Article No. jmbi.1999.2745 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (1999) 289, 371–384

JMB

The Solution Structure of the N-terminal Proteinase Domain of the Hepatitis C Virus (HCV) NS3 Protein Provides New Insights into its Activation and Catalytic Mechanism

Gaetano Barbato, Daniel O. Cicero, M. Chiara Nardi, Christian Steinkühler, Riccardo Cortese, Raffaele De Francesco and Renzo Bazzo*

IRBM ''P. Angeletti''
Via Pontina km 30.600
00040 Pomezia, Roma, Italy

The solution structure of the hepatitis C virus (BK strain) NS3 protein N-terminal domain (186 residues) has been solved by NMR spectroscopy. The protein is a serine protease with a chymotrypsin-type fold, and is involved in the maturation of the viral polyprotein. Despite the knowledge that its activity is enhanced by the action of a viral protein cofactor, NS4A, the mechanism of activation is not yet clear. The analysis of the folding in solution and the differences from the crystallographic structures allow the formulation of a model in which, in addition to the NS4A cofactor, the substrate plays an important role in the activation of the catalytic mechanism. A unique structural feature is the presence of a zinc-binding site exposed on the surface, subject to a slow conformational exchange process.

*Corresponding author

Keywords: HCV; NS3; serine proteinase; structure; NMR

Introduction

Hepatitis C virus (HCV) is recognised as the principal etiologic agent of parenterally transmitted non-A, non-B hepatitis (NANB-H: Choo *et al*., 1989; Kuo *et al*., 1989). The virus establishes a chronic infection that persists for decades in at least 85% of the infected individuals and up to 70% develop chronic active hepatitis. Chronic infection ultimately leads to the development of liver cirrhosis and hepatocellular carcinoma. Neither a vaccine against viral infection nor effective therapy for HCV associated chronic hepatitis has been developed to date. With an estimated world-wide population of infected people of more than 150 million, HCV represents one of the most widely spread and challenging viral infections to block.

The HCV virion has a positive strand RNA genome of about 9.6 kb that encodes a polyprotein of about 3000 amino acid residues (Houghton, 1996). The genetic organisation of HCV is similar to that of Flavi and Pestiviruses and it was classified as a separate genus of the Flaviviridae family. The analysis of the sequence of HCV reveals that it exists in at least six major genotypes and 11 subtypes (Simmonds, 1994). However, all known HCV polyprotein sequences share at least 71% identity. The structural protein ''core'' and the envelope glycoproteins E1 and E2 are released from the N-terminal portion of the polyprotein by action of cellular peptidases, while the non-structural proteins involved in the replication of HCV are released by the action of two virus-encoded proteinases: NS2-3 and NS3 (for reviews, see Bartenschlager, 1997; Neddermann *et al*., 1997). NS2-3 is a zinc-dependent proteinase that performs a single proteolytic cut to release the N terminus of NS3. The proteolytic cleavage at the NS3/NS4A, NS4A/NS4B, NS4B/NS5A, NS5A/NS5B junctions is mediated by a serine proteinase contained within the N-terminal 180 amino acid residues of NS3. The C-terminal (residues 180-630) of NS3 has been demonstrated to possess helicase activity. It has also been shown that the action of a cofactor, NS4A, enhances the activity of the serine proteinase in all the cleavages, *via* the formation of an

Abbreviations used: HCV, hepatitis C virus; NOE, nuclear Overhauser enhancement; NOESY, NOE spectroscopy; SCR, structurally conserved regions.

E-mail address of the corresponding author: bazzo@irbm.it

0022-2836/99/220371–14 $30.00/0

NS3/NS4A complex. Interaction with the NS4A cofactor is required to perform the cleavages at NS3/NS4A, NS4A/NS4B and NS4B/NS5A junctions but the proteinase in its uncomplexed state is still able to cleave at the NS5A/NS5B boundaries, although with a much lower activity.

Since the NS3 proteinase is involved in the maturation process of the virus, the study of the structure of this enzyme is of crucial importance from a pharmacological point of view, in that it can give a strong impulse to the design of inhibitors that may prevent its action and thus block the viral replication and spread.

The crystallographic structures of the free enzyme (Love *et al*., 1996; for simplicity in the following we refer to it as ns3) and its complex with a peptide representing the central region of the NS4A protein cofactor complex (Kim *et al*., 1996; Yan *et al*., 1998; we refer to it as ns3-4a) have been solved. The overall topology is similar in both structures, and forms an N-terminal (approximately residues 1-93) and a C-terminal (residues 94-180) six-stranded anti-parallel β-barrel. The barrels are packed like those of chymotrypsin-like serine proteinases. The catalytic site is formed by the triad of residues H57, D81 and S139, and is found in the crevice between the two domains. In addition to the β-barrels, there are two helical segments: α1 (residues 56-60), comprising the catalytic histidine residue, and α2 (residues 131-137) present also in the ns3 and ns3-4a structures. Two additional helices, α0 (residues 13-21) and α3 (residues 172-180), are formed only in the ns3-4a structure (Kim *et al*., 1996; Yan *et al*., 1998).

The commonly accepted mechanistic model of action of the serine proteinases implies a relay mechanism of hydrogen bonds involving, on one side, the carboxylate moiety of the Asp and the δ HN of the His residues and, on the other side, the ε N of the His and the γ HO of the Ser residues. This relay of H-bonds activates the γ O of the Ser residue, which can produce the nucleophilic attack on the C atom of the scissile bond (Fersht, 1984; Polgar, 1989; Lesk & Fordham, 1996).

In the ns3 structure, the carboxyl group of D81 is positioned far from and points away from H57 (Love *et al*., 1996) impairing the H-bond formation. Conversely, in the ns3-4a structure, the side-chain of D81 is within hydrogen-bonding distance from the H57 imidazole group (Kim *et al*., 1996; Yan *et al*., 1998). By comparison of the ns3 and ns3-4a structures it has been inferred that D81 is correctly positioned as a member of the canonical catalytic triad as a consequence of the presence of the NS4A cofactor (Love *et al*., 1998).

This conclusion, reasonable as it may seem from the crystallographic data alone, is challenged by new available biochemical evidence. In fact, the pH-dependence of the hydrolysis reaction studied in the presence of substrate and with or without NS4A titrates with the identical p*K* value of 7.2 (Landro *et al*., 1997). The authors conclude that the activation role of NS4A is not exerted by a perturbation of the pK_a values of the active-site residues involved in the catalysis, in contrast with the model proposed by Love *et al*. (1998). On the other hand, NMR pH titration of the catalytic H57 residue of the free enzyme gives a pK_a value of 6.8 (Urbani *et al*., 1998). Assuming that the pK measured by Landro *et al*. (1997) reflects the actual value for H57, the difference in pK_a value from that of the free enzyme could be due either to a direct role of the substrate itself in the pK_a alteration or to the different buffers used in the experiments. A kinetic analysis conducted on different types of substrate-like inhibitors in the absence and presence of the NS4A cofactor has shown that the action of NS4A peptide is exerted only on the P′-side of the substrate (Landro *et al*., 1997). From these findings, the authors conclude that NS4A modulates NS3 activity by alteration of the S′ subsites.

At the moment, despite knowledge of the crystallographic structures in the absence and presence of the NS4A cofactor, its mechanism of activation and its role on the catalytic triad relative orientation is not completely understood. Here, we illustrate the first solution structure of the NS3 proteinase domain (first N-terminal 180 residues of NS3 with the addition of a solubilising six residue tail with the amino acid sequence ASKKKK) obtained in the absence of the NS4A cofactor and based on NMR data. The novelty of our findings is that the global architecture responsible for the relative positioning of the catalytic residues is already present in the absence of the NS4A cofactor. This difference from the relative crystallographic structure ns3 accounts for all the biochemical evidence available to date. The action of the cofactor is then discussed in terms of stabilisation of the fold of the N-terminal region and by its influence on the substrate leaving-group S′ side, while an influence on the substrate recognition S side can be excluded. Also, a possible role of the substrate in the relative positioning of the catalytic triad is proposed and discussed.

A rather unusual structural feature of the NS3 enzyme is the presence of a zinc-binding site completely exposed to the solvent. We find that this site in solution undergoes a conformational exchange between an open and a closed conformation by switching the side-chain of H149 on the hundreds of milliseconds timescale.

Results and Discussion

Structure determination

The solution structure of the protease domain of the hepatitis C virus NS3 protein of the strain BK was solved by multidimensional heteronuclear NMR spectroscopy (Clore & Gronenborn, 1991a; Bax & Grzesiek, 1993) making use of uniformly labelled ^{15}N, $^{15}N/^{13}C$ and $^{2}H/^{15}N/^{13}C$ samples, as well as of a selectively ^{15}N[Leu], ^{15}N[Val] and ^{15}N[Ala]-labelled sample. Complete resonance assignment was obtained, except for 15 residues at the N terminus and few other signals. In fact, the

proton resonances of residues 6-21 appear to be broadened beyond detection by conformational exchange. Leucine residues 14, 15 and 21 resonances, for example, were identified as very broad peaks in a sample selectively labeled with leucine but it was not possible to identify them sequentially. The general quality of NMR data is shown in Figure 1, which depicts strips from (a) a 3D ^{13}C edited NOESY and (b) from a ^{15}N edited NOESY. The structure was calculated, excluding the first 21 residues, by simulated annealing (Nilges *et al.*, 1988) using the database of constraints shown in Table 1, where a summary of the structural statistics is given. In Figure 2, a stereoview of the overlay of the 20 lower-energy structures generated is shown. For simplicity we will refer in the following to the minimised average solution structure as nmr.

Description of the structure

The nomenclature used for trypsin β-strands and applied to the topology of the protein is shown in Figure 3(a). The sequence alignment of NS3 with several other chymotrypsin-like serine proteinases is proposed in Figure 3(b), on the basis of the structural overlay, following the guidelines indicated by Greer (1990). This alignment differs slightly from that proposed previously (Love *et al.*, 1996). The

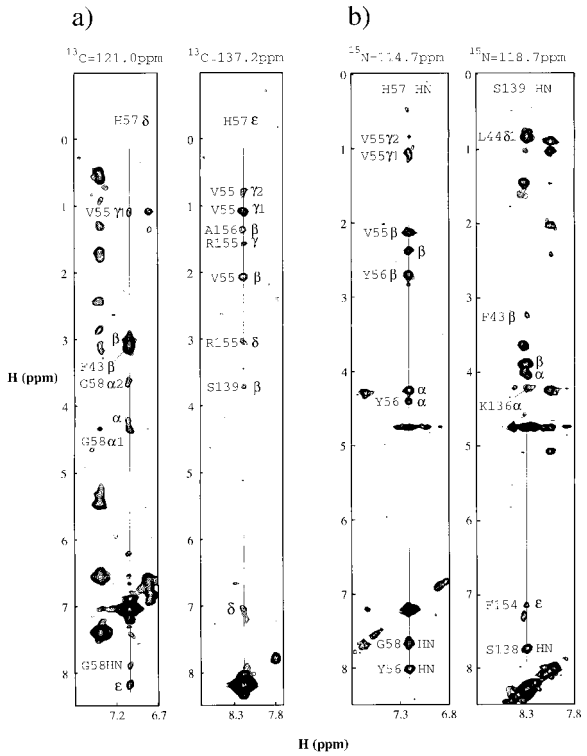

Figure 1. Examples of spectra showing the quality achieved with this protein. (a) Strips taken from a ^{13}C edited 3D NOESY show the aromatic H^{ε} and H^{δ} NOEs of the catalytic histidine residue. (b) Strips of the amide protons of both the catalytic histidine and serine residues from a ^{15}N edited 3D NOESY.

Table 1. Experimental restraints and structural statistics

A. *NMR constraints*		
NOE	Total	2476
	Intra	945
	Inter short distance (<*i*+3)	521
	Inter long-range (>*i*+3)	1010
Generic	Total	70
	H-bond	64
	Zn-binding site	6
Dihedral	Total	83
	φ	37
	χ1	43
Stereospecific	Methylene groups	31/219
	Methyl groups	50/66
B. *Structure statistics*		
R.m.s. deviations from experimental constraints[a]		
Distance (Å)		0.076 ± 0.003
Dihedral(deg.)		1.331 ± 0.148
$^{13}C^{\alpha}$		1.49 ± 0.05
$^{13}C^{\beta}$		1.09 ± 0.06
Deviations from idealized geometry		
Bonds (Å)		0.005 ± 0.0006
Angles (deg.)		0.761 ± 0.032
Impropers (deg.)		0.543 ± 0.01
Coordinates precision referred to mean structure (Å)		
Residues SCR+helices		
Backbone		0.472 ± 0.089
All heavy atoms		1.147 ± 0.160
All residues		
Backbone		0.872 ± 0.097
All heavy atoms		1.306 ± 0.117
N-terminal residues SCR+helices		
Backbone		0.554 ± 0.140
All heavy atoms		1.020 ± 0.292
C-terminal residues SCR+helices		
Backbone		0.233 ± 0.042
All heavy atoms		0.567 ± 0.091
C. *Ramachandran analysis*[b]		
% Residues in most favoured regions		70.4 ± 1.8
% Residues in allowed regions		26.7 ± 1.4
% Residues in generously allowed regions		3.0 ± 0.6
% Residues in disallowed regions		0.0 ± 0.0

[a] None of the structures exhibited distance violations greater than 0.5 Å or dihedral angle violations greater than 5°.

[b] The program PROCHECK (Laskowski *et al.*, 1993) was used to assess the overall quality of the structures. The residues with a heteronuclear NOE ^{15}N-$^{1}H < 0.6$ (total 28 residues) were excluded from the computations because of their intrinsic mobility.

structurally conserved regions (SCR) are indicated with boxes in the alignment. The overall sequence similarity is very low (<20%), but the general topology is well conserved. The NS3 proteinase, like the proteinase from Sindbis virus (Tong *et al.*, 1993) and from the Semliki Forest virus (Choi *et al.*, 1997), is a small proteinase (about 180 residues) and, as such, makes an economical use of loops, lacking all of a series of connecting elongations that are a common feature of cellular proteinases: we will see below that this has a peculiarly relevant consequence.

The N-terminal β-barrel appears to be less compact than the C-terminal one (Figure 2). Evidence for this is obtained from a comparison of the num-

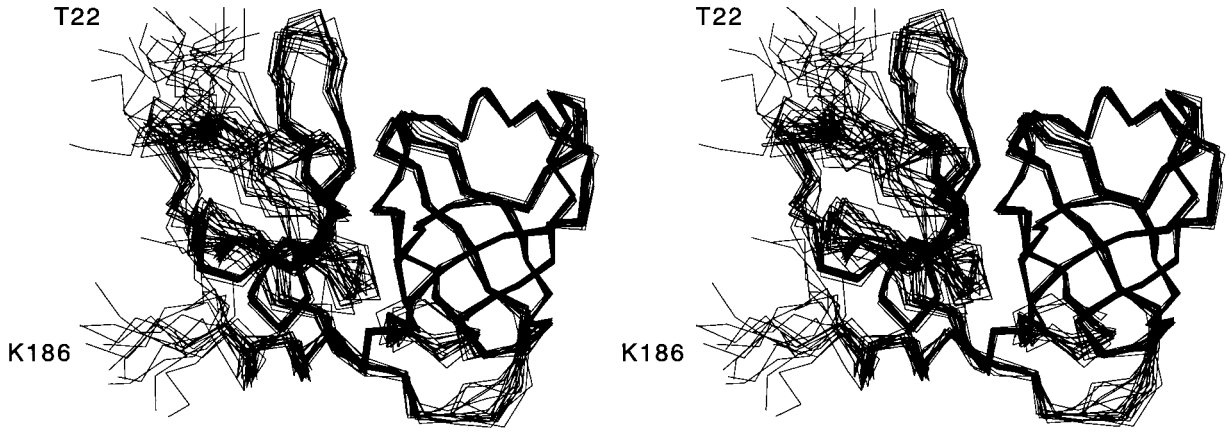

Figure 2. A stereoview of the 20 minimum-energy structures is shown. For the overlay, the SCR residues identifying the β-strands plus the helicoidal segments were used. Due to its peculiar mobility, see the NS4a interaction section, strand D1 was omitted in the calculation of the r.m.s.d. The first 21 residues were not included in the structure calculation, because for them no structural information was available.

ber of slow-exchanging amide protons between the two domains: 12 in the N-terminal and 20 in the C-terminal domain. In Figure 4(a), (b) and (c) the number of observed NOEs per residue, the backbone heavy atoms r.m.s.d. per residue, and the heteronuclear ^{1}H-^{15}N NOEs are reported, respectively. The total number of NOEs for the N-terminal barrel is 990, whereas for the C-terminal it is 1486. The ^{1}H-^{15}N NOE values in 17 residues in the N-terminal portion of the molecule are below 0.6 (indicating a high level of mobility), while this is the case for only six residues in the C-terminal domain (excluding the residues forming the solubilising tail). These differences are reflected in the poorer r.m.s.d. of the N-terminal barrel compared to the C-terminal one (Figure 4(b)).

Table 2 illustrates the pairwise comparison of the SCR residues between the nmr and the ns3 and ns3-4a structures. The r.m.s.d. values for the backbone heavy atoms are 1.45 and 1.18 Å, respectively. However, if we consider the N and C-terminal β-barrels separately, we get a better insight into the comparison. There is a substantial difference between the N-terminal β-barrel, with an r.m.s.d. of 1.68 Å (nmr/ns3), of 1.48 Å (nmr/ns3-4a) and of 1.98 Å (ns3/ns3-4a). The nmr structure looks different to a similar extent from either of the two crystallographic structures, thereby suggesting that, in the crystal of the uncomplexed enzyme, crystal forces are responsible for distorting the N-terminal β-barrel. The C-terminal β-barrel with an r.m.s.d. of 0.52 Å (nmr-ns3), of 0.56 Å (nmr/ns3-4a) and of 0.41 Å (ns3/ns3-4a) is very similar in all three structures. However, a major difference in this domain is given by the fact that in solution (thus in the absence of the NS4A peptide) we find helix α3 (residues 172-182), while this helix is absent from the ns3 (in the absence of NS4A peptide) structure. In Figure 5(a) a zoomed view of the overlay of the ns3 and nmr structures is shown, providing evidence for this difference, which is extremely relevant in the evaluation of the role played by NS4A peptide to activate the proteinase and will be further discussed in the catalytic triad and substrate-binding section.

Table 2 also reports the pairwise comparisons with several other serine proteinases belonging to the chymotrypsin family. Also in this case it is instructive to consider separately the two β-barrels. The fold of the N-terminal SCR residues in solution is similar to that of chymotrypsin (Blevins & Tulinsky, 1985), trypsin (Bode *et al.*, 1984) and elastase (Meyer *et al.*, 1988), with an r.m.s.d. of 1.37, 1.40 and 1.42 Å, respectively, while the ns3 structure gives 1.80, 1.91 and 1.84 Å, respectively, and the ns3-4a structure gives 2.12, 2.12 and 2.11 Å, respectively. One can conclude that, in absence of the NS4A peptide, the overall fold of the β-barrel is conserved; when forming the complex with the NS4A peptide, the N-terminal β-barrel undergoes a substantial change in conformation (specifically, the D1 strand and the loops preceding and following it; Figure 3(a)) that differentiates it locally from the other proteinases belonging to the same family.

The C-terminal β-barrel of all the NS3 structures, although very similar to each other, differs from those of the other chymotrypsin-like proteinases (r.m.s.d. >2.0 Å). This is partially due to the slightly different packing of the E2 strand, which is directly involved in substrate binding, and of the C2 strand, which is packing directly against it (Figure 3(a)). This difference can be accounted for by the peculiarity of the substrate recognition surface for NS3. In fact, substrate specificity studies have shown that the NS3 proteinase requires at least a decamer peptide spanning P6-P4′ for optimal activity. The substrate frame needed thus is unusually long for serine proteinases.

Incidentally, the only other viral serine proteinases, Sindbis virus (Tong *et al.*, 1993) and Semliki Forest virus core protein (Choi *et al.*, 1997), for which a structure is available in the Brookhaven database, are more similar to NS3 with an r.m.s.d. for the C-terminal β-barrel

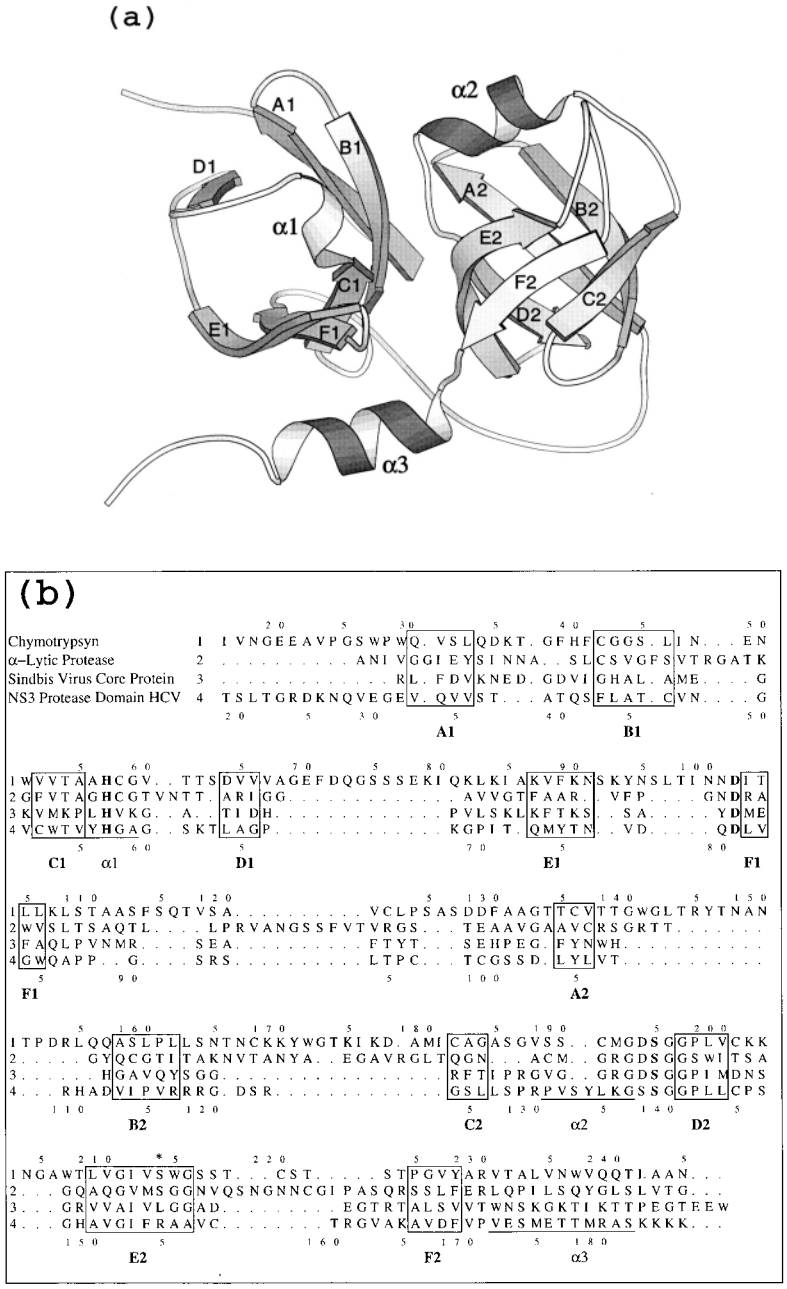

Figure 3. Topology and sequence alignment. (a) The topology of the NS3 proteinase domain is represented in a MOLSCRIPT view. The SC Regions are constituted by the strands A1-F1 in the N-terminal and A2-F2 in the C-terminal β-barrel. The helices are named α1, α2 and α3. (b) The structure alignment of NS3 with chymotrypsin-like serine proteinase. The alignment results from the best superposition of the C^{α} of the SCR residues, in the box. The upper numeration is referred to the chymotrypsin, while the bottom is referred to the NS3 proteinase domain. In bold are the residues of the catalytic triad, while the asterisk (*) at the top indicates the position of residue S214, which is strictly conserved in all cellular proteinases, and its corresponding residue in the two viral proteinases. Below the NS3 amino acid sequence is shown the corresponding secondary structure elements; strands are coincident with the boxed regions, while the helices are underlined.

ranging between 1.6 and 1.72 Å for all the pairwise comparisons.

The positioning of NS4A

The lower definition of the N-terminal β-barrel in the solution structure should be related to the absence of the NS4A protein cofactor. In Figure 6, a zoomed view of the overlay between the nmr (blue) and the ns3-4a (NS3 cyan, NS4A magenta) structures is shown. It has been shown by deletion mutagenesis experiments (Failla *et al*., 1995), and more recently by the ns3-4a crystallographic structure (Kim *et al*., 1996; Yan *et al*., 1998), Figure 6, that only the N-terminal β-barrel of the proteinase is involved in binding the NS4A peptide cofactor. According to these structures, the NS4A peptide cofactor is almost completely buried inside the core of the N-terminal β-barrel, where it forms a β-strand within a four-stranded β-sheet. Its companion strands are formed by residues 4-10 (strand A0) and 33-37 (A1). Residues 13-21 form a short α-helix (α0) that is likely to contribute to the stability of the NS3-NS4A complex in the crystal. This helix is very peculiar, since the residues exposed to the solvent are three leucine and two isoleucine residues (Figure 6). It is unlikely that such a structure could exist in solution. In this respect, it is interesting to note that only one of the monomers in the crystallographic asymmetric unit is folded in the way just described (Kim *et al*., 1996). In the second monomer, the first 30 residues do not have a

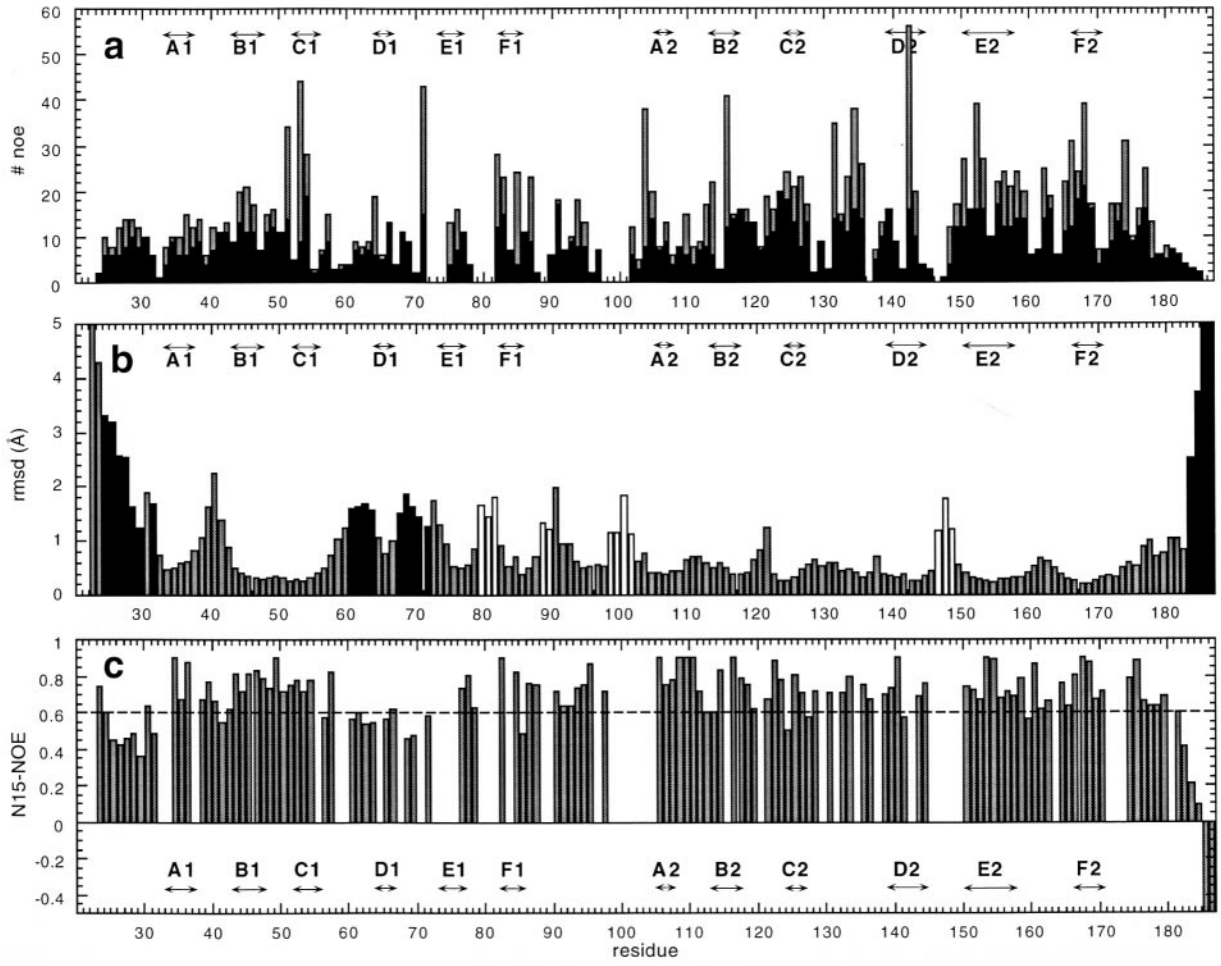

Figure 4. Correlation between the number of NOE constraints, mobility and r.m.s.d. The SCR strands are defined by the arrows and reported in all three panels to simplify reading the Figure. (a) A representation of the number of NOEs per residue, in black are represented the NOEs involving backbone protons, while in grey are represented those involving the side-chains proton. (b) Behaviour of the r.m.s.d. per residue. Regions in which the r.m.s.d. is poor can be related to low values of the ^{1}H-^{15}N NOE (<0.6) and are in black, or to a lack or low number (<10) of NOE constraints, and are in white. Only two regions do not follow this characteristic, namely residues 39-41 and 120-122, both are well-characterized turns but lack long-range NOEs. (c) Heteronuclear ^{1}H-^{15}N NOE value per residue number. The broken line drawn at the value of 0.6 indicates the mobile regions (<0.6).

defined structure. Despite these differences, both complexes contain the NS4A peptide in essentially the same position. It could be argued that more than one conformation can be assumed by the N-terminal 30 amino acid residues of NS3 even in the presence of NS4A. In our experiments, residues 6-21 signals are broadened almost beyond detection, moreover residues 25-31 exhibit a high mobility in solution (Figure 4(c)). Therefore, it seems reasonable to assume that some type of mobility is affecting all the N-terminal 31 residues. Evidence obtained by limited proteolysis experiments suggest indeed that the N-terminal region of NS3 is highly accessible in solution, even in the complex with the NS4A peptide cofactor (data not shown).

The N-terminal residues of the NS4A peptide contact directly the D1 strand and the preceding and following loops α1-D1 and D1-E1, respectively (Figure 6). In solution, these regions together with the strand D1 itself are characterised by a high degree of mobility (^{1}H-^{15}N NOE < 0.6; Figure 4(c)). From the overlay of the nmr and ns3-4a structures shown in Figure 6, it is evident that the whole region encompassing the residues 61-66 (loop α1-D1 and strand D1) are held down toward the F1 strand in the crystallographic structure, while in solution this strand is packing across the A1 strand, compensating for the absence of the NS4A peptide. From this comparison it can be concluded that one of the roles exerted by NS4A is to stabilize the strands A1, D1, the loop α1-D1 and D1-E1 in a more defined conformation thus compacting the whole N-terminal barrel. Its influence on the substrate-binding region is primarily due to the direct interaction and consequent conformational stabilisation of the strand A1 and the region α1-D1, which form the walls surrounding the P′-side of the substrate (see the next section).

Table 2. The r.m.s.d. comparison with several serine proteinases

N-terminal									
	2snv	5cha	2alp	1ntp	3est	2sga	3sgb	ns3	ns34a
nmr	1.91	1.37	1.86	1.40	1.42	1.70	1.82	1.68	1.48
2snv		1.45	2.12	1.60	1.50	2.25	2.03	1.77	2.45
5cha			1.81	0.36	0.25	1.67	1.73	1.80	2.12
2alp				1.78	1.78	1.09	0.37	2.02	2.17
1ntp					0.33	1.64	1.71	1.91	2.12
3est						1.66	1.70	1.84	2.12
2sga							1.07	1.96	1.77
3sgb								1.90	2.11
ns3									1.98
C-terminal									
	2snv	5cha	2alp	1ntp	3est	2sga	3sgb	ns3	ns34a
nmr	1.72	2.20	1.86	2.18	2.25	2.14	2.13	0.52	0.56
2snv		1.51	1.66	1.51	1.52	1.61	1.61	1.60	1.66
5cha			0.75	0.45	0.49	0.72	0.79	2.13	2.18
2alp				0.76	0.80	0.42	0.47	2.06	2.14
1ntp					0.63	0.69	0.73	2.11	2.18
3est						0.78	0.83	2.16	2.22
2sga							0.24	2.02	2.10
3sgb								2.02	2.10
ns3									0.41
all									
	2snv	5cha	2alp	1ntp	3est	2sga	3sgb	ns3	ns34a
nmr	2.09	2.08	2.15	2.03	2.09	2.06	2.11	1.45	1.18
2snv		1.66	2.06	1.74	1.63	2.07	1.99	1.95	2.27
5cha			1.51	0.46	0.52	1.40	1.44	2.04	2.31
2alp				1.44	1.47	0.83	0.47	2.14	2.25
1ntp					0.60	1.33	1.38	2.06	2.26
3est						1.40	1.44	2.08	2.31
2sga							0.77	2.12	2.03
3sgb								2.04	2.21
ns3									1.66

Root-mean-square deviation (Å) comparison of the SRC backbone residues heavy atoms for several serine proteinases. The names are given as Brookhaven PDB codes: 1bt7 (nmr); 1a1q (ns3); 1jxp (ns3-4a); 2snv (Sindbis virus core protein); 5cha (bovine alpha-chymotrypsin); 2alp (alpha lytic protease); 1ntp (bovine beta trypsin); 3est (porcine elastase); 2sga (*Streptomyces griseus* protease A); 3sgb (*Streptomyces griseus* protease B). Due to the high degree of similarity between the Sindbis virus and the Semliki Forest virus core proteins (1vcp) (r.m.s.d. 0.4 Å), the comparison are reported only for the former.

The high level of similarity between the C-terminal barrel of the solution and that of the crystallographic structures, is quantified by the low r.m.s.d. of the backbone heavy atoms for the SCR residues (Table 2). It can be concluded that the complex with the NS4A peptide has little, if any, influence on this region of the enzyme, which contains the recognition pockets for the P-side residues of the substrate.

These conclusions find experimental support from a steady-state kinetic analysis of inhibitor binding to the active site of the NS3 proteinase (Landro *et al*., 1997). In fact, two classes of competitive inhibitors could be identified: those interacting

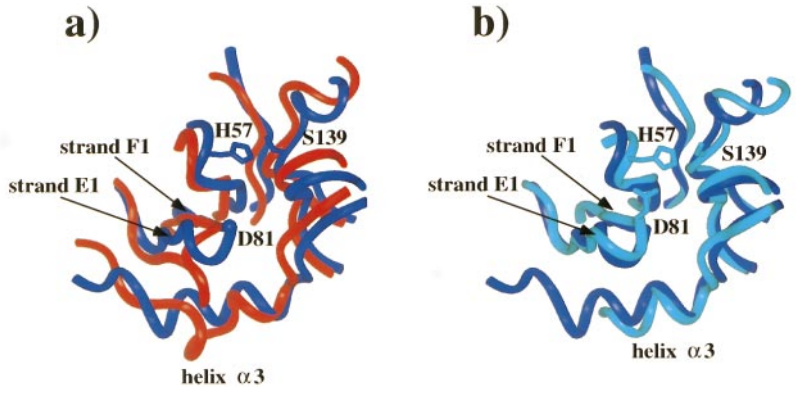

Figure 5. Pairwise superposition of the nmr (blue) with ns3 (red) and ns3-4a (cyan) structures. The superposition is obtained from the r.m.s.d. of all the SCR residues. The catalytic triad residues are shown on only one of the structures to facilitate the comparison. The position of D81 in the nmr structure is indicated by a dot, since its position is not accurately determined. The residues were not shown for ns3, since only the C^{α} coordinates are publicly available. (a) nmr/ns3: the strands E1-F1 containing the loop that bears the catalytic Asp are different. In the ns3 structure, the helix α3 is essentially absent. (b) nmr/ns3-4a: the strands E1-F1 bearing the Asp residue are similarly positioned, and the helix α3, which is packed mainly against the strand E1, is very similar in both structures.

only with the P binding pockets located on the C-terminal β-barrel and those compounds extending their interaction with the enzyme to both the P and P′ binding sites. The potency of the former category of inhibitors was not influenced by complex formation with NS4A. In contrast, the affinity of active-site ligands relying only on contacts with the P′ binding site on the N-terminal β-barrel was strongly impaired in the absence of NS4A.

The catalytic triad and substrate-binding region

The serine proteinases of the chymotrypsin family share a number of elements: a catalytic triad formed by residues Ser-His-Asp; a site for hydrogen bonding to a tetrahedral oxyanionic intermediate of the reaction (also called the oxyanion hole); a strand forming an antiparallel β sheet with the P-side of the polypeptide chain of the substrate (also called S site), which contributes also to the formation of a recognition pocket, and a leaving-group recognition site (also called S′ site). Each of these elements occurs in the different members of the family in an almost identical geometric relationship.

The alignment with other serine proteinases shows that these residues and the nearby positions are well conserved (Figure 3(b)). From the structural point of view, the relative position of H57 and S139 is similar in the nmr and ns3-4a structures and only slightly more apart than in the other serine proteinases (Table 3). On the contrary, in the ns3 structure ''the imidazole of H57 [...] is oriented toward S139 but is not close enough to form the H-bond observed in proteinase structures'' (Love *et al.*, 1996). The difference in distance (Table 3) is very likely due to distortions induced by the crystal forces, since in solution we also find that the observed pK_a value of 6.8 for H57 (Urbani *et al.*, 1998) (uncomplexed enzyme) is in agreement with the catalytic histidine residue being H-bonded with the catalytic serine, as expected from the canonical model of serine proteinases.

In the nmr structure, the position of D81 is not accurately determined because of the lack of experimental constraints. As a matter of fact, for the resonances of residues 79, 80 and 81 no inter-residue

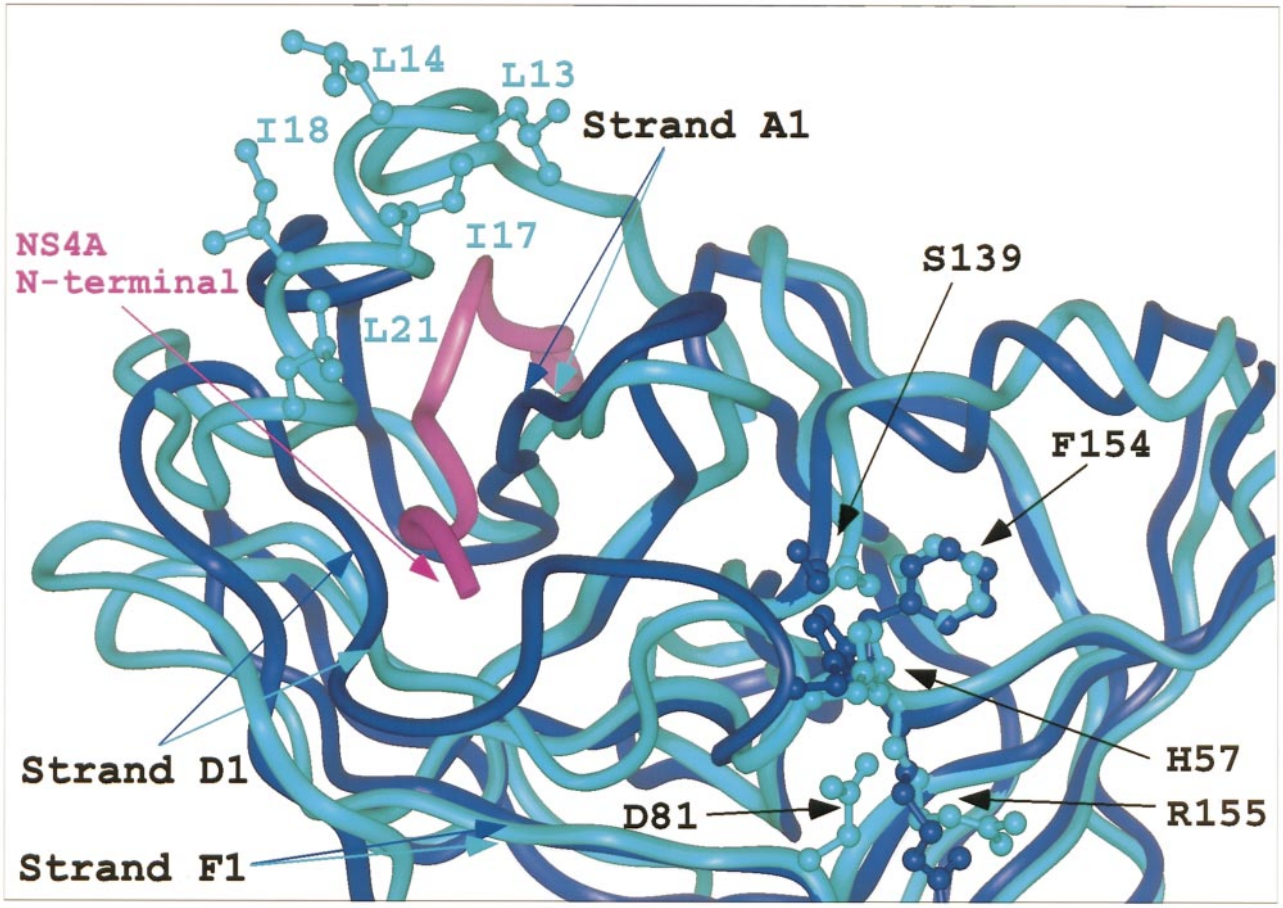

Figure 6. The role of NS4A; zoomed vision of the superposition of nmr/ns3-4a structures (blue/cyan-magenta). The catalytic triad residues as well as R155 side-chains are shown to facilitate the comparison. The role of NS4A seem to be to order the N-terminal 21 residues in an initial strand and a subsequent helix that turns around the N terminus of NS4A. However, this helix exposes hydrophobic residues to the solvent while packing inside the hydrophilic residues. A situation like this is not possible in solution. The positioning of NS4A has the consequence of order-

Table 3. Geometric relation of the catalytic residues in serine proteinases

	His-Ser (Å)	His-Asp (Å)	Asp-Ser (Å)
nmr	9.0	-	-
ns3	9.7	7.9	11.5
ns3-4a	9.3	6.5	10.5
2snv	8.8	7.0	10.6
5cha	8.5	6.5	10.5
1ntp	8.5	6.3	10.1
2alp	8.3	6.2	9.9
3est	8.4	6.5	9.8
2sga	8.5	6.2	9.9
3sgb	8.4	6.3	10.1

Comparison of the distances between C^{α} of the catalytic triad in NS3 structures and in several serine proteinases. The actual distances His-Asp and Asp-Ser for the nmr structures are not included, since the whole loop containing D81 undergoes substantial conformational averaging, so that a single evaluation of distance would not be meaningful.

NOE is observed and the signals arising from the amide protons of these residues are affected by their fast water exchange rates. The resulting calculated structures show that several allowed conformations of this loop are sampled in the simulated annealing computations, all of them solvent-exposed.

It is worthwhile to point out, that also in the ns3-4a crystal structure this loop is solvent-exposed and the temperature factors of the backbone atoms of the three residues 79-81 are around 50.0 $Å^2$, which indicates a high degree of mobility even in presence of the NS4A peptide. Moreover we find that, in the average solution structure of the isolated NS3 proteinase domain, the strands E1 and F1, which enclose the loop bearing D81, are positioned similarly to the crystal ns3-4a structure despite the absence of the cofactor, as clearly shown in the zoomed view of the overlay between the nmr and ns3-4a structures in Figure 5(b). These strands are held in position by the packing of the helix α3 (residues 172-182), which in solution is similar to that in the ns3-4a crystal (Figure 5(b)). On the contrary, in the ns3 structure this helix is absent (Figure 5(a)), strongly suggesting that the crystal packing is inducing a distorted conformation in this region. The consequence is that the strands E1 and F1 that position the D81 are distorted and thus its positioning in the ns3 structure is not compatible with a well-formed catalytic triad. Our results in solution thus are in contrast with the conclusions drawn by Love *et al.* (1998) based on the crystallographic structures alone. On the other hand, our results are in agreement with the biochemical evidence that the presence of NS4A is not affecting the pK_a of the catalytic residues (Landro *et al.*, 1997), indicating that the catalytic triads of the enzyme-substrate complex and of the ternary enzyme-substrate-NS4A complex must possess very similar geometries.

From all the above considerations, we conclude that other factors in addition to the presence of NS4A are to be invoked to stabilize the position of the D81 residue in a canonical catalytic triad configuration. The identification of such factors could be attempted by analysing the structures of several members of the chymotrypsin-like family of serine proteinases. Ideally, we can divide it into two subfamilies; namely, the short-chain, about 180 residues, and the long-chain proteinases, about 250 residues (Bazan & Fletterick, 1998). In all the solved long-chain structures there are two conserved characteristics: (i) the residue at position 214 is invariably serine, which forms an H-bond with the carboxylic group of the catalytic aspartate residue, thus helping in its correct positioning and alignment in respect to the catalytic histidine; (ii) the catalytic aspartate residue is shielded from the solvent by the hydrophobic residues at conserved positions. In Table 4 these characteristics are summarized for several long-chain serine proteinases and the short-chain proteinases for which a structure is available. They do not share these characteristics, and incidentally these proteinases are all of viral origin.

The protection from the solvent allows the direct observation of the δ HN proton NMR signal of the

Table 4. Summary of some serine proteinases that present both Ser in position 214 and the catalytic Asp sheltering from the solvent, and the few known exceptions

Serine proteinase	PDB entry	Category	Origin	214	Sheltering	Reference
Trypsin	1NTP	LC[a]	Mammalian	Ser	Y94, L99	Bode *et al.* (1984)
Chymotrypsin	5CHA	LC	Mammalian	Ser	Y94, I99	Blevins & Tulinsky (1985)
Kallikrein	2PKA	LC	Mammalian	Ser	F94, Y99	Bode *et al.* (1983)
Thrombin	1HXE	LC	Mammalian	Ser	Y94, L99	Bode *et al.* (1989)
Mast cell proeinase	3RP2	LC	Mammalian	Ser	Y94, N99	Remington *et al.* (1988)
Tonin	1TON	LC	Mammalian	Ser	Y94, L99	Fujinaga & James (1987)
Elastase	1HNE	LC	Mammalian	Ser	Y94, L99	Meyer *et al.* (1988)
Streptomyces griseus A	2SGA	LC	Bacterial	Ser	F94, Y171, V177	Moult *et al* (1985)
Streptomyces griseus B	3SGB	LC	Bacterial	Ser	F94, Y171, V177	Read *et al.* (1983)
Alpha lytitc protease	2ALP	LC	Bacterial	Ser	F94, Y171, V177	Fujinaga *et al.* (1985)
Sindbis core protein	2SNV	SC[a]	Viral	Leu		Tong *et al.* (1993)
Semliki Forest core protein	1VCP	SC	Viral	Leu		Choi *et al.* (1997)
NS3(1-180) HCV	1A1R	SC	Viral	Arg		Kim *et al.* (1996)
	1JXP					Yan *et al.* (1998)

[a] LC, long-chain, SC, short-chain Serine proteinases.

histidine residue involved in the H-bond with the carboxylic moiety of the aspartate residue (Frey *et al*., 1994). This proton resonates at an unusually low field chemical shift, in the range of 14.5-19.0 ppm depending on pH, and has been observed in several serine proteinases of the long-chain family (Markley, 1978; Bachovchin, 1985; Frey *et al*., 1994). In the case of the NS3 proteinase, however, despite all attempts, such a signal has until now not been observed. From both the nmr and ns3-4a structures, one could argue that this is due to the site being solvent accessible. We speculate that, in solution, the aspartate residue is unlikely to be engaged in an H-bond with the histidine residue, even in the presence of the NS4A cofactor. However, an environment similar to that of the long-chain subfamily could be partially created with the participation of the substrate. We observe, in fact, for the short-chain enzymes a tendency to have a hydrophobic bulky residue in position P2 (Tong *et al*., 1993; Ingallinella *et al*., 1998), whereas long-chain enzymes tend to have either glycine or short side-chain residues such as Ala (Yasutake & Powers, 1981; Chang, 1986; Coombs *et al*., 1996). A bulky residue in position P2 together with the aliphatic methylene groups of R155 (NS3) and of the L231 side-chain (Sindbis and Semliki viruses) could contribute to shelter the aspartic acid side-chain, thereby favouring the formation of a catalytic triad machinery that more closely resembles that observed in long-chain enzymes.

The pH titration data obtained in the presence of substrate (Landro *et al*., 1997) show that the value of pK_a 7.2 of the catalytic residues of the NS3 proteinase is NS4A-independent, but this value is different from that of 6.8 found for the catalytic histidine residue of the free enzyme (Urbani *et al*., 1998). One could speculate thus that a role in the enhancement of the pK_a value is played by the substrate itself; however, it is possible that the difference observed is due to the different buffer conditions used in the experiments. Preliminary spectroscopic evidence in agreement with the hypothesis of an involvement of the substrate itself in the conformational stabilization of the catalytic triad has been collected on substrate-based inhibitors (Cicero *et al*., 1999).

The residues preceding the catalytic serine residue form the oxyanion-stabilizing loop (residues 137-139; Figure 7). The S site comprises strand E2, which forms one side of the specificity pocket, with the A157 amide proton and carbonyl group accessible to form backbone to backbone H-bonds with the substrate P3 residue according to a classical chymotrypsin-like substrate interaction. The recognition pocket is shallow and apolar, being formed by the methyl groups of A157 and L135 on the side and by the F154 aromatic ring at the bottom (Figure 7). Features of this pocket have been predicted from modelling studies (Pizzi *et al*., 1994).

The S′ site is constituted by the ending of the A1 (T38) and the beginning of the B1 strands, as well as the A1-B1 short loop (Figure 7). Strand A1 packs directly with the NS4A-activating peptide in the ns3-4a structure, thus explaining a direct influence of the NS4A peptide on the P′ portion of the substrate (leaving group). A more detailed insight into the S binding site, the detailed identification of the residues involved and their role in the interaction with substrate-based inhibitors are presented in the accompanying paper (Cicero *et al*., 1999).

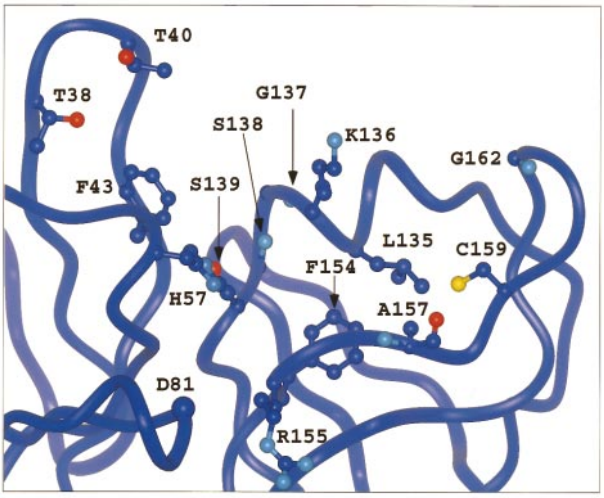

Figure 7. Zoomed view of the substrate interaction region. The S′ and S regions range approximately from strand A1 (identified by the position of T38) to loop E2-F2 (G162), encompassing the rather flat surface defined by the E2 strand (F154 bottom of recognition pocket, A157 H-bond candidate with substrate P3 partner) on one side and the α2 helix (oxyanion hole G137-S139; L135 delimiting the top side of the recognition pocket) on the other side.

The Zn-binding site

In a previous study, we showed that a zinc ion is required for the structural integrity and activity of the NS3 proteinase and, from modelling studies, the coordination site was predicted to be comprised of loops F1-A2 (C97 and C99) and B2-C2 (C145 and H149: De Francesco *et al*., 1996). Subsequent publication of the crystallographic structures confirmed the predictions. In the crystallographic structures, the zinc coordination is essentially tetrahedral. However, the imidazole moiety and the zinc atom are too distant to be directly bound, and the authors postulated the presence of a water molecule acting as a bridge (Love & Hostomska, 1996; Kim *et al*., 1996).

The zinc-binding site in solution, as determined by NMR, is illustrated by Figure 8. Our data identify the N^{δ} of H149 as the one involved in the binding to the zinc ion. This structural result is in agreement with our previous findings on the tautomeric state of H149, with N^{ε} being in the α state and N^{δ} being in the β state (Urbani *et al*., 1998). We

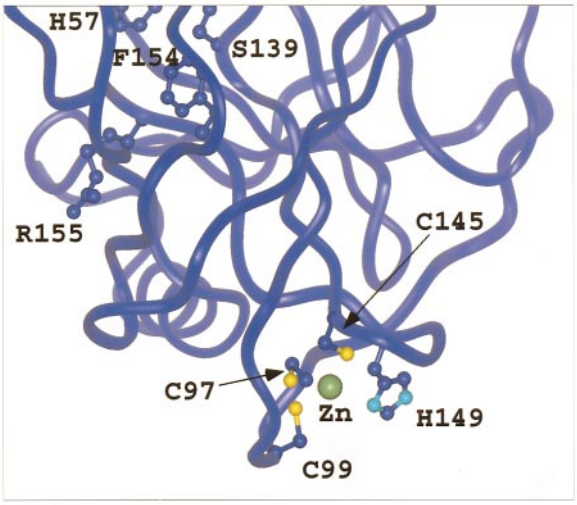

Figure 8. Zoomed view of the zinc-binding site. The residues chelating the zinc ion are shown; i.e., it is evident the unusually long stretch that links C145 to H149. This longer linker could be required to allow the conformational switch of H149 and leave space accessible to the zinc ion.

also found that the imidazole moiety of H149 modulates the accessibility of the zinc ion, allowing an ''open'' and a ''closed'' conformation in the protonated and unprotonated state, respectively, which interconvert on the 100 ms timescale (Urbani *et al*., 1998). In this respect, one should notice that in the ns3 structure (Love *et al*., 1996) H149 is postulated to participate in metal coordination in only two of the three molecules in the asymmetric unit, whereas in the third one the imidazoyl side-chain moves away. In the experimental conditions used in this structural study, the closed conformation is dominant. In the NMR-derived structure, the imidazole moiety was not subject to specific restraints that would force the coordination with the zinc atom (see Materials and Methods) and in all the resulting structures it is positioned at a distance too far to directly chelate the zinc atom. This result is in agreement with our previous findings, in which H149 is ligated to the metal using the δ1 N through a hydroxyl group (Urbani *et al*., 1998).

Most of the chymotrypsin-like proteinases have disulphide bridges that are believed to maintain the relative orientations of the residues involved in catalysis (Lesk & Fordham, 1996). Disulphide bridges present in these extracellular serine proteinases are unlikely to be stable in the reducing intracellular milieu. Since NS3 is an intracellular proteinase, we proposed that the zinc-binding site is used to stabilise the relative orientation of the two β-barrel domains, thus indirectly influencing the position of the catalytic triad, which is located in the crevice between the domains (De Francesco *et al*., 1996).

However, we observe also that the histidine residue is highly conserved in all the HCV strains and HCV-related viruses. This fact, together with the unusual location of the site, i.e. completely solvent-exposed, and its dynamic behaviour in solution may suggest that there could be a biologically relevant function, not yet clarified, linked to this zinc-binding site. Hepatitis C is a small virus that encodes only six non-structural proteins. Therefore, as observed in other viruses, multi-functional proteins are a strategy to minimise the number of agents needed during viral replication. Then the optimisation of the role of the zinc ion and the peculiar features of its binding site would be just part of the same biomolecular strategy.

Materials and Methods

Expression, purification and solubilisation

Escherichia coli cells BL21(DE3) were transformed with a plasmid containing the cDNA coding for the serine proteinase domain of NS3 under the control of the bacteriophage T7 gene 10 promoter. A solubilisation tag (ASKKKK) was inserted at the C terminus of the NS3 enzyme sequence (Steinkühler *et al*., 1998). The ^{15}N, $^{15}N/^{13}C$ and $^{15}N/^{13}C/^{2}H$ uniformly labelled samples were obtained allowing the cells to grow in M9 minimum medium supplemented with 1 g/l [^{15}N]ammonium sulphate (Martek), 2.0 g/l [^{13}C]glucose (Martek). A further addition to the medium of 6.8 mg/l $ZnCl_2$ was necessary, since the protein is a zinc-binding proteinase. For the perdeuterated sample, growth was carried out in 99 % $^{2}H_2O$ (Martek). The ^{15}N selectively labelled sample (Leu, Val and Ala) was obtained by incorporation of 0.33 g/l each of ^{15}N selectively labelled Leu, Val and Ala in M9 modified medium. The NMR samples were prepared by dialysis against a buffer containing 20 mM sodium phosphate, 4 % deuterated glycerol (Isotec Inc.), 3 mM DTT, 1.5 mM Chaps (pH 6.3), 5 mM NaN_3. The protein concentration was in the range of 0.7-0.9 mM. The aggregation state of the samples was verified with dynamic light-scattering and sedimentation equilibrium studies. The solutions were monodispersed and the protein behaviour was compatible with a monomeric state in solution (S. Di Marco & M. Sollazzo, unpublished results).

Data collection and assignment

All the NMR experiments were acquired at 298 K on a Bruker AMX 500 MHz, Varian Unity Plus 600 MHz, Bruker DMX 600, and 800 MHz all equipped with *z*-shielded gradient triple resonance probes. Spectral assignments were obtained using the following 3D experiments: CT-HNCO, CT-HNCA, CT-HNCOCA, CBCACONH, HNCAHA, HN(COCA)HA, HACACO, (H)CCH-COSY, H(C)CH-COSY, H(C)CH-TOCSY, (H)CCH-TOCSY, C-CONH-TOCSY, ^{15}N edited ^{1}H TOCSY (Clore & Gronenborn, 1991a,b; Bax & Grzesiek, 1993). The following 3D experiments were acquired for the perdeuterated sample: CT-HNCA, CT-HNCOCA, HNCOCACB, HNCACB (Yamazaki *et al*., 1994).

NOE-type 3D experiments were acquired (Clore & Gronenborn, 1991a): on the perdeuterated sample, ^{15}N edited NOESY (80 ms); on the double-labelled sample, two sets of ^{13}C edited NOESY (80, 100 and 150 ms), one

optimized for aliphatic and another set for aromatic residues; on the ^{15}N single-labelled sample, ^{15}N edited NOESY (80 and 150 ms) and ^{15}N edited ROESY (20 ms). On the ^{15}N selectively labelled Leu, Ala and Val sample, ^{15}N edited NOESY (100 ms) and TOCSY (24 ms) were also acquired.

Coupling constants and stereospecific assignments: on the ^{15}N sample, 3D experiment HNHA (Kuboniwa *et al*., 1994); on the double labelled sample, 3D HAHB (Grzesiek *et al*., 1994) experiment, and 2D CO and N decoupled (Hu *et al*., 1997). On the 10% ^{13}C-labelled sample was acquired a CT-HSQC for the stereospecific assingment of methyl groups of Leu and Val (Neri *et al*., 1989).

Spectra were processed using NMRPipe (Delaglio *et al*., 1995) and analysed using NMRView (Johnson & Blevins, 1994) software packages.

Structure calculation

Approximate interproton distances were derived from the multidimensional NOE spectra (Clore & Gronenborn, 1991a). NOEs were grouped into three distance ranges 1.8-2.8 Å (1.8-3.0 Å for NOEs involving HN protons), 1.8-3.4 Å (1.8-3.6 Å for NOEs involving HN protons) and 1.8-5.0 Å (1.8-6.0 for NOEs involving methyl groups); 0.6 Å was added to the upper bounds of the strong and medium NOEs involving methyl groups. No constraint was included for the zinc-binding site during the early stages of the calculation. After verification that, in each structure, the ligands were always disposed approximately in a tetrahedral geometry, the zinc atom was subsequently incorporated into the calculations using six distance and one valence angle restraint involving the cysteine residues and the explicit zinc atom (Omichinski *et al*., 1990). H149 was not constrained to bind the zinc atom, since our previous results suggested that the coordination could be mediated through a hydroxyl group (Urbani *et al*., 1998). Protein backbone hydrogen-bonding restraints ($d_{NH\text{-}O} = 1.6\text{-}2.9$ Å, $d_{N\text{-}O} = 2.4\text{-}3.6$ Å) within areas of regular secondary structure were introduced during the final stages of refinement using standard NMR criteria based on backbone NOE $^3J_{HN\alpha H}$ coupling constants, supplemented with secondary ^{13}C shifts. The ϕ, ψ and $\chi 1$ torsion angle restraints were derived from homo- and heteronuclear three-bond coupling constant data, employing as minimum ranges $\pm 25^\circ$, $\pm 40^\circ$ and $\pm 30^\circ$, respectively. The structures were calculated by simulated annealing with the program X-Plor 3.851 (Brunger, 1993) on a SGI O2 R10000 platform, using a protocol described by Omichinski *et al*. (1997). Figures and statistical analysis were generated using the program InsightII (Molecular Simulations Inc.).

Brookhaven Protein DataBank

The coordinates of the final 20 simulated annealing structures have been deposited in the Brookhaven Protein DataBank, accession code 1bt7.

Acknowledgements

We thank Professor C. Griesinger, of the Frankfurt University Centre of Biomolecular NMR, (Frankfurt-am-Main) for access to the Large Scale Facility 800 and 600 MHz instruments, and Dr Marcus Maurer and Dr Teresa Carlomagno for the recording of the ^{13}C and ^{15}N NOESY at 800 MHz, and one of the ^{13}C edited H(C)CH TOCSY experiments on the 600 MHz. G.B. thanks all the people working at that facility for their kindness during his stay there. We thank Dr Bruce Johnson, (Merck Research Laboratories, New Jersey, USA) for his help in the collection of data on the perdeuterated sample. We are deeply indebted to Professor Stephan Grzesiek (Forschungszentrum, Jülich, Germany) for his help and recording of the experiments aimed at determining the dihedral angles.

References

Bachovchin, W. W. (1985). Confirmation of the assignment of the low field proton resonance of serine proteinases by using specifically nitrogen-15 labeled enzyme. *Proc. Natl Acad. Sci. USA,* **82**, 7948-7951.

Bartenschlager, R. (1997). Molecular targets in inhibition of hepatitis C virus replication. *Ann. Chem. Chemother.* **8**, 281-301.

Bax, A. & Grzesiek, S. (1993). Methodological advances in protein NMR. *Acct. Chem. Res.* **26**, 131-138.

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

Blevins, R. A. & Tulinsky, A. (1985). The refinement and the structure of the dimer of α-chymotrypsin at 1.67 Å resolution. *J. Biol. Chem.* **260**, 4264-4275.

Bode, W., Chen, Z., Bartels, K., Kutzbach, C. & Schmidt, G. (1983). Refined 2.0 Å X-ray structure of porcine pancreatic kallikrein A. *J. Mol. Biol.* **164**, 237-282.

Bode, W., Walter, J., Huber, R., Wenzel, H. R. & Tschesche, H. (1984). The refined 2.2 Å (0.22 nm) X-ray crystal structure of the ternary complex formed by bovine trypsinogen, valine-valine and the Arg15 analogue of bovine pancreatic trypsin inhibitor. *Eur. J. Biochem.* **144**, 185-190.

Bode, W., Mayr, I., Baumann, U., Huber, R., Stone, R. S. & Hofsteenge, J. (1989). The refined 1.9 Å X-ray structure of human alpha-thrombin. *EMBO J.* **8**, 3467-3475.

Brunger, A. T. (1993). *X-Plor Version 3.1: A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.

Chang, J. Y. (1986). The structures and proteolytic specificities of autolysed human thrombin. *Biochem. J.* **240**, 797-802.

Choi, H. K., Lu, G., Lee, S., Wengler, G. & Rossmann, M. G. (1997). Structure of Semliki forest virus core protein. *Proteins: Struct. Funct. Genet.* **27**, 345-359.

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

Cicero, D. O., Barbato, G., Koch, U., Ingallinella, P., Bianchi, E., Nardi, M. C., Steinkhüler, C., Cortese, R., Matassa, V., De Francesco, R., Pessi, A. & Bazzo, R. (1999). Structural characterization of the interactions of optimized product inhibitors with the protease domain of human hepatitis C virus by NMR and modelling studies. *J. Mol. Biol.*

Clore, G. M. & Gronenborn, A. M. (1991a). Structures of larger proteins in solution: three and four dimensional heteronuclear NMR spectroscopy. *Science,* **252**, 1390-1399.

Clore, G. M. & Gronenborn, A. M. (1991b). Applications of three and four dimensional heteronuclear NMR

spectroscopy to protein structure determination. *Prog. NMR Spectrosc.* **23**, 43-92.

Coombs, G. S., Dang, A. T., Madison, E. L. & Corey, D. R. (1996). Distinct mechanisms contribute to stringent substrate specificity of tissue-type plasminogen activator. *J. Biol. Chem.* **271**, 4461-4467.

De Francesco, R., Urbani, A., Nardi, M. C., Tomei, L., Steinkühler, C. & Tramontano, A. (1996). A zinc site in viral serine proteinases. *Biochemistry,* **35**, 13282-13287.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR,* **6**, 277-293.

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

Fersht, A. (1984). *Enzyme Structure and Mechanism*, 2nd. edit., W. H. Freeman, San Francisco, CA.

Frey, P. A., Whitt, S. A. & Tobin, J. B. (1994). A low-barrier hydrogen bond in the catalytic triad of serine proteinases. *Science,* **82**, 1927-1930.

Fujinaga, M. & James, M. N. G. (1987). Rat submaxillary gland serine proteinase tonin structure resolution and refinement at 1.8 Å resolution. *J. Mol. Biol.* **195**, 373-396.

Fujinaga, M., Delbaere, T. J., Brayer, G. D. & James, M. N. G. (1985). Refined structure of α-lytic proteinase at 1.7 Å resolution. *J. Mol. Biol.* **183**, 479-502.

Greer, J. (1990). Comparative modeling methods: Application to the family of the mammalian serine proteinases. *Proteins: Struct. Funct. Genet.* **7**, 317-334.

Grzesiek, S., Kuboniwa, H., Hinck, A. P. & Bax, A. (1994). Multiple quantum line narrowing for measurements of Hα-Hβ *J* couplings in isotopically enriched proteins. *J. Am. Chem. Soc.* **117**, 5312-5315.

Houghton, M. (1996). Hepatatis C virus. In *Fields Virology* (Fields, B. N., Knipe, D. M. & Howley, P. M., eds), 3rd. edit., pp. 1035-1058, River Press, New York.

Hu, J. S., Grzesiek, S. & Bax, A. (1997). Two dimensional NMR methods for determining χ1 angles of aromatic residues in proteins from three bond $J_{C\text{-}C\gamma}$ and J $N_{C\gamma}$ couplings. *J. Am. Chem. Soc.* **119**, 1803-1804.

Ingallinella, P., Altamura, S., Bianchi, E., Taliani, M., Ingenito, R., Cortese, R., De Francesco, R., Steinkühler, C. & Pessi, A. (1998). Potent inhibitors of human hepatitis C virus NS3 proteinase are obtained by optimising the cleavage products. *Biochemistry,* **37**, 8906-8914.

Kim, J. L., Morgenstern, K. A., Lin, C., Fox, T., Dwyer, M. D., Landro, J. A., Chambers, S. P., Markland, W., Lepre, C. A., O'Malley, E. T., Harbeson, S. L., Rice, C. M., Murcko, M. A., Caron, P. R. & Thomson, J. A. (1996). Crystal structure of the hepatitis C virus NS3 proteinase domain complexed with a synthetic NS4A cofactor peptide. *Cell,* **87**, 343-355.

Kuboniwa, H., Grzesiek, S., Delaglio, F. & Bax, A. (1994). Measurement of HN-H alpha *J* coupling in calcium free calmodulin using new 2D and 3D water flip back methods. *J. Biomol. NMR,* **4**, 871-878.

Kuo, G., Choo, Q.-L., Alter, H. J., Gitnick, G. L., Redecker, A. G., Purcell, R. H., Myamura, T., Dienstag, J. L., Alter, M. J., Syevens, C. E., Tagtmeier, G. E., Bonino, F., Colombo, M., Lee, W.-S., Kuo, C., Berger, K., Shister, J. R., Overby, L. R., Bradley, D. W. & Houghton, M. (1989). An assay for circulating antibodies to a major etiologic virus human non-A non-B hepatitis. *Science,* **244**, 362-364.

Johnson, B. & Blevins, R. A. (1994). NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR,* **4**, 603-614.

Landro, J. A., Raybuck, S. A., Luong, Y. P. C., O'Malley, E. T., Haberson, S. L., Morgenstern, K. A., Rao, G. & Livingston, D. J. (1997). Mechanistic role of an NS4A peptide cofactor with the truncated NS3 proteinase of hepatitis C virus: elucidation of the NS4A stimulatory effect via kinetic analysis and inhibitor mapping. *Biochemistry,* **36**, 9340-9348.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). Procheck: a program to check the stereochemical quality of protein structure. *J. Appl. Crystallog.* **26**, 283-291.

Lesk, A. & Fordham, W. D. (1996). Conservation and variability in the structures of serine proteinases of chymotrypsin family. *J. Mol. Biol.* **258**, 501-537.

Love, R. A., Parge, H. E., Wickersham, J. A., Hostomsky, Z., Habuka, N., Moomaw, E. W., Adachi, T. & Hostomska, Z. (1996). The crystal strcture of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site. *Cell,* **87**, 331-342.

Love, R. A., Parge, H. E., Wickersham, J. A., Hostomsky, Z., Habuka, N., Moomaw, E. W., Adachi, T., Margosiak, S., Dagostino, E. & Hostomska, Z. (1998). The conformation of hepatitis C virus NS3 proteinase with and without NS4A: a structural basis for the activation of the enzyme by its cofactor. *Clin. Diagnost. Virol.* **10**, 151-156.

Markley, J. L. (1978). Hydrogen bonds in serine proteinases and their complexes with protein proteinase inhibitors. *Biochemistry,* **17**, 4648-4656.

Meyer, E., Cole, G. & Radhakrishnan, R. (1988). Structure of native porcine pancreatic elastase at 1.65 Å resolution. *Acta Crystallog. sect. B,* **44**, 26-38.

Moult, J., Sussman, F. & James, M. N. G. (1985). Electron density calculation as an extension of protein structure refinement, *Streptomyces griseus* proteinase A at 1.5 Å resolution. *J. Mol. Biol.* **182**, 555-566.

Neddermann, P., Tomei, L., Steinkühler, C., Gallinari, P., Tramontano, A. & De Francesco, R. (1997). The non structural proteins of the hepatitis C virus: structure and functions. *Biol. Chem.* **378**, 469-476.

Neri, D., Szyperski, T., Otting, G., Senn, H. & Wüthrich, K. (1989). Specific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of 434 repressor by biosynthetically directed fractional ^{13}C labeling. *Biochemistry,* **28**, 7510-7516.

Nilges, M., Clore, G. M. & Gronenborn, A. M. (1988). Determination of three dimensional structures of proteins from interproton distance data by hybrid distance geometry simulated annealing calculations. *FEBS Letters,* **229**, 317-324.

Omichinski, J. G., Clore, G. M., Appella, E., Sakaguchi, K. & Gronenborn, A. M. (1990). High resolution three dimensional structure of a single zinc finger from a human enhancer binding protein in solution. *Biochemistry,* **29**, 9324-9334.

Omichinski, J. G., Pedone, P. V., Felsenfeld, G., Gronenborn, A. M. & Clore, G. M. (1997). The solution structure of a specific GAGA factor-DNA complex reveals a modular binding mode. *Nature Struct. Biol.* **4**, 122-132.

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

Polgar, L. (1989). *Mechanisms of Proteinase Action*, CRC Press, Boca Raton, FA.

Read, R. J., Fujinaga, M., Sielecki, A. R. & James, M. N. G. (1983). Structure of the complex of *Streptomyces griseus* proteinase B and the third domain of the turkey ovomucoid inhibitor at 1.8 Å resolution. *Biochemistry,* **22**, 4420-4433.

Remington, S. J., Woodbury, R. G., Reynolds, R. A., Matthews, B. A. & Neurath, H. (1988). The structure of rat mast cell proteinase II at 1.9 Å resolution. *Biochemistry,* **27**, 8097-8105.

Simmonds, P. (1994). Variability of hepatitis C virus. In *Current Studies in Hematology and Blood Transfusion* (Reesink, H. W., ed.), pp. 12-35, Kargel, Basel, Switzerland.

Steinkühler, C., Biasol, G., Brunetti, M., Urbani, A., Koch, U., Cortese, R., Pessi, A. & De Francesco, R. (1998). Product inhibition of the hepatitis C virus NS3 proteinase. *Biochemistry,* **37**, 8899-8905.

Tong, L., Wengler, G. & Rossmann, M. G. (1993). Refined structure of Sindbis virus core protein and comparison with other chymotripsin-like serine proteinase structures. *J. Mol. Biol.* **230**, 228-247.

Urbani, A., Bazzo, R., Nardi, M. C., Cicero, D. O., De Francesco, R., Steinkühler, C. & Barbato, G. (1998). The metal binding site of the hepatitis C virus NS3 proteinase. *J. Biol. Chem.* **273**, 18760-18769.

Yamazaki, T., Lee, W., Arrowsmith, C. H., Muhandiram, D. R. & Kay, L. E. (1994). A suite of triple resonance experiments for the backbone assignments of ^{15}N, ^{13}C, ^{2}H labeled proteins with high sensitivity. *J. Am. Chem. Soc.* **116**, 11655-11666.

Yan, Y., Li, Y., Munshi, S., Sardana, V., Cole, J., Sardana, M., Steinkuehler, C., Tomei, L., De Francesco, R., Kuo, L. C. & Chen, Z. (1998). Complex of NS3 proteinase and NS4A peptide of BK strain hepatitis C virus: a 2.2 Å resolution structure in a hexagonal crystal form. *Protein Sci.* **7**, 837-847.

Yasutake, A. & Powers, J. C. (1981). Reactivity of human leukocyte elastase and porcine pancreatic elastase toward peptide 4-nitroanilides containing model desmosine residues. *Biochemistry,* **20**, 3675-3679.

Edited by P. E. Wright

(Received 15 December 1998; received in revised form 19 March 1999; accepted 22 March 1999)