Cell, Vol. 87, 343–355, October 18, 1996, Copyright ©1996 by Cell Press

Crystal Structure of the Hepatitis C Virus NS3 Protease Domain Complexed with a Synthetic NS4A Cofactor Peptide

J.L. Kim,* K. A. Morgenstern,* C. Lin,* T. Fox,* M.D. Dwyer,* J.A. Landro,* S.P. Chambers,* W. Markland,* C.A. Lepre,* E.T. O'Malley,* S.L. Harbeson,* C.M. Rice,† M.A. Murcko,* P.R. Caron,* and J.A. Thomson*

*Vertex Pharmaceuticals Incorporated
130 Waverly Street
Cambridge, Massachusetts 02139-4242

†Department of Molecular Microbiology
Washington University School of Medicine
St. Louis, Missouri 63110-1093

Summary

An estimated 1% of the global human population is infected by hepatitis C viruses (HCVs), and there are no broadly effective treatments for the debilitating progression of chronic hepatitis C. A serine protease located within the HCV NS3 protein processes the viral polyprotein at four specific sites and is considered essential for replication. Thus, it emerges as an attractive target for drug design. We report here the 2.5 Å resolution X-ray crystal structure of the NS3 protease domain complexed with a synthetic NS4A activator peptide. The protease has a chymotrypsin-like fold and features a tetrahedrally coordinated metal ion distal to the active site. The NS4A peptide intercalates within a β sheet of the enzyme core.

Introduction

Infection by hepatitis C viruses (HCVs) is a compelling human medical problem. The virus was formally identified by molecular cloning and sequencing around 1989 (Alter et al., 1989; Choo et al., 1989; Kuo et al., 1989). HCV has now emerged as the causative agent for most cases of non-A, non-B hepatitis, with an estimated human seroprevalence of 1% globally (Purcell, 1994; Van der Poel, 1994). Four million individuals may be infected in the United States alone (Alter and Mast, 1994). Upon first exposure to HCV, approximately 20% of infected individuals develop acute clinical hepatitis while others appear to resolve the infection spontaneously. However, in most instances the virus establishes a chronic infection that persists for decades (Iwarson, 1994). This usually results in recurrent and progressively worsening liver inflammation, often leading to more severe disease states such as cirrhosis and hepatocellular carcinoma (Saito et al., 1990; Kew, 1994). The prospects for effective anti-HCV vaccines remain uncertain and the only established therapy for HCV disease is interferon treatment. However, interferons have significant side effects (Renault and Hoofnagle, 1989; Janssen et al., 1994) and induce long term remission in only a fraction (~25%) of cases (Weiland, 1994). For these reasons, there is considerable interest in developing more effective anti-HCV therapies.

Early analyses of the HCV single-positive strand RNA genome indicated a relationship to the flaviviruses and pestiviruses (Miller and Purcell, 1990; Choo et al., 1991), and HCV was assigned a new genus in the family Flaviviridae (Francki et al., 1991). More recently it has become clear that there are multiple HCV genotypes (Bukh et al., 1993; Simmonds, 1994) that all share the same essential features. The HCV genome is about 9.4 kb in length and consists of a highly conserved 5′ untranslated region followed by a single open reading frame that encodes a polyprotein of 3010 to 3033 amino acids (Kato et al., 1990; Choo et al., 1991; Takamizawa et al., 1991). All known HCV polyprotein sequences share at least 71% identity. The structural proteins (envelope and core components) are clustered in the N-terminal portion of the polyprotein, followed by the nonstructural (NS) proteins, which represent the essential catalytic machinery for viral replication (Hijikata et al., 1991; Bartenschlager et al., 1993; Grakoui et al., 1993a, 1993c; Tomei et al., 1993). The polyprotein architecture is shown in Figure 1, together with a summary of the proteolytic processing events that generate the mature viral proteins.

In vivo processing of the HCV structural proteins is probably facilitated by host cell signal peptidases associated with the lumen of the endoplasmic reticulum (Hijikata et al., 1991; Lin et al., 1994a; Mizushima et al., 1994). In contrast, processing of the nonstructural proteins seems to be orchestrated by two viral gene products. The NS2/NS3 junction is cleaved by a zinc-dependent protease associated with NS2 and the N-terminus of NS3 (Grakoui et al., 1993b; Hijikata et al., 1993a). The remaining four cleavages (at the NS3/NS4A, NS4A/NS4B, NS4B/NS5A, and NS5A/NS5B sites) are mediated by a separate serine protease that resides in the N-terminal one-third of NS3 (Bartenschlager et al., 1993; Eckart et al., 1993; Grakoui et al., 1993a, 1993b, 1993c; Hijikata et al., 1993a, 1993b; Tomei et al., 1993; Manabe et al., 1994). Questions remain concerning the functions of NS4B and NS5A, but the roles of the other nonstructural proteins are becoming clearer. NS5B appears to be the viral RNA polymerase (Takamizawa et al., 1991; Behrens et al., 1996) and the C-terminal two-thirds of NS3 an NTP-dependent helicase (Miller and Purcell, 1990; Suzich et al., 1993; Jin and Peterson, 1995; Kanai et al., 1995; Kim et al., 1995). NS4A is a 54-residue amphipathic peptide, with a hydrophobic N-terminus and a hydrophilic C-terminus (Failla et al., 1994). It appears to serve multiple functions: acting as a cofactor for the NS3 protease and possibly assisting in the membrane-localization of NS3 and other viral replicase components (Lin et al., 1995; Tanji et al., 1995a; Shimizu et al., 1996).

Great efforts have been made to characterize the mechanisms by which the NS3 serine protease domain processes the downstream gene products. To summarize, the NS3/NS4A cleavage appears to occur in *cis*, whereas the NS4A/NS4B, NS4B/NS5A and NS5A/NS5B processing events probably occur in *trans*. NS4A seems critical to this scheme of events, enhancing the proteolytic efficiency at all of the sites. Many deletion mapping

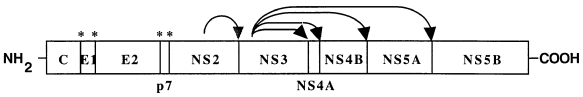

Figure 1. HCV Polyprotein Processing

The locations of the HCV structural and nonstructural proteins are marked on a diagram of the 3011 amino acid polyprotein. Cleavages between the structural proteins by cellular peptidases are marked by asterisks. Cleavage between NS2 and NS3 is mediated by the NS2/NS3 protease. The NS3 protease is resposible for cleavages between NS3 and NS4A, NS4A and NS4B, NS4B and NS5A, and NS5A and NS5B.

and mutagenesis experiments have been performed recently to probe the regions of NS3 and NS4A that interact. However, to date, there is no direct structural information about either component.

Here we present the X-ray crystal structure, at 2.5 Å resolution, of a recombinant truncated HCV NS3 protease domain (tNS3) complexed with a synthetic peptide that encompasses the essential NS3-binding region of NS4A (Lin et al., 1995). We believe that the information provided by this structure will have far-reaching consequences for the global effort to develop more effective HCV therapies.

Results and Discussion

Structure Determination and Analysis

The structure of the truncated NS3 protease domain:NS4A peptide (tNS3:NS4A) complex was determined at 2.7 Å resolution by multiple isomorphous replacement and is presently refined at 2.5 Å resolution (Table 1). The rhombohedral crystals used in this study (space group R32; a = b = 225.0 Å, c = 75.5 Å) contain two tNS3:NS4A complexes per asymmetric unit. The refined model consists of residues 2–180 of NS3 (residues 1028–1206 of the HCV polyprotein) and residues 21–39 of NS4A (residues 1678–1696 of the polyprotein) in the first complex (complex A), NS3 residues 29–180 (1055–1206) and NS4A residues 21–36 in the second complex (complex B), 2 zinc atoms, and 130 water molecules. The N-terminal 28 residues of tNS3 (excluding the N-terminal T7 tag) are disordered in complex B and have not been modeled. The current R value is 21.6% (free R value = 26.1%) for data between 6.0 and 2.5 Å with $|F| > 1.0\sigma|F|$. Representative electron density for the refined model is shown in Figure 2. Atomic temperature factors for the two complexes are closely matched (average B = 29.2 Å^2 and 29.8 Å^2, respectively), and neither complex contains nonglycine main-chain torsion angle values located in unfavorable regions of a Ramachandran plot. Apart from the disordered N-terminus in complex B, the two tNS3:NS4A complexes in the asymmetric unit are very similar, with a root-mean-square deviation (rmsd) of 0.6 Å for 159 equivalent Cα positions (polyprotein residues 1057–1205 from NS3 and residues 22–31 of NS4A). The similarity is even greater within the core of the complex, increasing our confidence that this structure is representative of the protease domain of the full-length NS3 protein.

Structure of tNS3

The three-dimensional structure of the tNS3:NS4A complex reveals that the HCV NS3 serine protease domain adopts a chymotrypsin-like fold. The complex consists of two structural domains, each containing a twisted β sheet incorporating a "Greek key" motif (Figure 3). The C-terminal domain (residues 1120–1206) contains the conventional six-stranded β barrel, common to nearly all members of the chymotrypsin family, followed by a structurally conserved α helix. The core of this barrel is

Table 1. Data Collection and Refinement Statistics

Data set	Native	$HgCl_2$(1)	$HgCl_2$(2)	HgFuran	PCMBS	K_2PtCl_4(1)	K_2PtCl_4(2)	UO_2OAc_2
Resolution (Å)	2.5	2.9	2.8	3.0	2.8	3.0	3.0	2.7
Unique reflections	24,236	15,624	17,410	14,032	16,951	13,822	15,630	18,611
Redundancy	2.9	2.5	2.6	2.8	2.1	1.7	1.9	1.8
Completeness (%)	96	96	97	95	94	94	96	93
R_{sym} (%)	5.7	8.9	7.3	9.2	7.7	5.8	5.8	5.7
MIR analysis:								
Resolution (Å)	3.1	3.1	3.3	3.1	3.2	3.6	3.6	3.6
R_{iso} (%)		13.3	6.9	12.2	8.6	8.6	12.0	9.1
Number of sites		5	4	5	3	4	4	3
Phasing power		2.34	1.67	2.36	1.39	1.45	1.33	0.82
R_{cullis}		0.52	0.60	0.54	0.61	0.60	0.62	0.67
Mean overall figure of merit	0.62							
Refinement Statistics:								
Resolution (Å)	3.93	3.28	2.92	2.68	2.50	Total		
R factor (%)	17.9	20.0	22.9	26.9	32.1	21.6		
R_{free}[a] (%)	23.2	24.8	26.8	31.4	35.5	26.1		
Reflections ($F>1.0\sigma F$)	4,195	4,131	3,978	3,805	3,597	21,635		
Nonhydrogen atoms						2,815		
rms bond lengths (Å)						0.007		
rms bond angles (°)						1.48		

$R_{sym} = \Sigma|I - \langle I\rangle|/\Sigma I$, where I = observed intensity, $\langle I\rangle$ = average intensity obtained from multiple observations of symmetry related reflections.
$R_{iso} = \Sigma||F_{PH}|-|F_P||/\Sigma|F_P|$, where $|F_P|$ = protein structure factor amplitude, $|F_{PH}|$ = heavy-atom derivative structure factor amplitude.
Phasing power = rms ($|F_H|/E$), where $|F_H|$ = heavy-atom structure factor amplitude and E = residual lack of closure.
[a] Free R value was calculated using 9% of the data.

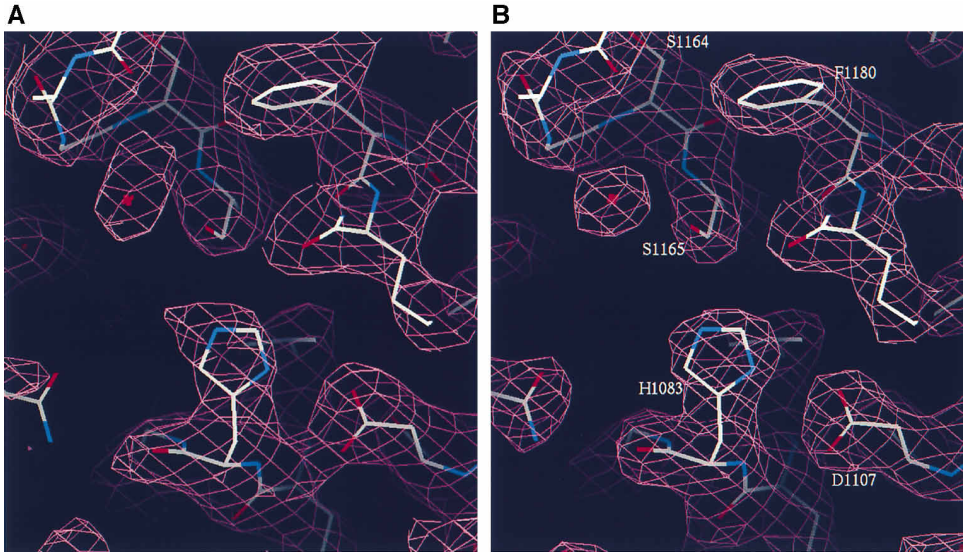

Figure 2. Electron Density Maps Around the Active Site of tNS3:NS4A

(A) Electron density map at 2.7 Å, calculated using density modified phases. The map is contoured at 1.2σ and superimposed on the refined coordinates constituting the active site.
(B) $2F_o$-F_c electron density map at 2.5 Å, calculated using phases from the refined model. The map is contoured at 3.0σ. Atoms are color coded by element type: carbon in white, nitrogen in blue, and oxygen in red. Active site residues His-1083, Asp-1107, and Ser-1165 are labeled, as well as Phe-1180, which defines part of the P1 substrate-binding pocket. A well-ordered water molecule occupies the oxyanion hole.

packed with hydrophobic residues that emanate from all six strands of the β barrel. The N-terminal domain (residues 1027–1119) contains eight β strands rather than six, including one strand contributed by NS4A. This array of β strands gives rise to a β sheet that superimposes with most of the distorted barrel found in the N-terminus of chymotrypsin. Active site residues His-1083 and Asp-1107 reside in the N-terminal domain of the tNS3 protease, whereas Ser-1165 is located in the C-terminal domain. All three residues are situated in a cleft separating the two domains. Additionally, the C-terminal domain contains a tetrahedrally coordinated metal ion, modeled here as a zinc, located at one end of the β barrel. This zinc ion appears to play a structural role in the enzyme by stabilizing the loops preceeding β strands A_2 and E_2.

Structural comparisons between tNS3:NS4A and other serine proteases currently in the Protein Data Bank (PDB) indicate that it shares the greatest similarity with the Sindbis virus core protein (SCP) (Tong et al., 1993), with an rmsd of 1.6 Å for 75 Cα residues that comprise the core and active site of each enzyme (Figure 4B). The structural agreement is better in the C-terminal domain, where the six β strands from each enzyme superimpose reasonably well. Similarly, the six β strands in the C-terminal domain of chymotrypsin closely match those in tNS3 (Figure 4C). On the other hand, only β strands A_1, B_1, C_1, and F_1 in the N-terminal domain of tNS3:NS4A superimpose with the equivalent strands in either SCP or chymotrypsin. The β strand formed by the NS4A peptide is in the vicinity of strand D_1 of chymotrypsin in their structural alignment, but this strand does not hydrogen-bond to strand E_1 as is found in the β barrel of chymotrypsin. The reported structure of SCP is missing over 100 residues at its N-terminus, so it is unclear whether this protein has any structural equivalents to NS4A, β strand A_0, or helix α_0 of tNS3:NS4A (Figure 5).

Interactions with NS4A

The NS4A peptide forms a β strand that lies between strands A_0 and A_1 of tNS3 (Figure 3B). All but two of the main-chain carbonyl and amide groups of residues Val-23 to Leu-31 of NS4A form hydrogen bonds with either the N-terminal region of tNS3 or strand A_1. This includes hydrogen bonds to the side chains of Arg-1037 and Glu-1058. Several side chains from NS4A are buried in hydrophobic pockets formed by tNS3 and contribute to the hydrophobic core of the N-terminal domain (Figure 6A). A total of 2400 $Å^2$ of surface area is buried by the interaction of NS4A with tNS3. Thus, NS4A should be considered an integral structural component of the complex and probably plays a significant role in stabilizing tNS3.

A wealth of biochemical and mutational data exists characterizing the interaction of NS4A with the NS3 protease domain. Multiple experiments have demonstrated that a central hydrophobic region of NS4A, encompassing residues Gly-21 to Lys-34, is sufficient for NS3 binding and protease activation. Indeed, a synthetic peptide corresponding to this region substitutes for longer NS4A sequences in NS3 protease activation (Lin et al., 1995). The structure of the tNS3:NS4A complex provides an explanation for these results. All of the contacts observed in the complex occur in the region spanning residues 21 to 32 of NS4A. Interestingly, the requirement for only the central region of NS4A for full protease activation has been demonstrated for the full-length NS3 enzyme as well as the isolated protease domain (Shimizu

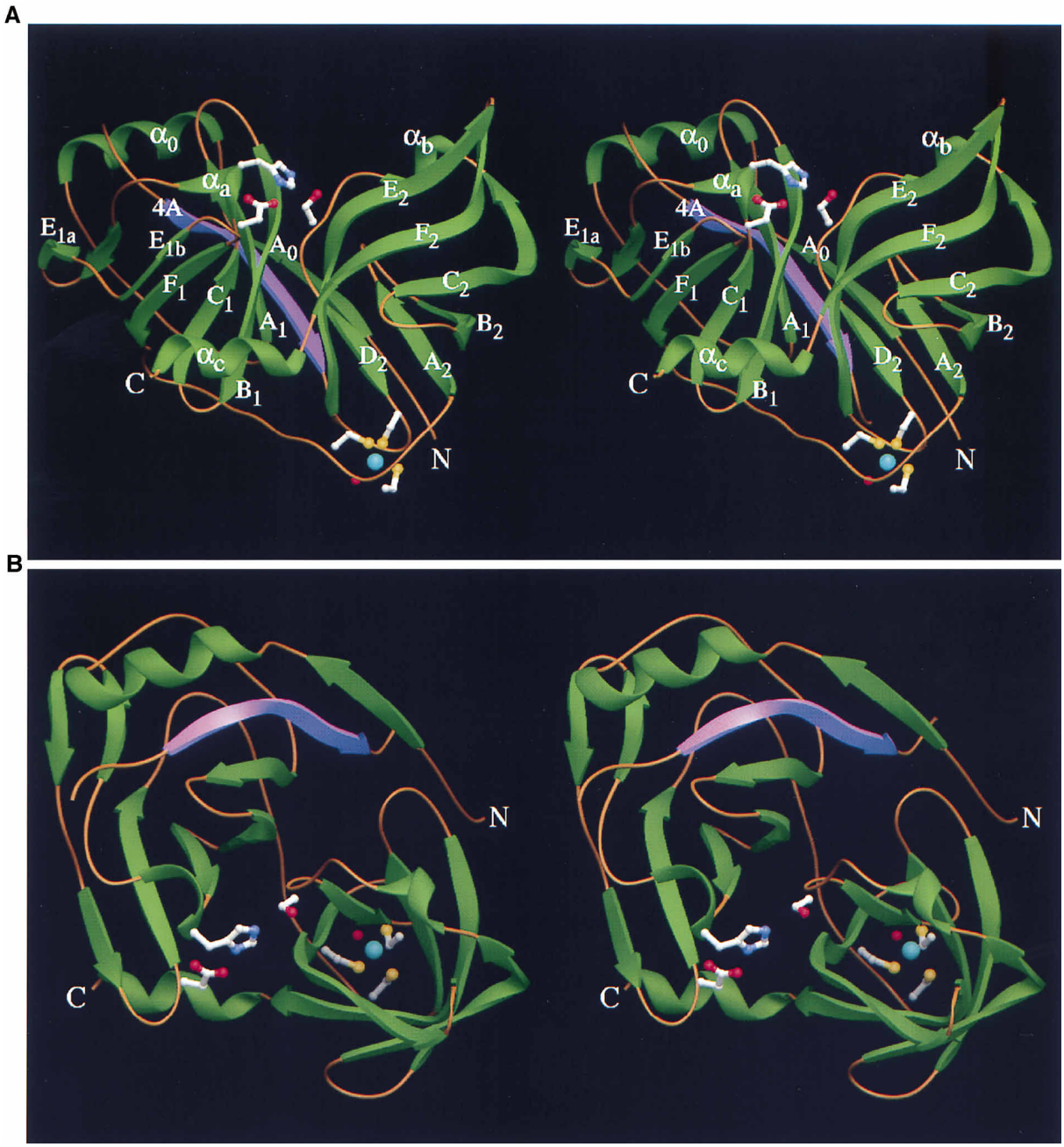

Figure 3. Stereo Ribbon Diagrams of the tNS3:NS4A Complex

(A) View into the active site cleft of the enzyme. The N-terminal domain of the complex is on the left and the C-terminal domain on the right. Secondary structural elements are labeled according to the convention used for chymotrypsin. Side-chains of active site residues His-1083, Asp-1107, and Ser-1165, along with Zn^{2+} ligands Cys-1123, Cys-1125, and Cys-1171 are displayed in ball-and-stick representation. Zn^{2+} is colored cyan and its H_2O ligand red. The β strand formed by NS4A is shown in magenta.

(B) View from above the active site, approximately 90° from (A). This view is approximately down the axis of the C-terminal β barrel, with the tetrahedrally coordinated Zn^{2+} at the bottom. The location of the NS4A peptide within the fold of the enzyme is readily apparent. This figure was generated using Ribbons (Carson, 1991).

et al., 1996), suggesting that the same interactions observed here occur in the full-length NS3:NS4A complex.

The interactions of NS4A with tNS3 can be divided into two components: those with the N-terminal 30 residues of tNS3 and those with the core of tNS3. The former are found in only one of the two tNS3:NS4A complexes in the asymmetric unit, whereas the latter interactions are found in both. Despite this difference, the structure of NS4A is nearly identical in complex A and B. Mutational analyses within the central region of NS4A have identified several residues that are critical for complex formation with NS3 and protease activation (Bartenschlager et al., 1995; Lin et al., 1995; Shimizu et al., 1996). A strong correlation exists between the amount

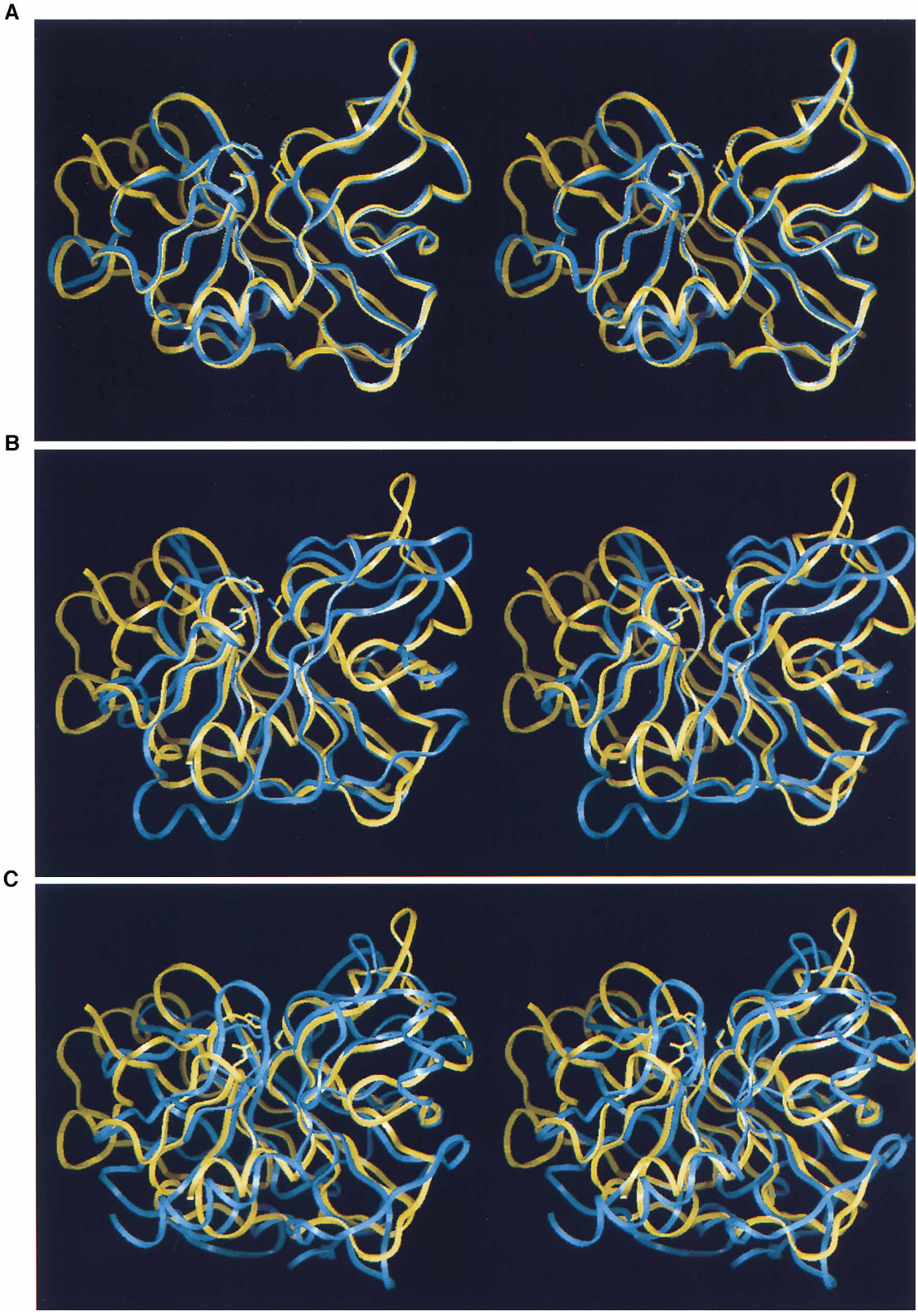

Figure 4. Structural Conservation between HCV Protease and Chymotrypsin-Fold Proteases

The structure of HCV tNS3 monomer A (yellow) was superposed on monomer B in (A), Sindbis virus capsid protease (PDB entry 2SNV) in (B), and chymotrypsin (PDB entry 5CHA) in (C). The orientation of the molecules is similar to that in Figure 2A. The side chains of the catalytic triads are also shown.

```
          A0        α0              A1          B1       C1  *αa
         <-->    <-------->      <--->      <------>   <--><-->
NS3   PITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAT..QTFLATCIN.GVCWTVYHGAGT 1087
2SNV  ..............................RLFDVKNEDGDVIGHALAME.GKVMKPLHVKG.  144
5CHA        CGVPAIQPVLSIVNGEEAVPGSWPWQVSLQDKTGFHFCGGSLINENWVVTAAHCGVT   61

        D1                    E1              *  F1
       <->            <-->  <->   <->         <--->
NS3   R...............TIASPKGPVIQMYTNV......DQDLVGWPAPQG......SRSLTP 1122
2snv  ..TIDH.............PVLSKLKFT.KSS......AYDMEFAQLPVNMR....SEAFTY  180
5cha  TSDVVVAGEFDQGSSSEKIQKLKIAKVFK.NSKYNSLTINNDITLLKLSTAASFSQTVSAVCL  123
NS4A  GSVVIVGRIVLSGKPAIIP

                A2                B2                        C2
               <--->             <---->                   <------->
NS3   CTC.....GSSDLYLVTR........HADVIPVRRRG..............DSRGSLLSPRP 1157
2SNV  TSE....HPEGFYNWHH.............GAVQYSGG..............RFTIPR...  207
5CHA  PSASDDFAAGTTCVTTGWGLTRYTPDRLQQASLPLLSNTNCKKYWGTKIKDAMICAGASG...  187

      αb    *  D2          E2          F2      αc
     <-->     <->       <-------->   <------><----->
NS3   ISYLKGSSGGPLLCPA...GHAVGLFRAAVCTRGVTKAVDFIPVENLETTMR            1206
2SNV  GVGGRGDSGRPIMDNS...GRVVAIVLGGADE.GTRTALSVVTWNSKGKTIKTTPEGTEEW    264
5CHA  VSSCMGDSGGPLVCKKNGAWTLVGIVSWGSSTCSTSTPGVYARVTALVNWVQQTLAAN       245
```

Figure 5. Structure-Based Sequence Alignment of HCV tNS3, Sindbis capsid, and Chymotrypsin

Residues in bold have a Cα rmsd of less than 2 Å in the overlay of all three structures. The catalytic triad residues are denoted by asterisks. Ligands to the zinc ion in HCV are underlined. The secondary structure elements are marked above the alignment and are labeled consistent with chymotrypsin notation. NS4A is located in the approximate location of strand D_1 in chymotrypsin.

of buried surface area of a given NS4A residue in the complex and the severity of the effect of mutating that residue on NS3 protease activation and complex stability (Figure 6B). Those NS4A residues that are buried in the core of tNS3, Val-23, Ile-25, Ile-29, and Leu-31, are most sensitive to mutation. The only exception is Gly-27, whose mutation to serine has a severe effect (Shimizu et al., 1996) despite its small buried surface area. Analysis of the structure indicates that this absolutely conserved glycine is situated such that mutations would result in steric clashes with the backbone of NS3 residue 1035.

The Role of NS4A

The activation of the NS3 protease by NS4A may be a result of NS4A incorporation into the N-terminal domain β sheet, which is extended by up to two strands in the complex. This could lead to alterations in the active site by providing a more rigid and precise framework for residues that form the "prime-side" substrate-binding channel. In addition, activation via direct interactions of NS4A with the prime-side residues of the substrate are possible, although these would occur more than 10 Å away from the cleavage site. This is in contrast to the mechanism of activation in chymotrypsin, where proteolytic processing of the polypeptide chain leads to the creation of a new N-terminus that interacts with "non-prime-side" of the substrate-binding channel. This results in the formation of a cavity that can accommodate the P1 residue of the substrate (see Appel, 1986, for review). The tNS3 protease has an inherently open conformation in the analogous region. Other serine proteases such as α lytic protease and subtilisin require a propeptide, which can function in *trans*, to obtain full activity (Li et al., 1995). These propeptides are believed to act as molecular chaperones by acting as competitive active site inhibitors around which the enzyme folds.

Experiments both in vivo and in vitro indicate that NS4A promotes the membrane association of NS3 and stabilizes it from degradation by cellular proteases (Hijikata et al., 1993b). In addition, NS4A has been found to stimulate the phosphorylation of NS5A, which is also believed to be membrane-associated (Tanji et al., 1995b). Hydrophobicity analysis of the 54 amino acid NS4A protein suggests that the N-terminus may associate with membranes (Failla et al., 1994; Bartenschlager et al., 1995; Tanji et al., 1995a). More detailed analyses using multiple sequence alignments (Rost et al., 1995) support the hydrophobicity plots and predict that the N-terminal 20 residues of NS4A form a transmembrane helix. The NS3 protease activation region immediately follows this, suggesting that NS4A holds the NS3 protease domain very close to the membrane and may serve as a molecular tether that anchors the HCV replication machinery complex together at the cellular membrane.

tNS3 N-Terminal Region

As shown in Figure 4A, the two molecules of tNS3 in the asymmetric unit are very similar, with a Cα rmsd of 0.6 Å for residues 1057–1206 and a side-chain rmsd of 1.4 Å over this same range. They differ mainly at the N-terminus with one molecule having well-defined structure extending to residue 1028 and the other only to residue 1055. In spite of this difference, both complexes contain NS4A in essentially the same position, with a Cα rmsd of 0.8 Å for residues 21–31. NS4A becomes part of the N-terminal domain β sheet and is located in approximately the same location as strand D_1 in SCP and chymotrypsin. In complex A, the N-terminus of NS3 forms a β strand (A_0), an α helix (α_0), and an extended chain leading into strand A_1. This region is stabilized by main-chain hydrogen bonds between residues 1030–1034 and NS4A, along with hydrophobic interactions between α_0 and Val-24 and Val-26 of NS4A. Additionally, a salt bridge between residues flanking α_0 (Arg-1037

(B) Surface area of NS4A in contact with NS3. The surface area of each residue of NS4A buried by tNS3 is plotted. The solid line represents contacts found in monomer B in which the N-terminal 28 residues of NS3 are disordered. The dotted line depicts contacts in monomer A and thus includes contacts with the NS3 N-terminus. Replacement of residues Val-23, Ile-25, Gly-27, Arg-28, Ile-29, and Ile-31 by alanine, serine, or aspartate cripples NS4A (Bartenschlager et al., 1995; Lin et al, 1995; Shimizu et al., 1996). In contrast, replacement of more exposed residues Gly-21, Ser-22, Val-24, Val-26, Val-30, Ser-32, or Gly-33 has minimal effect.

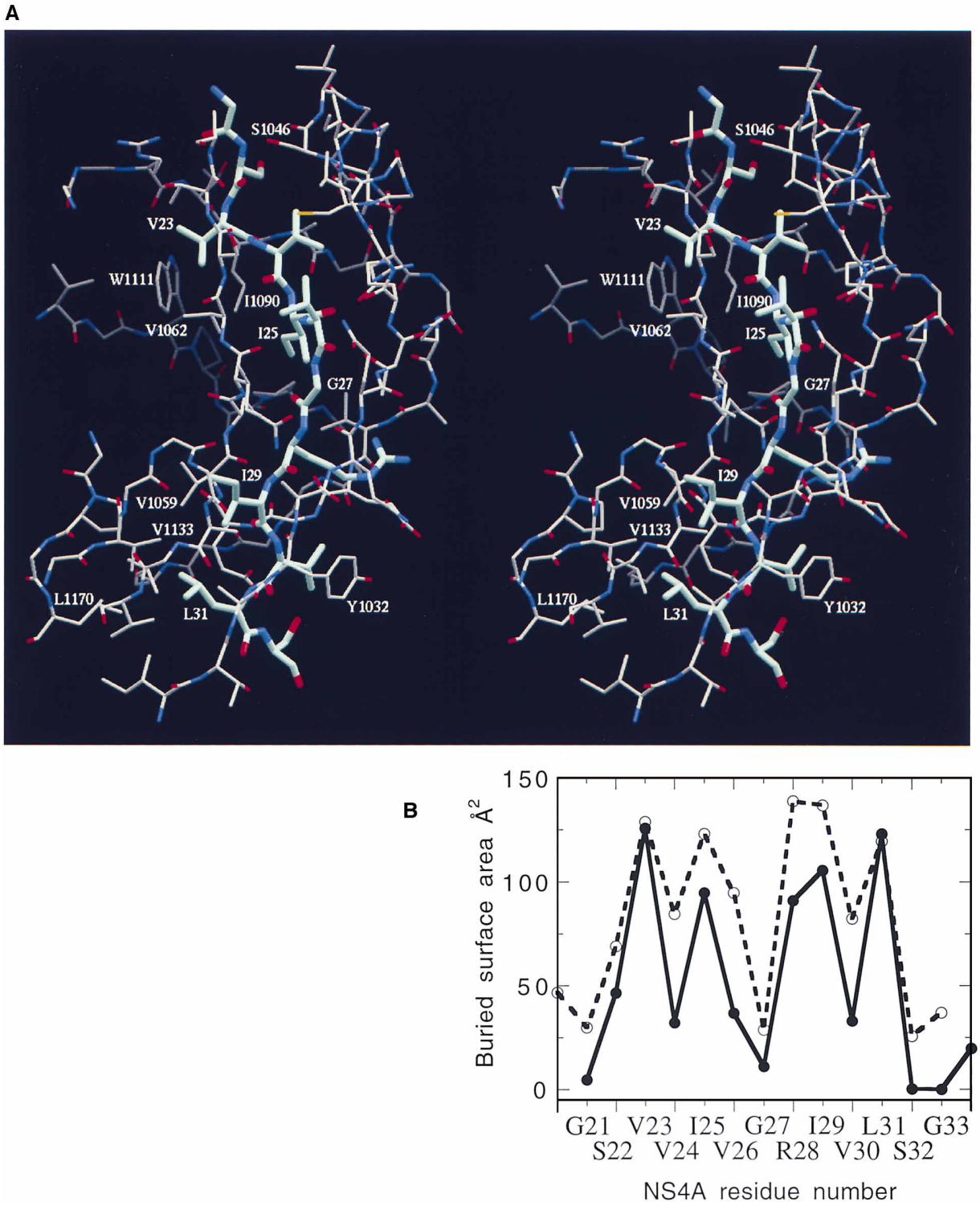

Figure 6. NS4A-tNS3 Interactions

(A) Residues within the essential region of NS4A (thick bonds) make extensive hydrophobic contacts with tNS3. Val-23 is buried in a hydrophobic pocket formed by side-chain atoms of Val-1062, Thr-1064, Ala-1085, Arg-1088 (Cg), Ile-1090, and Trp-1111. Ile-29 is located in a hydrophobic pocket formed by Ala-1031, Ala-1033, Val-1059, Ile-1061, Leu-1070, Val-1133, Ala-1137, and NS4A side-chain Leu-31. Ile-25 and Leu-31 also make extensive hydrophobic contacts with the core of tNS3, while Ser-22, Val-24, Val-26, Arg-28, and Val-30 interact with residues Ile-1029 to Thr-1045 in the N-terminal region of tNS3. NS4A also forms numerous main-chain hydrogen bonds with strands A_0 and A_1. This figure was generated using Ribbons (Carson, 1991).

(*Legend for Figure 6 continued on previous page*)

and Asp-1051) helps stabilize the fold of this region and direct the polypeptide chain toward strand A_1.

NS4A binding significantly inhibits cellular degradation of NS3 (Hijikata et al., 1993b; Tanji et al., 1995a), and this may result from NS4A ordering the NS3 N-terminal region. The interaction of NS4A peptide with the core of NS3 buries 1650 Å^2 of surface area while the interaction with the N-terminal region buries an additional 745 Å^2. The removal of NS4A would cause disruption of the N-terminal domain β sheet and would likely result in either disordering of the N-terminal region or repacking of this region against the core of tNS3, as they currently share only 250 Å^2 of buried surface area. The stabilization of the N-terminus of tNS3 by a cofactor may be similar to the stabilization of the N-terminus of trypsin by calcium, which binds to a loop between strands D_1 and E_1 (Gomez et al., 1977).

Deletions of the N-terminus of NS3 have shown that this region is important for complex formation with NS4A. NS3 constructs beginning at the true N-terminus (Ala-1027) can be activated and coimmunoprecipitated with NS4A as shown by in vitro translation and cell culture experiments. NS3 N-terminal truncations of 7–15 residues (constructs beginning between amino acids 1034 and 1042) result in products that can still be activated by NS4A but appear less competent in complex formation as judged by immunoprecipitation experiments (Bartenschlager et al., 1995; Failla et al., 1995; Koch et al., 1996). These N-terminal residues of NS3 encompass β strand A_0, which packs onto NS4A, and removing these residues reduces the total amount of tNS3 surface area contact with NS4A by approximately one-third. NS3 constructs that begin at residue 1048 or 1049 are appreciably less stable, and in most cases cannot be activated by NS4A (Satoh et al., 1995; Tanji et al., 1995a; Koch et al., 1996). In these proteins strand A_0 and helix α_0 are missing. Of note is the deletion of Ser-1046, which is buried against NS4A in the complex and has been shown by mutagenesis to be critical for NS4A activation (Koch et al., 1996). Replacement of this residue by a hydrophobic residue, such as valine, results in a protease that can still be activated by NS4A, whereas replacement with a glutamic acid residue eliminates the NS4A activation. This is in good agreement with the structure. Proteases truncated at residue 1049 are also inherently unstable. This is likely due to the lack of significant interactions of residues 1049–1055 with either tNS3 or NS4A peptide. Truncations to residue 1055 retain some residual activity, but are not activated by NS4A (Failla et al., 1995).

Peptide Activation of Other Proteases

The activation of the HCV NS3 protease by a short peptide fragment of NS4A is reminiscent of the activation mechanism of the human adenovirus-2 L3 23K protease (AVP) by an 11 amino acid cofactor encoded by the C-terminus of the pVI protein (pVIc) (Webster et al., 1993). In AVP, the peptide binds to form a 1:1 complex and increases k_{cat} by 350-fold (Mangel et al., 1996). The structure of the AVP-pVIc complex demonstrated that this 204 residue cysteine protease has a unique fold that is grossly similar to those of papain and subtilisin (Ding et al., 1996). The pVIc peptide binds at the surface of the AVP, quite distant from the active site, and forms the sixth strand of a β sheet that comprises the core of the enzyme. It has been proposed that the activation of AVP by pVIc is due to alterations in the conformation of the active site, propagated by interactions of the peptide with noncontiguous segments of the protease, analogous to what we have proposed for tNS3 activation by NS4A.

Sequence analysis and mutagenesis experiments indicate that a chymotrypsin-like serine protease resides within the NS3 N-terminal regions of other flaviviruses (Bazan and Fletterick, 1989). Homology modeling of these proteases suggests that they have similar overall folds to that of the HCV NS3 protease (P. R. C., unpublished data). However, flaviviruses require NS2B rather than NS4A for protease activation (Chambers et al., 1993; Falgout et al., 1993), and there is no sequence similarity between the HCV NS4A activation region and NS2B. Therefore, elucidation of the activation mechanisms of these viral proteases will require further biochemical or structural characterization.

Zinc-Binding Site

The HCV NS3 protease structure has a metal ion tetrahedrally coordinated by Cys-1123, Cys-1125, Cys-1171, and through a water molecule to His-1175 (Figures 3 and 7). This metal was originally modeled as zinc, based on comparisons of its surrounding architecture with zinc-binding motifs found in the PDB. Preliminary results obtained via atomic absorption spectroscopy confirm the presence of zinc in samples of tNS3:NS4A (details will be published elsewhere; T. F., M. D. D., K. A. M., J. A. T., unpublished data). The existence of this binding site was previously predicted from sequence analysis and model building (P. R. C., unpublished data). The ligands to the zinc are conserved among the known HCV and hepatitis G virus (HGV) NS3 sequences, but there is no discernible evidence for metal-binding motifs in the analogous regions of other flavivirus proteases (P. R. C., unpublished data). Individual mutations of any of the HCV NS3 metal-coordinating cysteine residues to alanine greatly reduces protease activity, whereas mutation of the histidine has much less effect (Hijikata et al., 1993a; S. P. C. et al., unpublished data). This is consistent with the indirect interaction of His-1175 with the zinc ion. An essential zinc has previously been described for rhinovirus 2A protease (Sommergruber et al., 1994). This protein is believed to also have a chymotrypsin-like fold, although the nucleophilic residue is a cysteine rather than a serine. Sequence alignments of rhinovirus 2A protease and the corresponding regions in other picornaviruses suggests that a zinc could be coordinated by three cysteines and a histidine in approximately the same locations as those seen in HCV NS3 protease. Mutations of the corresponding zinc ligands proposed in poliovirus 2A protease also reduce activity (Yu and Lloyd, 1992).

No clear evidence has been found for zinc-activation of the HCV NS3 protease activity in vitro, and EDTA appears to be a very weak inhibitor of NS3 protease activity (Lin and Rice, 1995), suggesting that the ion is

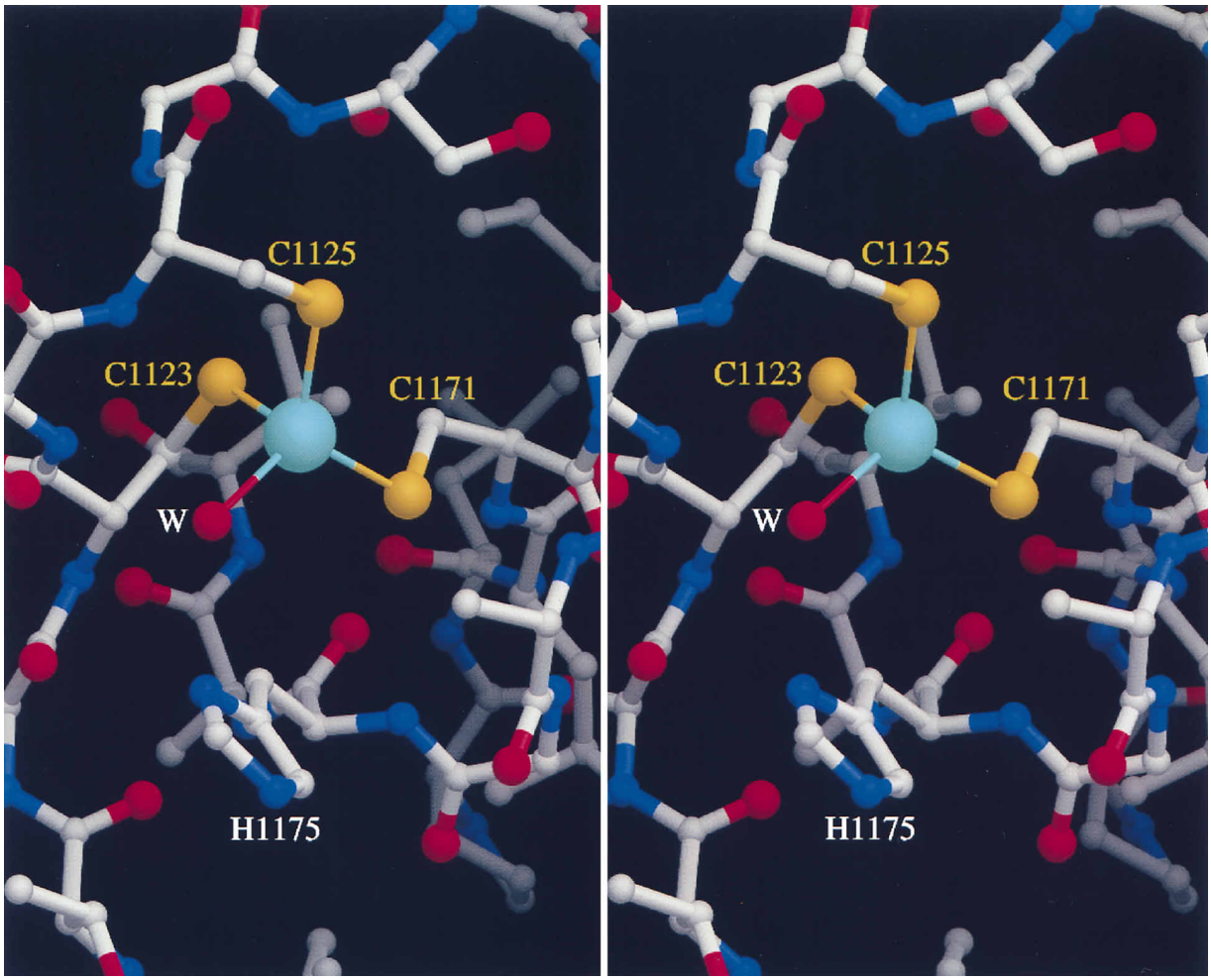

Figure 7. Stereoview of the Zinc Metal Binding Site

The Zn^{2+} coordination displays near perfect tetrahedral geometry. Coordination to the sulphur and oxygen ligands is depicted with narrow lines. Zn^{2+}-sulphur distances to Cys-1123, Cys-1125, and Cys-1171 are 2.29 Å, 2.34 Å, and 2.29 Å, respectively, and the Zn^{2+}-OH_2 distance is 2.37 Å. Atoms are color coded by element type: carbon in gray, nitrogen in blue, oxygen in red, sulphur in yellow, and zinc in cyan. This figure was generated using Ribbons (Carson, 1991).

tightly bound and may require partial denaturation of the protein for its removal. Similar observations have also been reported for the rhinovirus 2A protease (Sommergruber et al., 1994). Because it is located more than 20 Å away from the catalytic Ser-1165, the bound zinc ion is presumed to perform a structural rather than a catalytic role. One possible role of the zinc ion may be to induce stability at the active site through β strand D_2, which separates the active site Ser-1165 and one of the zinc ligands, Cys-1171. The zinc-binding site is also located near the interface of the two domains in the protease and may have an effect on the relative conformations of these domains, which both contribute residues forming the active site and substrate-binding groove.

The NS3 Active Site

The catalytic machinery of the HCV NS3 serine protease appears to be quite similar to that of other serine proteases with chymotrypsin-like folds. The location of the catalytic triad is highly conserved, as are the positions of the backbone amides of Gly-1163 and Ser-1165, which form the oxyanion hole. Further, the twisted strand encompassing residues 1181–1184 superimposes well with the corresponding residues in chymotrypsin (residues 214–217) and other enzymes of this family. This strand has been shown to make hydrogen bonds with the P3 carbonyl and the P1 and P3 amides of peptidomimetic serine protease inhibitors such as aldehydes and trifluoromethylketones (Edwards and Bernstein, 1994). From the standpoint of inhibitor design, however, there are several important differences between the HCV NS3 serine protease and other members of the chymotrypsin family. Comparisons with chymotrypsin and elastase indicate that the loop between strand E_1 and the catalytic Asp-1107, which forms part of the peptide substrate-binding channel, is shorter by six residues in NS3 protease. Wherever a shortened loop has been observed in this region in other cellular serine proteases, such as α lytic protease, it appears to be compensated for by an

altered conformation of the loop between strands B_2 and C_2 such that a well-defined substrate-binding channel is maintained (Fujinaga et al., 1985). In HCV tNS3, however, the B_2–C_2 loop is 14–15 residues shorter than in chymotrypsin and α lytic protease, resulting in a relatively solvent-exposed substrate-binding channel. Finally, in trypsin and thrombin there is a third loop spanning residues 60A–60G that is completely absent in HCV tNS3. Together, these loops have been shown to make crucial interactions with the P2, P3, and P4 moieties of inhibitors of all other members of this enzyme family including elastase, thrombin, trypsin, and factor Xa. The absence of these longer loops in HCV tNS3, and the contact points they provide, is likely to make the design of low molecular weight inhibitors quite challenging.

P1 Specificity Pocket

Cysteine is found at the P1 position in three of the four known cleavage sites for HCV NS3 protease. The fourth known P1 residue is threonine, which occurs at the NS3/NS4A site. This site is unique in that *cis* cleavage appears to be preferred (Bartenschlager et al, 1993; Tomei et al, 1993). Phe-1180 is located at the bottom of the substrate P1 pocket (Figure 1). It is clearly in position to make favorable van der Waals interactions with the P1 residue, and it is known that Cys–Phe interactions are favorable (Burley and Petsko, 1988). Failla et al. (1996) have demonstrated that the double mutant Phe1180Thr and Ala1183Gly increased the size of the S1 pocket, enabling the cleavage of substrate with a bulkier Phe residue at P1.

Based on homology modeling studies, the residues that form the P1 specificity pockets of flavivirus NS3 proteases are quite different to those of the HCV NS3 protease (P. R. C., unpublished data). Specifically, a conserved aspartate is found six residues before the catalytic serine and would be expected to be located at the bottom of the P1 pocket, exactly as in thrombin. This suggests a preference for arginine or lysine, consistent with the observed in vivo cleavage specificities of flavivirus proteases (Chambers et al., 1989).

tNS3–NS4A Interaction Site: Alternate Design Target

As discussed above, and as shown in Figure 6, the contact surface between tNS3 and NS4A is quite extensive and provides another possible site for the design of anti-HCV chemotherapeutic agents. Among 80 deposited NS4A sequences in GenBank, there are only two NS4A residues in the tNS3-binding region that are 100% conserved, Ile-25 and Gly-27. In addition, several other NS4A residues that make significant contact with tNS3 (Val-23, Ile-29, and Leu-31) are always hydrophobic. Finally, the principal tNS3 residues that make contact with the conserved residue Ile-25 of NS4A (Pro-1114, Ile-1090, Val-1055, and Val-1062) are also highly conserved among 28 NS3 sequences in GenBank. These data suggest that it may be possible to design relatively small, predominantly hydrophobic molecules to compete with NS4A binding and activation.

Conclusions

We have described the first atomic-resolution crystal structure of the HCV NS3 serine protease domain, complexed with an NS4A activation peptide. The structure reveals a number of important features. First, the enzyme has a chymotrypsin-like fold with structural similarity to the protease encoded by the Sindbis capsid protein. Second, the activation region of NS4A forms a β strand that intercalates into the N-terminal domain β sheet in the core of the enzyme. The extensiveness of the interactions between NS3 and NS4A suggests it should be considered as an alternative site for the design of anti-HCV drugs. Third, a zinc-binding motif is observed, which appears to have a structural role, given its distance from the enzyme active site. Fourth, several loops found in other chymotrypsin family proteases are missing from HCV. These loops normally play a critical role in defining the shapes of the non-prime-side substrate-binding pockets. In the HCV tNS3 protease, the absence of these loops renders the substrate-binding groove relatively featureless, and this constitutes a challenge for drug design efforts. It is therefore anticipated that the generation of high-resolution structural information for enzyme-inhibitor complexes may be crucial for the optimization of potent, drug-like NS3 protease inhibitors. Finally, the information derived from the HCV tNS3 protease structure may help in further understanding homologous viral proteases encoded by related pathogenic flaviviruses such as the hepatitis G, dengue, and yellow fever viruses.

In vitro HCV replication systems, which would allow unambiguous confirmation of this protease as an essential enzyme in viral replication, are not yet available. However, it is known that mutations that inactivate homologous NS3 protease in yellow fever virus result in the loss of infectivity (Chambers et al., 1990). Mutations that eliminate NS3-dependent cleavage sites in yellow fever virus also result in noninfectious clones (Nestorowicz et al., 1994; Chambers et al., 1995). This, combined with the observations that the polyproteins of all the flaviviruses appear to be cleaved at approximately the same location by their respective NS3 proteases, suggests proteolytic processing at multiple sites is important for HCV replication.

The structure we describe clarifies a multitude of observations derived from biochemical studies, modeling predictions, mutational analyses, and inhibitor studies; all suggesting that the NS3 protease had a simple chymotrypsin-like fold, yet displayed complex NS4A-dependent activation behavior. The intimate interactions between NS4A and NS3 clearly define NS4A as an integral part of the active enzyme structure and help to explain its role in stabilizing NS3. This structure also allows investigations in the hepatitis C field to be redirected toward more precise goals, focusing on the exact mechanism of activation by NS4A, the role of the structural zinc ion, the spatial relationships between the NS3 protease and helicase domains, and the mechanisms by which NS3 might interact with other viral replicative components. It is anticipated that the availability of this information will stimulate efforts to develop more effective anti-HCV therapies, whether they target the NS3 protease or not, and may also provide benefit to related medical problems.

Experimental Procedures

Expression and Purification of tNS3

The truncated NS3 serine protease domain (tNS3) was cloned from a cDNA of the hepatitis C virus H strain (Grakoui et al., 1993b). The first 181 amino acids of NS3 (residues 1027–1207 of the viral polyprotein) have been shown to contain the serine protease domain of NS3 that *trans*-processes all four downstream sites of the HCV polyprotein (Lin et al., 1994b), so we expressed a (His)6-fusion protein based on this tNS3. The plasmid pET-BS(+)/HCV/T7-NS3181-His was derived from pTM3/HCV/1027–1207 (NS3181) (Lin et al., 1994b), by using polymerase chain reaction to introduce epitope tags and new restriction sites. A T7-tag (ASMTGGQQMG), from the N-terminus of the gene 10 protein of the T7 bacteriophage (Tsai et al., 1992), was placed at the N-terminus of the tNS3 domain. Two linker residues (GS) were placed at the tNS3 C-terminus, followed by the (His)6-tag. E. coli JM109(DE3) cells, freshly transformed with the pET-BS(+)/HCV/T7-NS3181-His plasmid, were grown at 37°C in complex media supplement with 100 μg/ml ampicillin, in a 10 L fermentor (Braun). When the cell density reached an OD_{600} of 3–4 the temperature of the culture was rapidly reduced to 30°C, and induction was immediately initiated by the addition of 1 mM IPTG. Cells were harvested at 2 hr postinduction and flash frozen at −70°C prior to purification.

The tNS3 was purified from the soluble fraction of the recombinant E. coli lysates as follows, with all procedures being performed at 4°C unless stated otherwise. Cell paste (75–100 g) was resuspended in 15 volumes of 50 mM HEPES, 0.3 M NaCl, 10% glycerol, 0.1% β-octyl glucoside, 2 mM β-mercaptoethanol (pH 8.0). Cells were ruptured using a microfluidizer and the homogenate was clarified by centrifugation at 100,000 × g for 30 min. The supernatant was brought to 50 mM HEPES, 20 mM imidazole, 0.3 M NaCl, 27.5% glycerol, 0.1% β-octyl-glucoside, 2 mM β-mercaptoethanol (pH 8.0) and applied at 1.0 ml/min to a 7.0 ml Ni-Agarose affinity column, equilibrated in the same buffer. After loading, the column was washed with 10–15 volumes of equilibration buffer and the bound proteins were eluted with equilibration buffer containing 0.35 M imidazole. The protein was then size-fractionated on two columns in series (each 2.6 cm × 90 cm) packed with Pharmacia high resolution S100 resin and equilibrated with 25 mM HEPES, 0.3M NaCl, 10% glycerol, 0.1% β-octylglucoside, 2 mM β-mercaptoethanol (pH 8.0). The tNS3 fractions, identified by SDS–PAGE, were pooled and concentrated to 1 mg/ml using a Amicon Centriprep-10, and stored at −70°C. The tNS3 was thawed slowly on ice and the NS4A peptide (dissolved in the size-exclusion chromatography buffer) was added at a tNS3:NS4A-peptide molar ratio of 1:2. The sample was then diluted 2.5-fold with 15 mM MES, 0.5 M NaCl, 20 mM β-mercaptoethanol (pH6.5) and concentrated to ~2 ml (~2 mg/ml) by ultrafiltration. The sample was then diluted 2-fold with the pH 6.5 buffer and concentrated again to ~2 ml. This dilution process was repeated until it gave a >40-fold dilution of the original buffer constituents. The protein sample was then concentrated to 13.0 mg/ml and centrifuged at ~300,000 × g for 20 min at 4°C. Concentrations of the pure tNS3 and tNS3/4A complex were determined by UV absorption spectroscopy, using a molar absorption coefficient (A_{280}) of 17,700 $M^{-1} \cdot cm^{-1}$.

4A Peptide Synthesis and Purification

The HCV NS4A peptide was synthesized to span residues Gly-21 to Pro-39 of the viral cofactor (residues 1678–1696 of the HCV polyprotein), which incorporates the region reported to be essential for NS3 stimulation (Lin et al., 1995). Lysine residues were added to the termini to assist aqueous solubility, and a serine residue was substituted for Cys–22 (residue 1679 of the polyprotein of the HCV H strain). The peptide (H-KKGSVVIVGRIVLSGKPAIIPKK-OH•TFA salt) was prepared by the solid-phase peptide synthesis (Applied Biosystems 433A) beginning with N^{α}-Fmoc, N^{ϵ}-Boc-Lys Wang resin. N^{α}-Fmoc-protected amino acids were added sequentially using HBTU (2-(1H-benzotriazol-1-yl)1,1,3,3-tetramethyluronium hexafluorophosphate) with HOBt (1-hydroxybenzotriazole hydrate) as coupling agents in N-methylpyrrolidinone. Cleavage from the resin and global deprotection were accomplished with 95% trifluoroacetic acid and 5% water at room temperature for 1.5 hr (15 ml/g resin). The peptide was purified by preparative HPLC on a Waters Delta Pak C18, 15 μm, 300 Å column (30 mm × 300 mm) eluting with a linear gradient of acetonitrile (15%–40%) in 0.1% aqueous trifluoroacetic acid over 35 min (flow rate of 22 ml/min). Peptide purity was confirmed by analytical HPLC. The sequence was confirmed by direct N-terminal sequence analysis and matrix-assisted laser desorption mass spectrometry (Kratos MALDI I), which showed the correct $(M + H)^+$ and $(M + Na)^+$ molecular ions.

Crystallization and Data Collection

Crystals of the tNS3/4A complex were grown by hanging-drop vapor diffusion over a reservoir of 0.1 M MES, 1.8 M NaCl, 0.1 M sodium/potassium phosphate, 10 mM β-mercaptoethanol (pH 6.5). The crystals grew over the course of 2–3 weeks, to final dimensions of about 0.1 × 0.1 × 0.25 mm. They belong to space group R32, with unit cell dimensions a = b = 225.0 Å, and c = 75.5 Å, and contain two tNS3/4A complexes per asymmetric unit.

Statistics for data collection, heavy atom refinement, and crystallographic refinement are given in Table 1. All heavy atom soaks were done in hanging-drops over the same reservoir as used for crystallization. Crystals were transferred to a stabilizing solution (50 mM MES, 2.0 M NaCl, 0.1 M sodium/potassium phosphate, 10 mM β-mercaptoethanol, and 20% glycerol [pH 6.2]) and then frozen in a dry nitrogen gas stream at 100 K (Molecular Structure Corp., Houston, TX) for data collection. Data was acquired by oscillation photography on a Rigaku R-AXIS IIC phosphor imaging area detector mounted on a Rigaku RU200 rotating anode generator (MSC), operating at 50 kV and 100 mA. Measured intensities were integrated, scaled, and merged using the HKL software package (Z. Otwinowski and W. Minor).

Phasing, Model Building, and Refinement

Heavy atom positions were located by inspection and confirmed with difference Fourier syntheses. Heavy atom parameters were refined and phases computed to 3.1 Å using the program PHASES (Furey and Swaminathan, 1996). MIR phases were improved and extended to 2.7 Å by cycles of solvent flattening (Wang, 1985) combined with histogram matching (Zhang and Main, 1990) using the CCP4 crystallographic package (Collaborative Computation Project, 1994). The resulting electron density map displayed nearly continuous density for the protein backbone as well as strong side chain density. Approximately 80% of the model could be unambiguously built into this map (QUANTA 4.1, Molecular Simulations), and a single round of simulated annealing refinement in X-PLOR (Brunger, 1993) brought the R-factor to 29% and free R value to 33% (Brunger, 1992). The remainder of the model was built and refined in several steps, by first extending the resolution to 2.5 Å and then adding well-ordered water molecules. A final round of positional and individual temperature factor refinement brought the R-factor to 21.6% (free R value 26.1%) for 26,652 reflections between 6.0 and 2.5Å (F> 1σF). The eventual model consisted of tNS3 residues 1028–1206 and NS4A residues 21–39 in complex A, and tNS3 residues 1055–1206 and NS4A residues 21–36 for complex B, with 2 zinc atoms and 130 water molecules. A Ramachandran plot for the final model contained 91% of the residues in the most favored regions and 0% in disallowed or generously allowed regions. The rms deviations from ideality were 0.007 Å for bond lengths and 1.47° for bond angles.

Acknowledgments

Correspondence should be adressed to J. L. K. or J. A. T. We thank John Fulghum for assitance with fermentation; Matthew Fitzgibbon and Joyce Coll for related protein purification assistance; Mark Fleming for sequencing and microanalysis of protein crystals and samples; and Jim Griffith and Michael Sintchak for useful discussions and assistance with programs. We are also grateful to Dr. Bert Patterson, of the Department of Chemistry at Concordia University, for the timely metal ion analyses.

Finally we offer particular thanks to Dr. Vicki Sato for providing much of the momentum to initiate this work, for her continued enthusiasm and commitment to the HCV project, and for valuable comments on the manuscript.

Received September 6, 1996; revised September 26, 1996.

References

Alter, M.J., and Mast, E.E. (1994). The epidemiology of viral hepatitis in the United States. Gastroenterol. Clin. North Am. *23*, 437–455.

Alter, H.J., Purcell, R.H., Shih, J.W., Melpolder, J.C., Houghton, M., Choo, Q.-L., and Kuo, G. (1989). Detection of antibody to hepatitis C virus in prospectively followed transfusion recipients with acute and chronic non-A, non-B hepatitis. New Engl. J. Med. *321*, 1494–1500.

Appel, W. (1986). Chymotrypsin: molecular and catalytic properties. Clin. Biochem. *19*, 317–322.

Bartenschlager, R., Ahlborn-Laake, L., Mous, J., and 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., Lohmann, V., Wilkinson, T., and Koch, J.O. (1995). Complex formation between the NS3 serine-type proteinase of the hepatitis c virus and NS4A and its importance for polyprotein maturation. J. Virol. *69*, 7519–7528.

Bazan, J.F., and Fletterick, R.J. (1989). Detection of a trypsin-like serine protease in flaviviruses and pestiviruses. Virology *171*, 637–639.

Behrens, S.-E., Tomei, L., and De Francesco, R. (1996). Identification and properties of the RNA-dependent RNA polymerase of hepatitis C virus. EMBO J. *15*, 12–22.

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

Brunger, A.T. (1993). X-PLOR: A System for X-Ray Crystallography and NMR (New Haven, Connecticut: Department of Molecular Biophysics and Biochemistry, Yale University).

Bukh, J., Purcell, R.H., and Miller, R.H. (1993). At least 12 genotypes of hepatitis C virus predicted by sequence analysis of the putative E1 gene of isolates collected worldwide. Proc. Natl. Acad. Sci. USA *90*, 8234–8238.

Burley, S.K., and Petsko, G.A. (1988). Weakly polar interactions in proteins. Adv. Prot. Chem. *39*, 125–189.

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

Chambers, T.J., McCourt, D.W., and Rice, C.M. (1989). Yellow fever virus proteins NS2A, NS2B, and NS4B: identification and partial N-terminal amino acid sequence analysis. Virology *169*, 100–109.

Chambers, T.J., Weir, R.C., Grakoui, A., McCourt, D.W., Bazan, J.F., Fletterick, R.J., and Rice, C.M. (1990). Evidence that the N-terminal domain of nonstructural protein NS3 from yellow fever virus is a serine protease responsible for site-specific cleavages in the viral polyprotein. Proc. Natl. Acad. Sci. USA *87*, 8898–8902.

Chambers, T.J., Nestorowicz, A., Amberg, S.M., and Rice, C.M. (1993). Mutagenesis of the yellow fever virus NS2B protein: effects on proteolytic processing, NS2B-NS3 complex formation, and viral replication. J. Virol. *67*, 6797–6807.

Chambers, T.J., Nestorowicz, A., and Rice, C.M. (1995). Mutagenesis of the yellow fever virus NS2B/3 cleavage site: determinants of cleavage site specificity and effects on polyprotein processing and viral replication. J. Virol. *69*, 1600–1605.

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

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

Collaborative Computation Project (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. *D50*, 760–763.

Ding, J., McGrath, W.J., Sweet, R.M., and Mangel, W.F. (1996). Crystal structure of the human adenovirus proteinase with its 11 amino acid cofactor. EMBO J. *15*, 1778–1783.

Eckart, M.R., Selby, M., Masiarz, F., Lee, C., Berger, K., Crawford, K., Kuo, C., Kuo, G., Houghton, M., and 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.

Edwards, P.D., and Bernstein, P.R. (1994). Synthetic inhibitors of elastase. Med. Res. Rev. *14*, 127–194.

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

Failla, C., Tomei, L., and De Francesco, R. (1995). An amino-terminal domain of the hepatitis C virus NS3 protease is essential for interaction with NS4A. J. Virol. *69*, 1769–1777.

Failla, C.M., Pizzi, E., De Francesco, R., Tramontano, A. (1996). Redesigning the substrate specificity of the hepatitis C virus NS3 protease. Folding Design *1*, 35–42.

Falgout, B., Miller, R.H., and Lai, C.J. (1993). Deletion analysis of dengue virus type 4 nonstructural protein NS2B: identification of a domain required for NS2B-NS3 protease activity. J. Virol. *67*, 2034–2042.

Francki, R.I.B., Fauquet, C.M., Knudson, D.L., and Brown, F. (1991). Classification and nomenclature of viruses. fifth report of the international committee on taxonomy of viruses. Arch. Virol. [Suppl.] *2*, 223.

Fujinaga, M., Delbaere, L.T.J., Brayer, G.D., and James, M.N.G. (1985). Refined structure of alpha-lytic protease at 1.7 Å resolution. Analysis of hydrogen bonding and solvent structure. J. Mol. Biol. *184*, 479–502.

Furey, W., and Swaminathan, S. (1996). PHASES-95: a program package for the processing and analysis of diffraction data from macromolecules. Meth. Enzymol., in press.

Gomez, J.E., Birnbaum, E.R., Royer, G.P., and Darnall, D.W. (1977). The effect of calcium ion on the urea denaturation of immobilized bovine trypsin. Biochim. Biophys. Acta *495*, 177–182.

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

Grakoui, A., McCourt, D.W., Wychowski, C., Feinstone, S.M., and Rice, C.M. (1993b). A second hepatitis C virus-encoded proteinase. Proc. Natl. Acad. Sci. USA *90*, 10583–10587.

Grakoui, A., Wychowski, C., Lin, C., Feinstone, S.M., and Rice, C.M. (1993c). Expression and identification of hepatitis C virus polyprotein cleavage products. J. Virol. *67*, 1385–1395.

Hahm, B., Han, D.S., Back, S.H., Song, O.-K., Cho, M.-J., Kim, C.-J., Shimotohno, K., and Jang, S.K. (1995). NS3-4A of hepatitis C virus is a chymotrypsin-like protease. J. Virology *69*, 2534–2539.

Hijikata, M., Kato, N., Ootsuyama, Y., Nakagawa, M., and Shimotohno, K. (1991). Gene mapping of the putative structural region of the hepatitis C virus genome by in vitro processing analysis. Proc. Natl. Acad. Sci. USA *88*, 5547–5551.

Hijikata, M., Mizushima, H., Akagi, T., Mori, S., Kakiuchi, N., Kato, N., Tanaka, T., Kimura, K., and 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., and Shimotohno, K. (1993b). Proteolytic processing and membrane association of putative nonstructural proteins of hepatitis C virus. Proc. Natl. Acad. Sci. USA *90*, 10773–10777.

Iwarson, S. (1994). The natural course of chronic hepatitis. FEMS Microbiol. Rev. *14*, 201–204.

Janssen, H.L.A., Brouwer, J.T., van der Mast, R.C., and Schalm, S.W. (1994). Suicide associated with alfa-interferon therapy for chronic viral hepatitis. J. Hepatol. *21*, 241–243.

Jin, L., and Peterson, D.L. (1995). Expression, isolation, and characterization of the hepatitis C virus ATPase/RNA helicase. Arch. Biochem. Biophys. *323*, 47–53.

Kanai, A., Tanabe, K., and Kohara, M. (1995). Poly(U) binding activity

of hepatitis C virus NS3 protein, a putative RNA helicase. FEBS Lett. *376*, 221–224.

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

Kew, M.C. (1994). Hepatitis C and hepatocellular carcinoma. FEMS Microbiol. Rev. *14*, 211–220.

Kim, D.W., Gwack, Y., Han, J.H., and Choe, J. (1995). C-terminal domain of the hepatitis C virus NS3 protein contains an RNA helicase activity. Biochem. Biophys. Res. Commun. *215*, 160–166.

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

Kuo, G., Choo, Q.-L., Alter, H.J., Gitnick, G.L., Redeker, A.G., Purcell, R.H., Miyamura, T., Dienstag, J.L., Alter, M.J., Stevens, C.E., et al. (1989). An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. Science *244*, 362–364.

Li, Y., Hu, Z., Jordan, F., and Inouye, M. (1995). Functional analysis of the propeptide of subtilisin E as an intermolecular chaperone for protein folding. Refolding and inhibitory abilities of propeptide mutants. J. Biol. Chem. *270*, 25127–25132.

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

Lin, C., Lindenbach, B.D., Prágai, B.M., McCourt, D.W., and Rice, C.M. (1994a). 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.

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

Lin, C., Thomson, J.A., and Rice, C.M. (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. *69*, 4373–4380.

Manabe, S., Fuke, I., Tanishita, O., Kaji, C., Gomi, Y., Yoshida, S., Mori, C., Takamizawa, A., Yosida, I., and Okayama, H. (1994). Production of nonstructural proteins of hepatitis C virus requires a putative viral protease encoded by NS3. Virology *198*, 636–644.

Mangel, W.F., Toledo, D.L., Brown, M.T., Martin, J.H., and McGrath, W.J. (1996). Characterization of three components of human adenovirus proteinase activity in vitro. J. Biol. Chem. *271*, 536–543.

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

Mizushima, H., Hijikata, M., Tanji, Y., Kimura, K., and Shimotohno, K. (1994). Analysis of N-terminal processing of hepatitis C virus nonstructural protein 2. J. Virol. *68*, 2731–2734.

Nestorowicz, A., Chambers, T.J., and Rice, C.M. (1994). Mutagenesis of the yellow fever virus NS2A/2B cleavage site: effects on proteolytic processing, viral replication, and evidence for alternative processing of the NS2A protein. Virology *199*, 114–123.

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

Purcell, R.H. (1994). Hepatitis C virus: historical perspective and current concepts. FEMS Microbiol. Rev. *14*, 181–192.

Renault, P.F., and Hoofnagle, J.H. (1989). Side effects of alpha interferon. Semin. Liver Dis. *9*, 273–277.

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

Saito, I., Miyamura, T., Ohbayashi, A., Harada, H., Katayama, T., Kikuchi, S., Watanabe, T.Y., Koi, S., Onji, M., Ohta, Y., et al. (1990). Hepatitis C virus infection is associated with the development of hepatocellular carcinoma. Proc. Natl. Acad. Sci. USA *87*, 6547–6549.

Satoh, S., Tanji, Y., Hijikata, M., Kimura, K., and Shimotohno, K. (1995). The N-terminal region of hepatitis C virus nonstructural protein 3 (NS3) is essential for stable complex formation with NS4A. J. Virol. *69*, 4255–4260.

Shimizu, Y., Yamaji, K., Masuho, Y., Yokota, T., Inoue, H., Sudo, K., Satoh, S., and 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.

Simmonds, P. (1994). Variability of hepititis C virus genome. In Current Studies in Hematology and Blood Transfusion, H.W. Reesink, ed. (Basel: Karger), pp. 12–35.

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

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

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

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

Tanji, Y., Kaneko, T., Satoh, S., and Shimotohno, K. (1995b). Phosphorylation of hepatitis C virus-encoded nonstructural protein NS5A. J. Virol. *69*, 3980–3986.

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

Tomei, L., Failla, C., Vitale, R.L., Bianchi, E., and 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., and 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.

Tsai, D.E., Kenan, D.J., and Keene, J.D. (1992). *In vitro* selection of an RNA epitope immunologically cross-reactive with a peptide. Proc. Natl. Acad. Sci. USA *89*, 8864–8868.

Van der Poel, C.L. (1994). Hepatitis C virus. Epidemiology, transmission and prevention. In Hepatitis C virus. Current Studies in Hematology and Blood Transfusion, H.W. Reesink, ed. (Basel: Karger), pp. 137–163.

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

Webster, A., Hay, R.T., and Kemp, G. (1993). The adenovirus protease is activated by a virus-coded disulphide-linked peptide. Cell *72*, 97–104.

Weiland, O. (1994). Interferon therapy in chronic hepatitis C virus infection. FEMS Microbiol. Rev. *14*, 279–288.

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

Zhang, K.Y.J., and Main, P. (1990). The use of Sayre's equation with solvent flattening and histogram matching for phase extension and refinement of protein structures. Acta Crystallogr. *A46*, 377–381.