated in the same position as the analogous D1 strand in chymotrypsin (Blevins & Tulinsky, 1985). In our bound structure, the N-terminal portion of the NS4A peptide (D1′) assumes the D1 location, and hence dislodges D1 away from A1. The D1′ of NS4A now lies adjacent to helix α1. As shown in Figure 6, Val-23′ of D1′ is 4.3 Å away from Ala-59 of helix α1. The resultant hydrophobic interaction lends structural support and rigidity to α1 on which one of the catalytic triad residues, His-57, resides. Indeed, α1 and His-57 show a well-defined geometry in our structure, as indicated by clear electron density and a low temperature factor (see Fig. 2). In contrast, in the free enzyme "the stretch of residues from 57 to 63 has greater mobility than most loops in the structure" (Love et al., 1996). The conformations of the other two catalytic triad residues, Asp-81 and Ser-139, are also very well defined for both molecules in the asymmetric unit of the $P6_322$ crystal, as shown in Figure 6. The most glaring difference upon binding of NS4A peptide is that the carboxyl group of Asp-81 now points to the imidazole ring of His-57 at a distance of 2.7 Å (ND1-OD2). In the free enzyme, the side chain of Asp-81 is swung away from His-57 and forms a hydrogen bond with the guanidinium moiety of Arg-155 instead (Love et al., 1996). The conformational arrangement of the NS4A-free protease is obviously an unfavorable one for proton shuttle during catalysis. The side chain of Ser-139 in the bound structure is 3.3 Å away from the imidazole group of His-57 (OG-NE2), whereas these two groups are too far apart to engage in bonding interactions in the structure of the free NS3 protease (Love et al., 1996). The geometrical conformation of the triad in the NS4A-bound structure of NS3 protease is one as seen for other serine proteases.

Together, these observations offer an explanation of the activation of the catalytic activity of NS3 protease by the NS4A peptide. With our enzyme, the enhancement of the pseudo-second-order rate constant (k_{cat}/K_M) due to binding of the NS4A peptide is ~950-fold for catalytic hydrolysis of peptides mimicking the cleavage junctions of NS4A/NS4B.

Fig. 6. Stereoview of a ribbon representation depicting the displacement of D1 of the NS3 protease compared to the location of D1 in the structure of chymotrypsin (Blevins & Tulinsky, 1985), due to NS4A peptide binding. The position of D1 of chymotrypsin, shown in blue, is obtained by super-positioning the C_α backbone of chymotrypsin onto that of the NS3 protease. For clarity, only D1 of chymotrypsin is shown. In the free enzyme, D1 of NS3 is situated in a position as seen for the D1 strand of chymotrypsin (Love et al., 1996). In our bound structure, the N-terminal portion of NS4A peptide (D1′) dislodges D1 away from A1. The D1′ of NS4A now lies adjacent to helix α1, with its Val-23′ 4.3 Å away from Ala-59. The resultant hydrophobic interaction lends structural support and rigidity to α1 on which one of the catalytic triad residues, His-57, resides. The distances (in Å) between the active triad residues are shown in dashed lines.

Crystal packing and implications

A surprising structural feature of the co-crystal is that the hydrophobic residues on helix α0 (Leu-14, Ile-7, Ile-18, and Leu-21) all face outward from the $NS3_{1-180}$–$NS4A_{21'-34'}$ complex to create a hydrophobic patch on the surface of this part of the complex. As part of the crystal packing in the $P6_322$ space group, these hydrophobic residues interact with their counterparts related by a crystallographic threefold axis, forming a three-helix bundle. The cross-angles of the three-helix bundles formed by two independent complexes are 20° and 60°, respectively. Because inter-protein, α-helix bundles in solution are generally formed with helices eight or more turns in length, the existence of a three-turn, three-helix bundle is most likely due to crystal packing (Lovejoy et al., 1993; Yan et al., 1993). Thus, the three-helix bundle provides in the crystal a hydrophobic and stabilizing environment for α0 that otherwise does not exist in solution. At opposite ends of helix α0 are Arg-11 and Asp-25, engaged in a salt bridge which, in turn, stabilizes the N-terminal A0 strand and its interactions with the NS4A peptide as discussed above. This "cascade" of stabilizing interactions are shown in Figure 7A and B. From a structural point of view, the three-helix bundle is the inducing factor for the peptide–surface association that ensures a tight interaction between NS3 and NS4A in the protein crystal.

We note that one of the two NS3–NS4A complexes in the asymmetric unit in the trigonal R32 crystal form has its N-terminal 28 residues completely disordered despite the presence of the NS4A peptide (Kim et al., 1996). In the absence of their structure coordinates, we reason that the disorder is due to lack of stabilizing forces as seen here for the hexagonal crystal form. In view of the structures of the free protease, of the bound protease in the trigonal form, and of the bound protease in the hexagonal form, *we propose that the presence of NS4A peptide is necessary but insufficient for the proper folding of the N-terminal 28 residues of the NS3 protease*. A local hydrophobic environment is required for the formation of the helix α0. In our structure, the N-terminal 28 residues of NS3 provide 36% of interface between the $NS4A_{21'-34'}$ peptide and the protease. Without these residues, the buried surface area of the NS4A peptide is reduced from ~1,380 Å^2 to ~880 Å^2.

Experiments have also indicated that NS4A is a membrane-targeting protein that brings NS3 to the membrane of the endoplasmic reticulum (Hijikata et al., 1993b; Tanji et al., 1995). It has been predicted with the use of multiple sequence alignments that the first N-terminal 20 residues of NS4A form a hydrophobic trans-membrane helix (Rost et al., 1995). This proposal is consistent with our structure of the $NS3_{1-180}$–$NS4A_{21'-34'}$ complex. The first 20 helix-forming residues of the NS4A protein can be envisioned to be pointing away from the NS3–NS4A complex to act as a membrane anchor for NS3. The membrane association can be viewed to serve the same role to provide hydrophobic stability for the tight association between NS3 and NS4A in a manner as seen in the hexagonal crystal form. A schematic representation in Figure 7C shows a possible interaction between the N-terminal 20 residues of the NS4A membrane anchor domain, the membrane of the endoplasmic reticulum, and helix α0 of the NS3 protease. This attachment, ensuring proper and tight association of NS3 and NS4A due to their interaction as induced by the three-helix bundle shown for the complex of the truncated protease in the hexagonal crystal form, could be the spatial activation trigger point for polyprotein processing and viral maturation. Evidence has shown that the replication of the flavivirus is closely associated with host cell membranes (Rice et al., 1986). The structure of the $NS3_{1-180}$–$NS4A_{21'-34'}$ complex in our crystals may then represent its membrane-attached conformation.

Materials and methods

Preparation of the NS3 protease and NS4A peptide

Escherichia coli BL21DE3 (plysS) cells from Novagen were transformed with a plasmid under the control of the T7 promoter and containing the cDNA encoding residues 1026 to 1205, of the open reading frame of the BK-strain HCV, and a four-lysine tag. The amino acid sequence of the NS3 protease cloned is given in Figure 1A together with the sequence of the NS4A peptide.

Cells were grown at 37 °C in LB medium containing ampicillin to an optical density of 0.6–0.8 at 600 nm. Induction (Boehringer-Mannheim, Indianapolis, IN) was performed at 25 °C with the addition of 1 mM isopropyl-D(−)-thiogalactopyranoside. Cells from a 30-L culture were resuspended at 4 °C in a lysis buffer [25 mM sodium phosphate pH 7.5, 1 mM ethylene diamine tetraacetic acid (EDTA), 10% glycerol, 5 mM dithiothreitol (DTT)] and treated with 0.02 mg/mL DNase (Type IIS: bovine pancreas; Sigma, St. Louis, MO). Cell lysis was performed with use of a microfluidizer at 6 bar. The supernatant was chromatographed on a Hi-Load SP Sepharose High Performance column (XK50/20: Pharmacia Biotech, Piscataway, NJ) pre-equilibrated in a buffer containing 50 mM sodium phosphate (pH 6.5), 10% glycerol, 1 mM EDTA, and 5 mM DTT. The NS3 protease was eluted in a 0–1 M NaCl gradient. Fractions containing the protease were pooled and chromatographed on a Hi-trap Heparin 50 mL column (XK26/20, Pharmacia-Biotech, Piscataway, NJ) with a NaCl gradient. Fractions containing NS3 protease were concentrated on a Hi-Load SP Sepharose High Performance column (XK16/10: Pharmacia Biotech, Piscataway, NJ) using a buffer containing 25 mM N-2-hydroxyethyl-piperazine-N′-2-ethanesulfonic acid (pH 7.0), 5 mM DTT, and 10% glycerol. Protein fractions demonstrating >99% purity, as determined with N-terminal sequence analysis (Edman degradation on an Applied Biosystem model 470A gas phase sequencer), were pooled, concentrated to 25 mg/mL, and stored at −80 °C. Protein concentrations were determined with quantitative amino acid analysis.

N-Terminal sequencing data indicate that the expressed protein was missing its N-terminal alanine and the initiation methionine. The amino acid composition of the protease was as expected and its sequence was confirmed by the X-ray structure determined in this work. The purified protease displayed a very high catalytic activity. Using peptide substrates mimicking the cleavage junction of NS4A/NS4B, the catalytic efficiency of the $NS3_{1-180}$ protease, expressed in terms of k_{cat}/K_M, was 196,000 ± 9,000 M^{-1} s^{-1} in the presence of saturating $NS4A_{21'-34'}$.

A peptide mimicking the central portion of NS4A (residues 1678 to 1691 encoded in the open reading frame of HCV BK strain), flanked by two lysine residues, one each at the N- and C-termini, was purchased from Enzyme Systems Products. Its sequence, also given in Figure 1A, was confirmed with amino acid analysis.

Enzyme assays

Kinetic reactions were carried out in 50 mM HEPES at pH 7.5 in the presence of 10 mM DTT, 50% glycerol, 2–5 nM enzyme, and 10 μM NS4A peptide. The substrate used was 7-methoxycoumarin-4-acetyl-DEMEECASHLPYK-(ϵ-NHCOCH$_3$). The protease and

NS4A peptide were preincubated for 10 min prior to initiation of reaction with substrate. Steady-state and initial velocity conditions were strictly observed.

Protein crystals

Crystallization of the $NS3_{1-180}$ protease was performed in the presence of the $NS4A_{21'-34'}$ peptide. The protease (~22 mg/mL) was mixed with the synthesized NS4A peptide in a 1:1.2 molar ratio in a solution containing 10% glycerol, 5 mM DTT, and 25 mM HEPES buffer, pH 7. Hanging drops were prepared with use of a mother liquor containing 2.2 M ammonium sulfate, 0.2 M NaCl, and 0.1 M sodium cacodylate buffer at pH 6.5. Equal volumes of the protein solution and mother liquor were applied for the preparation of hanging drops. Hexagonal crystals, up to 0.5 × 0.5 × 0.4 mm in size, were obtained at 4 °C within several days. Protein derivatives were prepared with the protein crystals soaked in the mother liquor containing heavy-atom compounds.

Our $NS3_{1-180}$ protease crystallized in the hexagonal crystal form in the presence of $NS4A_{21'-34'}$ with a space group of $P6_322$. There were two protein complexes per asymmetric unit with a solvent content of 56% v/v. The unit cell dimensions at ambient temperature were $a = b = 98.8$ Å and $c = 167.6$ Å. At 100 K, the cell dimensions were $a = b = 97.0$ Å and $c = 167.1$ Å.

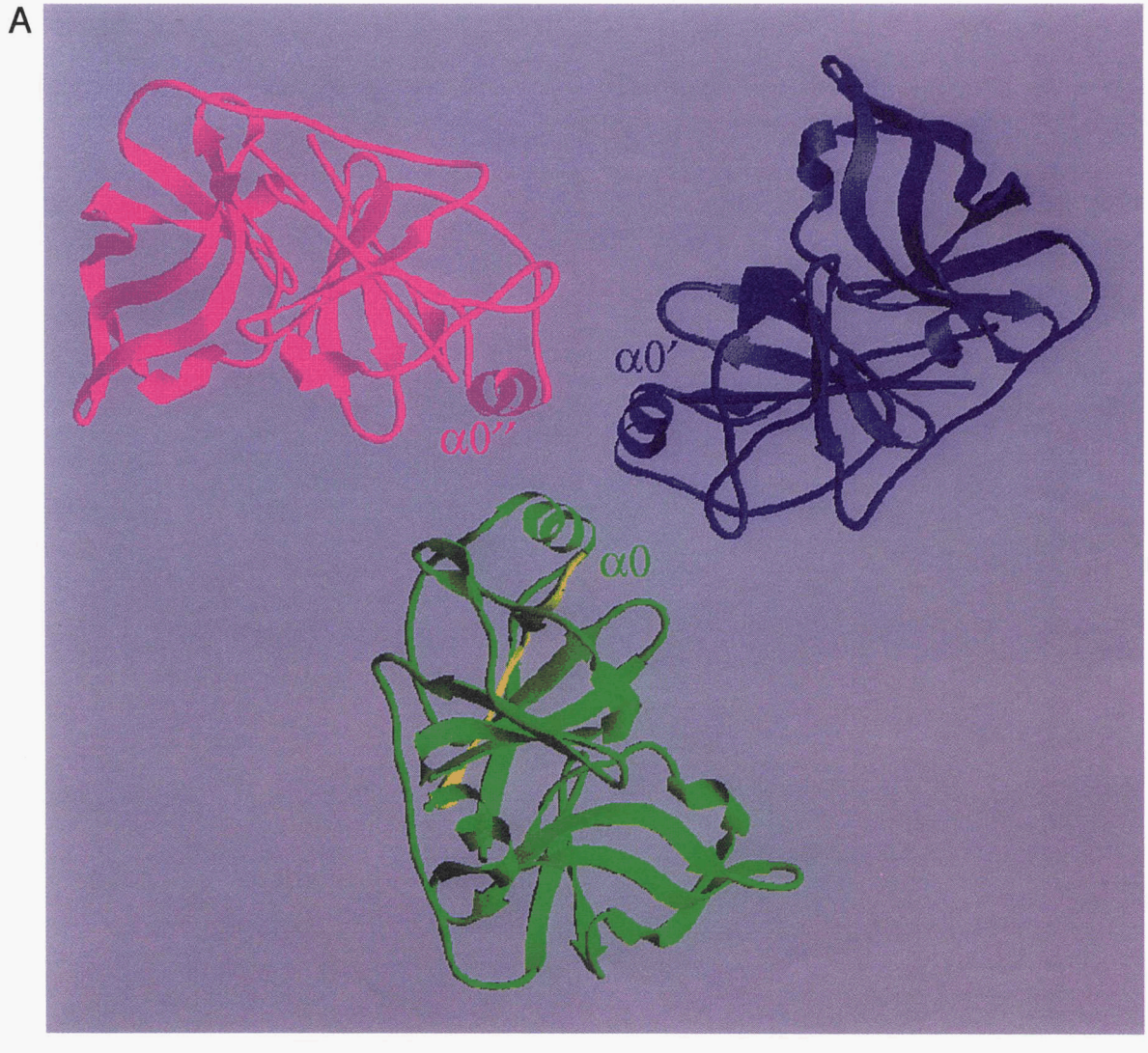

Fig. 7. Crystal packing of the $P6_322$ hexagonal crystal form of the NS3 protease bound with the NS4A peptide. **A:** A ribbon diagram (Carson, 1991) of three independent $NS3_{1-180}$–$NS4A_{21'-34'}$ complexes as related by a crystallographic threefold symmetry form a three-helix bundle through helices $\alpha 0$ ($\alpha 0'$ and $\alpha 0''$ from symmetric-related molecules in blue and magenta). **B:** For clarity, one complex in green and two helices, $\alpha 0'$ and $\alpha 0''$, of the symmetry-related molecules are shown in the ribbon diagram in blue and magenta. Helix $\alpha 0$ protrudes outward from the protease complex to create a hydrophobic patch at the molecular surface of a crystallographic threefold symmetry axis. A local hydrophobic environment is obtained due to the formation of a three-helix bundle between the symmetry-related complexes. The salt bridge between Arg-11 and Asp-25, at the opposite ends of the helix $\alpha 0$, is also shown. **C:** The proposed transmembrane helix of the N-terminal 20 amino acids is modeled in gold color. The membrane of the endoplasmic reticulum is proposed here to be in close proximity to the helix $\alpha 0$ to provide a similar hydrophobic environment as the three-helix bundle found in the hexagonal crystals (see text). (*Figure continues on facing page.*)

X-ray data collection

The native data at 2.8 Å resolution and four heavy-atom derivative data at 2.8 Å or 3.2 Å resolution were collected at ambient temperature with use of CuK_α irradiation generated by a rotating anode (Rigaku RU 200) at 50 kV/100 mA and focused with a double mirror system (Raxis II). A native data set at 2.2 Å resolution was collected at 100 K. Prior to data collection, the crystals were soaked in the mother liquor mixture containing a cryo-protective agent, glycerol (20% v/v), for 2 h. The frozen crystals retained the $P6_322$ space group, but its cell volume was decreased by ~2%. Diffraction data were processed with the program HKL (Otwinowski & Minor, 1997). Data statistics are shown in Table 1.

Crystallographic phasing

The structure of the $NS3_{1-180}$–$NS4A_{21'-34'}$ complex was determined with use of multiple isomorphous replacements (MIR). Two major sites of thimerosal derivatization were found with use of the program SHELEX (Sheldrick, 1991). Two other major sites and three minor sites were located with the difference Fourier method, phased with two and then four major sites, respectively. Three mercury derivatives (CH_3HgCl, C_2H_5HgCl, and CH_3HgI) were identified with the difference Fourier method, phased with four major sites of the thimerosal derivative. Five to seven heavy-atom sites were found for each mercury derivative and all shared the same major sites. The heavy atom parameters were refined by program HEAVY (Terwilliger & Eisenberg, 1983). Temperature factors of the thimerosal derivative were refined while those of other derivatives were fixed at the mean value of 40 $Å^2$. The anomalous signals from the thimerosal derivative were used. Because all derivatives were highly correlated, only the origin-removed Patterson function was used for refinement to avoid "cross-talks" between derivatives. The phases were calculated to 2.8 Å resolution (mean figure of merit = 0.58) and two derivatives had no contribution beyond 3.2 Å (Table 1).

Model building and refinement

The initial electron density map was generated with the MIR and solvent-flattened phases at 15–2.8 Å resolution (Yan et al., 1996). It revealed clear protein-solvent boundary and continuous main-chain peptide densities. There were only a few disjointed densities but the side chains were clearly seen with a mean figure of merit of 0.83. About 90% of the two independent complexes in the asymmetric unit was built into this map with the program CHAIN (Sack, 1988). A twofold local symmetry between the two protein complexes in the asymmetric unit was derived and then refined with the program PHASES (Furey & Swaminathan, 1997). Masks were derived from this partial model and some parts of the F1/A2 loop were filled in with dummy atoms to ensure that the masks covered the entire complex. The twofold averaged and solvent-flattened map at 2.8 Å resolution displayed smaller phase errors and revealed a better connectivity of the weak F1/A2 loop with a mean figure of merit of 0.82. A model for both complexes in the asymmetric unit was completed based on the averaged map.

The initial model gave an *R*-factor of 0.43 with the room temperature data set in the 15–2.8 Å resolution range. Several cycles

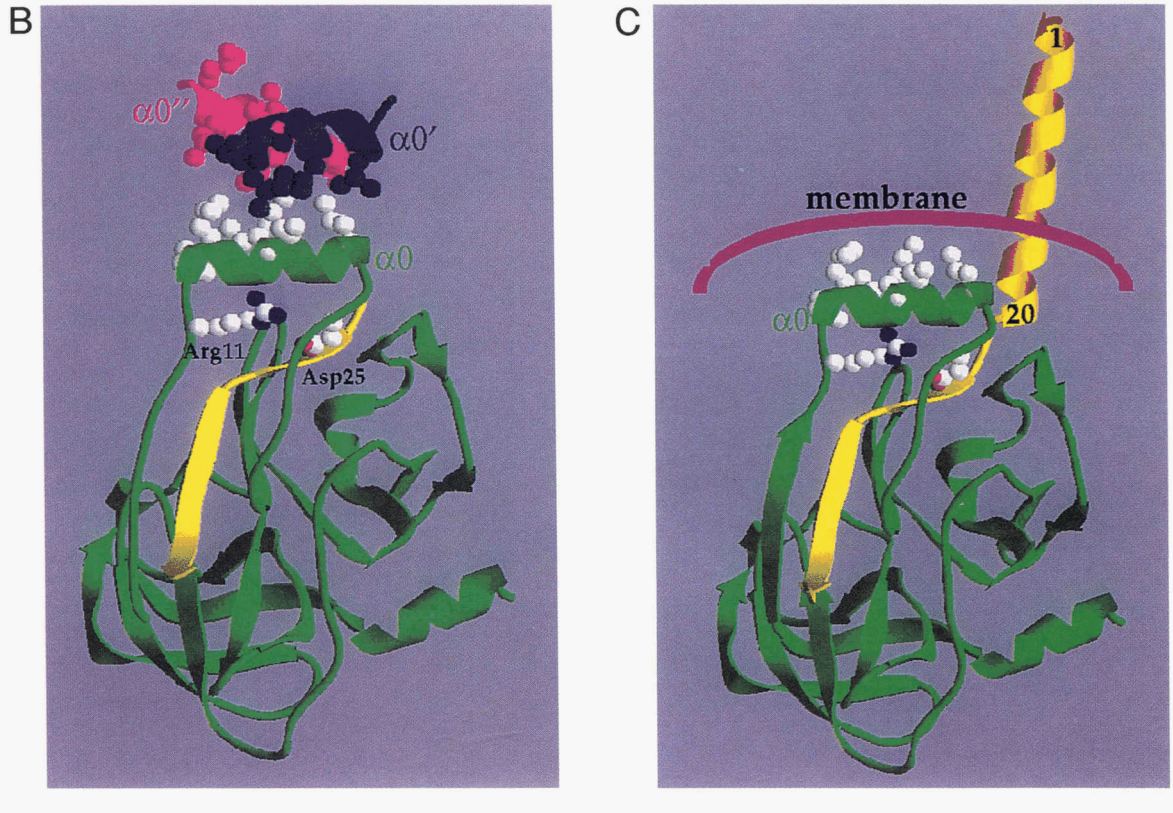

Fig. 7. *Continued.*

of refinement employing XPLOR (Brünger, 1992, 1993) and with restraints of noncrystallographic symmetry between the two protein complexes within the asymmetric unit reduced the R-factor to 0.22 and the R_{free} to 0.31 (Yan et al., 1996). At this point, the high resolution data at 2.2 Å was incorporated. The model was fit into the shrunken cell of the frozen crystals as one, two, and four rigid body segments at 4 Å and then 2.8 Å resolution. The model was found to shift by 1.2 Å in the *ab* plane perpendicular to the *a*-axis in the shrunken cell. There was no orientational change between the two independent protein complexes, nor was there any change between the two domains within the same complex. Coupled with manual adjustments, the model was refined at 2.8 Å and extended sequentially to 2.4 and 2.2 Å resolution. Water molecules were introduced to the model at 2.4 Å resolution based on the $F_o - F_c$ difference Fourier map. The noncrystallographic restraint was released during the final stage of refinement.[6]

Acknowledgments

We thank Mr. Jeff Blue and Mr. Brett Johns for help in protein purification. We also thank Dr. E. Emini for continued support and encouragement throughout the course of this work.

References

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.

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.

Bisceglie AM. 1995. Hepatitis C and hepatocellular carcinoma. *Semin Liver Dis 15*:64–69.

Blevins RA, Tulinsky A. 1985. The refinement and the structure of the dimer of α-chymotrypsin at 1.67Å resolution. *J Biol Chem 260*:4264–4275.

Blow DM. 1976. Structure and mechanism of chymotrypsin. *Acc Chem Res 9*:145–152.

Brünger AT. 1992. Free R value: A novel statistical quantity for assessing the accuracy of crystal structure. *Nature 355*:472–475.

Brünger AT. 1993. *X-PLOR: A system for X-ray crystallography and NMR*. New Haven, Connecticut: Yale University Press.

Carson M. 1991. Ribbons 2.0. *J Appl Crystallogr 24*:958–961.

Chen P, Tsuge H, Almassy RJ, Gribskov CL, Katoh S, Vanderpool DL, Margosiak SA, Pinko C, Matthews DA, Kan CC. 1996. Structure of the human cytomegalovirus protease catalytic domain reveals a novel serine protease fold and catalytic triad. *Cell 86*:835–843.

Choo Q-L, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M. 1989. Isolation of a cDNA clone derived from a blood-borne non-A non-B viral hepatitis genome. *Science 244*:359–362.

DeFrancesco R, Urbani A, Nardi MC, Tomei L, Steinkuehler C, Tramontano A. 1996. A zinc binding site in viral serine proteases. *Biochemistry 35*:13282–13287.

Eckart MR, Selby M, Masiarz F, Lee C, Berger K, Crawford K, Kuo C, Kuo G, Houghton M, Choo Q-L. 1993. The hepatitis C virus encodes a serine protease involved in the processing of the putative nonstructural proteins from the viral polyprotein precursor. *Biochem Biophys Res Commun 192*:399–406.

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

Furey W, Swaminathan S. 1997. PHASES-95: A program package for the processing and analysis of diffraction data from macromolecules. *Methods in Entymology 277*:590–620.

Grakoui A, Wychowski C, Lin C, Feinstone SM, Rice CM. 1993. Expression and identification of hepatitis C virus polyprotein cleavage products. *J Virol 67*:1385–1395.

Harrison SC. 1996. Peptide–surface association: The case of PDZ and PTB domains. *Cell 86*:341–343.

Hijikata M, Kato N, Ootsuyama Y, Nakagawa M, Shimotohno K. 1991. Gene mapping of the putative structural region of the hepatitis C virus. *Proc Natl Acad Sci USA 88*:5547–5551.

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

Hijikata M, Mizushima H, Tanji Y, Komoda Y, Hirowatari Y, Akagi T, Kato N, Kimura K, Shimotohno K. 1993b. Proteolytic processing and membrane association of putative nonstructural proteins of hepatitis C virus. *Proc Natl Acad Sci USA 90*:10773–10777.

Houghton M. 1996. Hepatitis C viruses. In: Fields BN, Knipe DM, Howley PM, eds. *Fields virology*, 3rd ed. New York: Raven Press. pp 1035–1058.

Huber R, Bode W. 1978. Structure basis of the activation and action of trypsin. *Acc Chem Res 11*:114–122.

Kleywegt GJ, Brünger AT. 1996. Checking your imagination: Applications of the free R value. *Structure 4*:897–904.

Kim JL, Morgenstern KA, Lin C, Fox T, Dwyer MD, Landro JA, Chamber, SP, Markland W, Lapre CA, O'melley ET, Harbeson SL, Rice CM, Murco MA, Caron PR, Thomso, JA. 1996. Crystal structure of the hepatitis C Virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. *Cell 87*:343–355.

Kraulis PJ. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr 24*:946–950.

Kuo G, Choo Q-L, Alter HJ, Gitnick GL, Redeker AG, Purcell RH, Miyamura T, Dienstag JL, Alter MJ, Stevens CE, Tagtmeier GE, Bonino F, Colombo M, Lee WS, Kuo C, Berger K, Shister JR, Overby LR, Bradley DW, Houghton M. 1989. An assay for circulating antibodies to a major etiologic virus of human non-A non-B hepatitis. *Science 244*:362–364.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr 26*:283–291.

Liddington RC, Yan Y, Moulai J, Sahli R, Benjamin TL, Harrison SC. 1991. Structure of simian virus 40 at 3.8 Å resolution. *Nature 354*:278–284.

Lin C, Pragai BM, Grakoui A, Xu J, Rice CM. 1994. Hepatitis C virus NS3 serine proteinase: Trans-cleavage requirements and processing kinetics. *J Virol 68*:8147–8157.

Lin C, Thomson JA, Rice CM. 1995. A central region in the hepatitis C virus NS4A protein allows formation of an active NS3–NS4A serine proteinase complex in vivo and in vitro. *J Virol 67*:4373–4380.

Love RA, Parge HE, Wickersham JA, Hostomsky Z, Habuka N, Moomaw EW, Adachi T, Hostomska Z. 1996. The crystal structure of the hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site. *Cell 87*:331–342.

Lovejoy B, Choe S, Cascio D, McRorie DK, DeGrado WF, Eisenberg D. 1993. Crystal structure of a synthetic triple-stranded α-helical bundle. *Science 259*:1288–1293.

Otwinowski Z, Minor W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol 276*:307–326.

Qiu X, Culp JS, Dilella AG, Hellmig B, Hoog SS, Janson CA, Smith WW, Abbel-Meguid SS. 1996. Unique fold and active site in cytomegalovirus protease. *Nature 383*:275–279.

Rice CM, Strauss EG, Strauss JH. 1986. In: Schlesinger S, Schlesinger MJ, eds. *Togaviridae and Flaviviridae*. New York: Plenum Press. pp 279–326.

Rost B, Casadio R, Fariselli P, Sander C. 1995. Transmembrane helices predicted at 95% accuracy. *Protein Sci 4*:521–533.

Sack JS. 1988. CHAIN: A crystallographic modeling program. *Mol Graphics 6*:224–225.

Saito I, Miyamura T, Obayashi A, Harada H, Katayama T, Kikuchi S, Watanabe Y, Koi S, Onji M, Ohata Y, Choo Q-L, Houghton M, Kuo G. 1990. Hepatitis C virus infection is associated with the development of hepatocellular carcinoma. *Proc Natl Acad Sci USA 87*:6547–6549.

Sheldrick GM. 1991. Heavy atom location using SHELXS-90. In: Wolf W, Evans PR, Leslie AGW, eds. *Isomorphous replacement and anomalous scattering*. Warrington, UK: SERC Daresbury Laboratory. Proceedings of the CCP4 study weekend 25–26 January 1991. pp 23–28.

Shieh HS, Kurumbail RG, Stevens AM, Stegeman RA, Sturman EJ, Pak JY, Wittwer AJ, Palmier MA, Wiegand RC, Holwerda BC, Stallings WC. 1996. Three-dimensional structure of human cytomegalovirus protease. *Nature 389*:279–282.

Shimizu Y, Kamaji K, Masuho Y, Yokota T, Inoue H, Sudo K, Satoh S, Shimotohno K. 1996. Identification of the sequence on NS4A required for enhanced cleavage of the NS5A/5B site by hepatitis C virus NS3 protease. *J Virol 70*:127–132.

Takamizawa A, Mori C, Fuke I, Mannabe S, Murakami S, Fujita J, Onoshi E, Andoh T, Yoshida I, Okayama H. 1991. Structure and organization of the

[6]The coordinates can be found in the Brookhaven Protein Data Bank under the code: 1jxp.

hepatitis C virus genome isolated from human carriers. *J Virol 65*:1105–1113.

Tanji Y, Hijikata M, Satoh S, Kaneko T, Shimotohno K. 1995. Hepatitis C virus-encoded nonstructural protein NS4A has versatile functions in viral protein processing. *J Virol 69*:1575–1581.

Terwilliger TC, Eisenberg D. 1983. Unbiased three-dimensional refinement of heavy-atom parameters by correlation of origin-removed Patterson functions. *Acta crystallogr A39*:813–817.

Tomei L, Failla C, Santolini E, De Francesco R, La Monica N. 1993. NS3 is a serine protease required for processing hepatitis C virus polyprotein. *J Virol 67*:4017–4026.

Tomei L, Failla C, Vitale RL, Bianchi E, De Francesco R. 1996. A central hydrophobic domain of the hepatitis C virus NS4A protein is necessary and sufficient for the activation of the NS3 protease. *J Gen Virol 77*:1065–1070.

Tong L, Wengler G, Rossman MG. 1993. Refined structure of Sindbis virus core protein and comparison with other chymotrypsin-like serine proteinase structures. *J Mol Biol 230*:228–247.

Tong L, Qian C, Massariol MJ, Bonneau PR, Cordingley MG, Lagace L. 1996. A new serine-protease fold revealed by the crystal structure of human cytomegalovirus protease. *Nature 383*:272–275.

Yan Y, Winograd E, Viel A, Cronin T, Harrison SC, Branton D. 1993. Crystal structure of the repetitive segments of spectrin. *Science 262*:2027–2030.

Yan Y, Stehle T, Liddington RC, Zhao H, Harrison SC. 1995. Structure determination of simian virus 40 and murine polyomavirus by a combination of 30-fold and 5-fold electron-density averaging. *Structure 4*:157–164.

Yan Y, Munshi S, Li Y, Sardana V, Blue, J, Johns B, Cole J, Steinkuehler C, Tomei L, De Francesco R, Kuo LC, Chen Z. 1996. Crystal structure of hepatitis C virus (HCV) NS3 protease–NS4A$_{21-34}$ complex at 2.8 Å resolution. Proceedings of the 4th International Meeting on Hepatitis C Virus and Related Viruses, Kyoto, Japan. p 8.