doi:10.1016/j.jmb.2005.09.074 J. Mol. Biol. (2005) 354, 854–871

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

# Dual Modes of Modification of Hepatitis A Virus 3C Protease by a Serine-derived β-Lactone: Selective Crystallization and Formation of a Functional Catalytic Triad in the Active Site

**Jiang Yin[1]†, Ernst M. Bergmann[1,3]†, Maia M. Cherney[1], Manjinder S. Lall[2] Rajendra P. Jain[2], John C. Vederas[2] and Michael N. G. James[1,3]***

[1]*CIHR Group in Protein Structure and Function Department of Biochemistry University of Alberta Edmonton, Alta. Canada T6G 2H7*

[2]*Department of Chemistry University of Alberta Edmonton, Alta. Canada T6G 2G2*

[3]*Alberta Synchrotron Institute University of Alberta Edmonton, Alta. Canada T6G 2E1*

**Corresponding author*

Hepatitis A virus (HAV) 3C proteinase is a member of the picornain cysteine proteases responsible for the processing of the viral polyprotein, a function essential for viral maturation and infectivity. This and its structural similarity to other 3C and 3C-like proteases make it an attractive target for the development of antiviral drugs. Previous solution NMR studies have shown that a Cys24Ser (C24S) variant of HAV 3C protein, which displays catalytic properties indistinguishable from the native enzyme, is irreversibly inactivated by *N*-benzyloxycarbonyl-L-serine-β-lactone (1a) through alkylation of the sulfur atom at the active site Cys172. However, crystallization of an enzyme–inhibitor adduct from the reaction mixture followed by X-ray structural analysis shows only covalent modification of the ε2-nitrogen of the surface His102 by the β-lactone with no reaction at Cys172. Re-examination of the heteronuclear multiple quantum coherence (HMQC) NMR spectra of the enzyme–inhibitor mixture indicates that dual modes of single covalent modification occur with a $\geq$3:1 ratio of S-alkylation of Cys172 to N-alkylation of His102. The latter product crystallizes readily, probably due to the interaction between the phenyl ring of the *N*-benzyloxycarbonyl (N-Cbz) moiety and a hydrophobic pocket of a neighboring protein molecule in the crystal. Furthermore, significant structural changes are observed in the active site of the 3C protease, which lead to the formation of a functional catalytic triad with Asp84 accepting one hydrogen bond from His44. Although the 3C protease modified at Cys172 is catalytically inactive, the singly modified His102 $N^{\varepsilon 2}$-alkylated protein displays a significant level of enzymatic activity, which can be further modified/inhibited by *N*-iodoacetyl-valine-phenylalanine-amide (IVF) (in solution and in crystal) or excessive amount of the same β-lactone inhibitor (in solution). The success of soaking IVF into HAV 3C–1a crystals demonstrates the usefulness of this new crystal form in the study of enzyme–inhibitor interactions in the proteolytic active site.



*Keywords:* hepatitis A; picornavirus; 3C protease; inhibitor and antiviral drug design; β-lactone

## Introduction

The picornavirus family (*Picornaviridae*) has over 200 members and contains many important pathogens such as poliovirus (PV), hepatitis A virus (HAV), foot-and-mouth disease virus (FMDV) and human rhinovirus (HRV).[1] Hepatitis A occurs most commonly in developing countries and usually spreads through contamination of drinking water

† J.Y. & E.M.B. contributed equally to this work.

Abbreviations used: r.m.s.d., root-mean-square difference; FMDV, foot-and-mouth disease virus; HRV, human rhinovirus; PV, poliovirus; IVF, *N*-iodoacetyl-Val-Phe-amide; Cbz, *N*-benzyloxycarbonyl; HMQC, heteronuclear multiple quantum coherence.

E-mail address of the corresponding author: michael.james@ualberta.ca

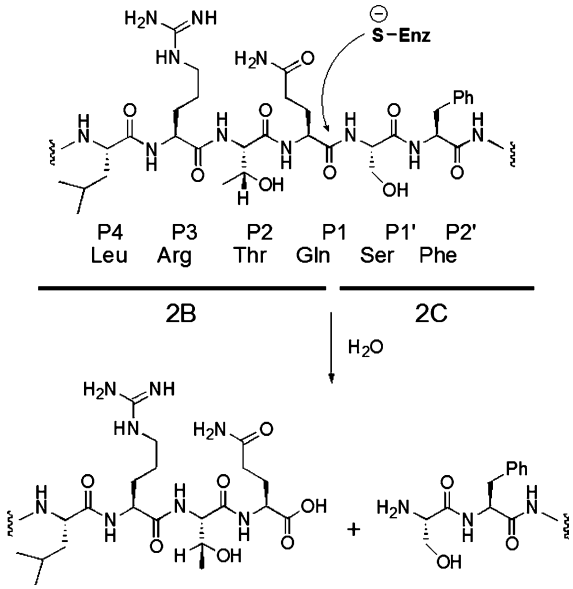



**Figure 1.** A preferred cleavage site of HAV 3C proteinase showing the sequence at the 2B–2C junction of the 250 kDa viral polyprotein.

or food.[2] The genomes of picornaviruses are of plus polarity, which, upon entry into the cytoplasm of a host cell, are translated into ~250 kDa polyproteins that subsequently undergo multiple cleavages by the 2A or 3C proteases to produce mature viral proteins. In HAV, the 3C protease is the key enzyme for co- and post-translational processing of the viral polyprotein and is of critical importance for viral genome replication.[3,4] Hence it is an appealing target for the development of new antiviral agents. This is especially true because many other viruses, such as members of the *Coronaviridae* family, including the causative agent of severe acute respiratory syndrome (SARS), require 3C-like proteases for processing the viral polyprotein into mature viral proteins.[5]

Crystal structures for several picornaviral 3C proteases such as those in HAV,[6–8] HRV2&14,[9,10] PV[11] and FMDV[12] have been reported. In all cases, these cysteine proteases show three-dimensional structures that are similar to those of the chymotrypsin family of serine proteases with the active site serine being replaced by a cysteine residue. The thiolate of Cys172 of HAV 3C protease is the nucleophile that attacks the substrate carbonyl carbon with general-base assistance from His44.[4] Formation of the tetrahedral intermediate is promoted by an electrophilic oxyanion hole. The enzyme typically binds the substrate at four or five residues ($P_5$ to $P_1$; Figure 1) on the N-terminal side of the scissile bond and two or three residues ($P_1'$ to $P_3'$) following the site of cleavage. Hydrogen bonding between the carbonyl oxygen in the side-chain of the $P_1$ glutamine of a substrate and $N^{\varepsilon 2}$ of His191 in the $S_1$ enzyme subsite is a key recognition event.[4,13] An octapeptide substrate mimicking the 2B/2C junction in the HAV polyprotein has a $k_{cat}$ value of 1.8 $s^{-1}$ and a value $K_m$ of 2.1 mM at pH 7.5, which makes HAV 3C a weakly binding and slow-acting protease with good specificity.[13]

Among the compounds that inhibit HAV 3C protease,[14,15] *N*-benzyloxycarbonyl (Cbz) L-serine β-lactone (1a) (Figure 2) is a potent irreversible inactivator with $k_{inact} = 0.70\ min^{-1}$, $K_i = 1.84 \times 10^{-4}$ M and $k_{inact}/K_i = 3800\ M^{-1}\ min^{-1}$.[16,17] We previously inactivated a C24S variant of HAV 3C, wherein a non-essential surface residue had been replaced and which displayed catalytic properties indistinguishable from the native enzyme, with a $^{13}C$-labeled β-lactone 1b.[16] Analysis using heteronuclear multiple quantum coherence (HMQC) NMR spectroscopy clearly showed that the main reaction of this enzyme with 1b involves nucleophilic attack of the active site cysteine (Cys172) on the β-position of the lactone ring (Figure 2) based on $^1H$ and $^{13}C$ chemical shifts of the labeled methylene group. We now report that crystallization of the reaction mixture of 1a and HAV 3C (C24S) enzyme affords a 3C protein that, while keeping a free thiol at Cys172, is modified only at the $N^{\varepsilon 2}$ atom of a surface histidine, namely His102. This species, which is formed by a minor reaction pathway and is barely detectable by HMQC NMR spectroscopy, crystallizes readily. The appearance of dual modes of enzyme modification, though inherently not surprising from a chemical reactivity point of view, provides a cautionary note for the use of single spectroscopic techniques (i.e. crystallographic analysis or NMR) to evaluate the mechanism of enzyme inhibition.

Previously reported crystal structures of HAV 3C protease showed that the putative third member of the catalytic triad, Asp84, does not form an interaction with the general base catalyst, His44, suggesting a dyad catalytic mechanism.[7,8,18] On the other hand, the crystal structures of 3C proteases from PV, HRV2, HRV14 and FMDV all showed a functional assembly of the three catalytic residues similar to that of the chymotrypsin family enzymes.[9–12] Here, we report the crystal structure

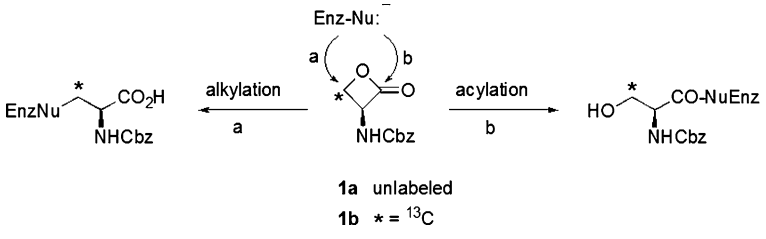



**Figure 2.** Possible modes of nucleophilic attack on *N*-Cbz-L-serine β-lactone (1). Only path a is observed with HAV 3C protease, with the nucleophiles (Nu:) being either the sulfur of Cys172 or the ε2-nitrogen of His102.

of HAV 3C protease solved at the highest resolution known to date for 3C and 3C-like viral proteases. In the new structure, the side-chain of Asp84 is in the right orientation and forms a hydrogen bond with the imidazole ring of His44, producing a cononical catalytic triad in chymotrypsin-like proteases. As the active site is unoccupied in the crystallized protease, we subsequently soaked a second inhibitor (IVF) into the pregrown HAV 3C–1a crystals and successfully obtained the complex structure. We believe that the HAV 3C protein modified at His102 may assist future crystallographic studies of active site binding of other inhibitors.

## Results

### Reaction of *N*-Cbz-L-serine β-lactone with HAV 3C protease C24S

The required irreversible inhibitor, *N*-Cbz-L-serine β-lactone (1a), is readily synthesized by Mitsunobu cyclization of commercially available *N*-Cbz-L-serine.[19] Depending on the particular species and the solvent, nucleophiles can attack such β-lactones either at the β-carbon to give alkylated product or they can react at the carbonyl carbon to give acylated material (Figure 2).[19,20] "Soft" nucleophiles such as thiolate tend to attack the β-carbon whereas "hard" nucleophiles such as alkoxide or hydroxide target the carbonyl.[21] Nitrogen nucleophiles may attack at either site.[22] In previous studies, serine or threonine residues in proteins were shown to attack β-lactones at the carbonyl to give acylated products.[23–27] β-Lactone 1a, like other α-amino-β-lactones lacking a β-substituent, is reasonably robust in acid but displays low stability in basic aqueous media wherein hydrolysis to the corresponding hydroxy acid (*N*-Cbz-L-serine) is fairly rapid.[21] However, its half-life in phosphate buffer at pH 7.5 is 76 min, which is sufficiently long for the inhibition studies with HAV 3C protease C24S variant.[16]

This protease variant still has the intact active site Cys172 and displays the same catalytic properties as the wild-type enzyme. The enzyme activity is easily monitored using a fluorometric assay at an enzyme concentration of 0.1 μM and a fluorogenic peptide substrate Dabcyl-GLRTQSFS-Edans. The β-lactone 1a irreversibly inactivates HAV 3C protease with a $k_{inact}$ value of 0.70 $min^{-1}$. This process is not affected by the presence of a tenfold molar excess of dithiothreitol, indicating that the inhibition is quite specific.[16] We have previously shown that the corresponding ring-opened hydrolysis product, *N*-Cbz-L-serine, exhibits no significant inhibition of HAV 3C at a concentration of 100 μM. In our earlier studies we used electrospray mass spectrometry to show that the mass difference between enzyme–inhibitor complex (24,101 Da) and the enzyme alone (23,882) is 219 Da, which is within experimental error of the calculated mass of one inhibitor molecule (221 Da). To determine the mechanism of inhibition,

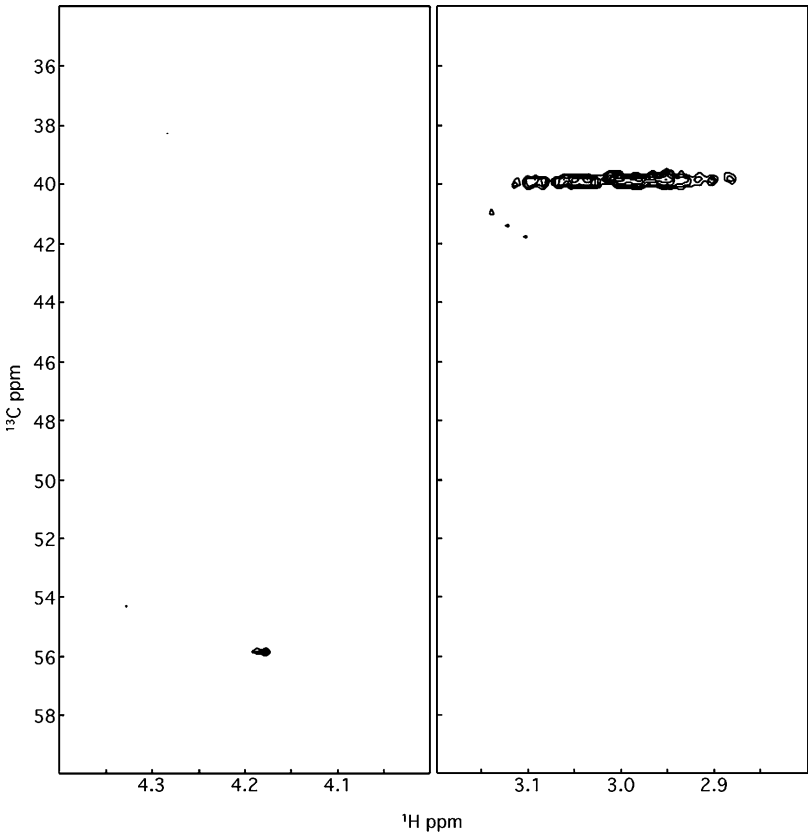
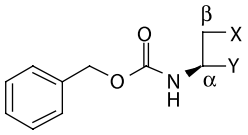
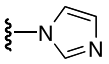



**Figure 3.** Sections of $^{1}H/^{13}C$ 600 MHz HMQC spectra of HAV 3C protease C24S modified by a single molecule $^{13}C$-labeled *N*-Cbz-L-serine β-lactone (1b). The left hand panel shows the labeled methylene signal for the enzyme modified at His102, whereas the right hand panel shows the analogous signal for enzyme modified at the active site Cys172. The signal in the right hand panel appears very broad on the proton axis because the contour slice was cut low to be comparable to that used to detect the weak signal in the left hand panel.

**Table 1.** $^{1}H$ and $^{13}C$ chemical shifts of 1a and model compounds




| Compound | X | Y | β-Carbon (ppm) | β-Hydrogen atoms (ppm)[a] |
|---|---|---|---|---|
| 1a[b] | O ——— CO | | 68 | 4.4 |
| *N*-Cbz-L-Ser[b] | OH | $CO_2H$ | 62 | 3.8 |
| 2[c] | SMe | $CO_2H$ | 37 | 2.9 |
| 3[c] | OH | COSEt | 64 | 3.9 |
| 4[c] | OH | $CO_2Me$ | 62 | 4.h |
| 5[c,d] | OMe | $CO_2H$ | 73 | 3.7 |
| 6[e] |  | $CO_2H$ | 52 | 4.4 |

[a] The average values of chemical shifts of diastereotopic β-hydrogen atoms are given as these were not usually resolved in HMQC spectra.
[b] $^{1}H/^{13}C$ HMQC conditions: $^{2}H_2O$ at pD 5.0, 6% DMSO-$d_6$.
[c] 20 mM $Na_2PO_4/^{2}H_2O$ at pD 7.5, 6% DMSO-$d_6$.
[d] Racemic.
[e] $CD_3OD$ used due to limited solubility in $^{2}H_2O$.

$^{13}C$-labeled β-lactone (1b) was synthesized and allowed to inactivate HAV 3C (C24S) protease. The enzyme–inhibitor complex was then analyzed by HMQC NMR and a new cross-peak at $\delta_c$ 40 ppm (Figure 3) was detected. Comparison to $^{1}H$ and $^{13}C$ chemical shift model compounds (Table 1) clearly showed that most of the inhibitor 1b had reacted at its β-carbon and alkylated the active site thiol of Cys172.[16] However, when X-ray crystallographic analysis (see below) indicated that reaction had occurred at the ε2-nitrogen of the surface His102, the original HMQC NMR spectra were re-examined and a weak cross-peak at $\delta_c$ 55.9 and $\delta_H$ 4.18 could be seen (Figure 3). Its position is slightly displaced from the corresponding chemical shifts ($\delta_c$ 52 and $\delta_H$ 4.4) of the model adduct (Table 1) made by reaction of 1a with imidazole, but this is likely due to the presence of nearby protein residues, the additional substitution on the imidazole ring as well as solvent effects. Approximate quantitation of cross-peak intensities for the sulfur and histidine adducts suggests that the ratio of sulfur to nitrogen alkylation is considerably greater than 3:1.

The crystallization of the minor component with His102 modified proceeded readily. The resulting 3C–1a crystals could not be easily redissolved under non-denaturing conditions. Some level of catalytic activity was observed when a single 3C–1a crystal was crushed and resuspended for protease activity assay using the fluorometric method as described (Figure 4). This activity can be irreversibly destroyed by a dipeptide inactivator, *N*-iodoacetyl-Val-Phe-amide, which is known to alkylate the active site cysteine 172 based on both biochemical and crystallographic evidence.[8,28] Prolonged exposure of the solution containing the two singly modified forms of HAV 3C protease to a 100-fold excess of inhibitor 1a leads to complete loss of the enzymatic activity and multiple covalent modification (i.e. 1:2 enzyme/inhibitor complex) based on electrospray mass spectrometry (data not shown).

## Crystallization and structure determination

The formation of the complex of HAV 3C protease variant C24S with *N*-Cbz-L-Ser-β-lactone (1a) was carried out in the presence of 25-fold excess of the inhibitor at a slightly acidic pH (pH5.4). The complex was crystallized by the hanging drop vapor diffusion method at room temperature. Crystals grew in two to three days to a size of 0.3 mm × 0.1 mm × 0.1 mm. They belong to the orthorhombic space group $P2_12_12_1$ (Table 2). At a synchrotron X-ray source, these crystals diffract to quite high resolution (often beyond 1.5 Å).

The structure of HAV 3C–1a complex was determined by the method of molecular replacement using a previous HAV 3C structure (PDB code 1HAV, molecule A) as search probe. After carefully rebuilding and refining the protein model, an

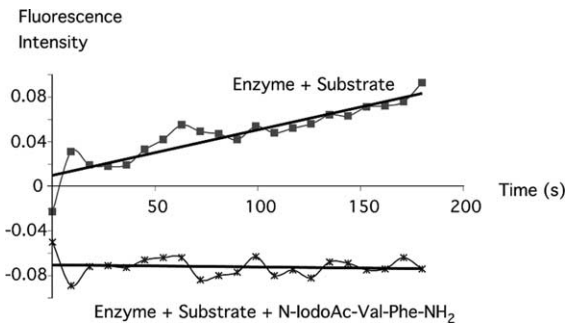



**Figure 4.** Assay of activity of redissolved crystals of HAV 3C protease C24S modified at the ε2-nitrogen of His102 by alkylation with *N*-Cbz-L-serine β-lactone (1a). The assay was done using fluorogenic peptide substrate Dabcyl-GLRTQSFS-Edans without inhibitor and also with addition of an excess of the irreversible inactivator, *N*-iodoacetyl-Val-Phe-amide.

**Table 2.** Crystallographic statistics of data collection and refinement, the cell constants of the two HAV 3C crystal structures previously solved are also listed

| Crystal | 2CXV | 2A4O | 1QA7 | 1HAV |
|---|---|---|---|---|
| 3C variant | C24S | C24S | C24S F82A | C24S |
| 3C-inhibitor complex | 3C–1a | 3C–1a–IVF | 3C–IVF | none |
| Space group | $P2_12_12_1$ | $P2_12_12_1$ | $P2_1$ | $P1$ |
| $a$ (Å) | 44.25 | 43.91 | 50.56 | 53.60 |
| $b$ (Å) | 56.02 | 56.07 | 78.36 | 53.55 |
| $c$ (Å) | 80.65 | 81.29 | 105.29 | 53.20 |
| α (deg.) | 90 | 90 | 90.00 | 99.08 |
| β (deg.) | 90 | 90 | 97.50 | 129.00 |
| γ (deg.) | 90 | 90 | 90.00 | 103.31 |
| No. molecules/asymmetric unit | 1 | 1 | 4 | 2 |
| A. *Data collection* | SSRL Beamline 9.2 | ALS Beamline 8.3.1 | | |
| Resolution range (Å) | 1.4–20.0 (1.40–1.44) | 1.55–46.19(1.55–1.63) | | |
| Wavelength (Å) | 1.1000 | 1.11589 | | |
| Observations | 163,242 | 118,281 | | |
| Unique reflections | 39,711 (3279) | 29,065 (4088) | | |
| $I/\sigma(I)$ | 13.6 (5.8) | 7.3 (1.6) | | |
| Data completeness (%) | 97.7 (98.0) | 97.6 (95.8) | | |
| $R_{merge}$ | 0.037 (0.211) | 0.052 (0.345) | | |
| B. *Refinement* | | | | |
| No. reflections used | 36,721 (2601) | 27,436 (1907) | | |
| Resolution range (Å) | 1.4–18.97 (1.4–1.44) | 1.55–10.0 (1.55–1.59) | | |
| Free set size (%) | 6.6 | 4.9 | | |
| No. atoms | 1982 | 1958 | | |
| No. water molecules | 332 | 258 | | |
| $R_{factor}$ (%) | 16.6 (18.4) | 17.8 (28.5) | | |
| $R_{free}$ (%) | 18.7 (23.3) | 19.8 (31.9) | | |
| Mean $B$ value | 16.38 | 19.61 | | |
| C. *r.m.s.d. from ideal geometry* | | | | |
| Bond length (Å) | 0.006 | 0.01 | | |
| Bond angle (deg.) | 0.985 | 1.19 | | |
| Chirality | 0.064 | 0.081 | | |
| Ramachandran plot outliers | Asp36, Asp84 | Asp36, Asp84 | Asp36 | Asp36, Asp149 |

$R_{merge} = \Sigma_h\Sigma_i|I_{hj} - I_h|/\Sigma_h\Sigma_i I_{hj}$, where $I_h$ is the weighted mean intensity of the symmetry-related reflections $I_{hj}$. $R_{working} = \Sigma_h||F_c| - |F_o||/\Sigma_h F_o$, where $F_o$ and $F_c$ represent the observed and calculated structural factors, respectively. $R_{free}$ is $R_{working}$ calculated with the reference set. Parentheses indicate values for the highest resolution shell.

examination of the $2|F_o|-|F_c|$ and $|F_o|-|F_c|$ difference maps presented a rather surprising find: the active site of the protease is free of extra densities other than those of water molecules, whereas additional electron density corresponding to *N*-Cbz-L-serine-β-lactone (1a) is unambiguously associated with His102, a surface residue on the opposite side of the enzyme from its active center (Figure 5(c)). The extra electron density is continuous from the $2|F_o|-|F_c|$ density contour of the $N^{\varepsilon 2}$ atom of His102, indicating that a covalent modification of the HAV 3C protease at this residue has occurred. Fitting a ring-opened product of the inhibitor as result of alkylation (Figure 2, path a) into the density next to His102 led to a further reduction in both the $R_{working}$ and $R_{free}$ factors. The final model consists of 212 residues of HAV 3C with the bound inhibitor (1a) and 247 water molecules. Overall $R_{working}$ and $R_{free}$ factors are 17.9% and 19.8%, respectively. The two Ramachandran plot outliers, residues Asp84 (phi, 68.8°; psi, −73.3°) and Asp36 (phi, 48.4°; psi, −127.1°), are in well-defined electron density (Figure 5(a) and (b)). As Asp36 participates in the formation of the RNA binding site of HAV 3C, its unusual conformation is perhaps tolerated for the function-related interactions involving its side-chain atoms.[7] The unusual stereochemistry of Asp84 and its involvement in the assembly of a functional catalytic triad are further discussed in the following sections. Other crystallographic statistics are listed in Table 2.

## Overall structure

The overall structure of HAV 3C–1a complex adopts the canonical chymotrypsin fold widely observed in proteases from several viral families.[5,9,18,29–34] Two structurally related β-barrels, each consisting of seven strands, form the basic scaffold of the protease. The residues connecting strands B2 and C2 in the C-terminal β-barrel form a β-hairpin inserted into the narrow groove between the N and C-terminal barrels with its "tip" (the β-turn between the two antiparallel strands in the β-hairpin) situated adjacent to the catalytic residues (Figure 6). The β-hairpin in HAV 3C is the longest among known structures of picornaviral proteases[8] (Figure 7(b)). The residues in this substructure of HAV 3C and their counterparts in related viruses usually display a higher degree of flexibility as evidenced by relatively large values in the temperature factors published with their structures. Additionally, among various

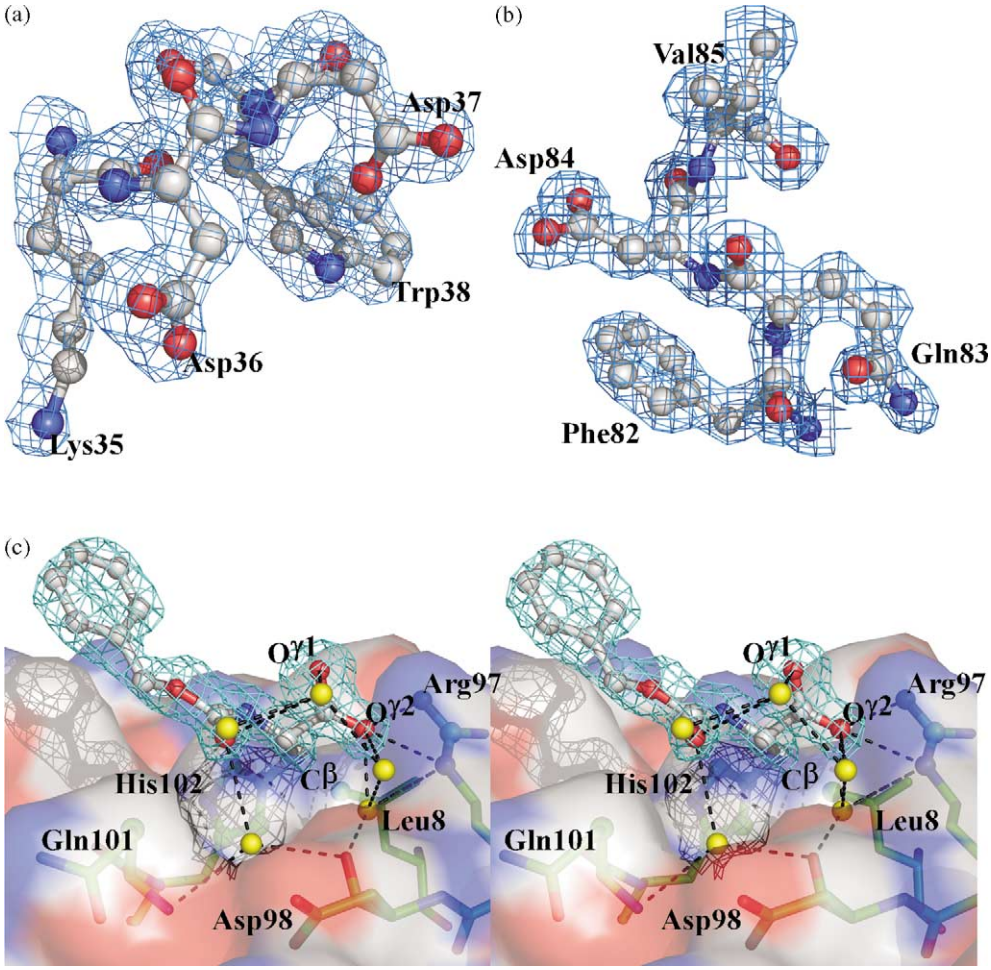



**Figure 5.** The electron densities around residues Asp36, Asp84 and the bound 1a inhibitor. (a) Residues 35–38 are shown in ball and stick and colored according to atom type (grey for carbon, red for oxygen and blue for nitrogen). Their corresponding $2|F_o|-|F_c|$ difference map is contoured at 1sigma level. (b) $2|F_o|-|F_c|$ difference map of residues 82–84 contoured at 1sigma level. (c) $|F_o|-|F_c|$ difference map (3 sigma, cyan mesh) shows the well-defined density for the 1a inhibitor on the surface (transparent, atom coloring as in (a)) of 3C protease. The hydrogen bonds involved are depicted in black broken lines. Five ordered water molecules (yellow spheres) were observed to form hydrogen bonds to 1a and surrounding residues.

crystal structures of HAV 3C, this part of the enzyme exhibits greater structural divergence that can be viewed as the result of a rotational movement of the β-hairpin hinged at its base residues (residues Glu138 and Asp158) (Figure 7(a)). The intrinsic flexibility in the β-hairpin region of HAV 3C may be explained in part by the fact that 15 out of its 21 (aa 138–158) residues are polar or charged, contributing to its relatively higher propensity to be exposed to solvent (with its tip farther away from the proteolytic active site). Many of the residues in the β-hairpin are involved in the formation of substrate specificity pockets (the S sites); thus their spatial orientations with respect to the residues in the active site may directly affect the binding of substrate or *vice versa*.

## Structural alignment with previously solved HAV and related viral 3C structures

The new crystal structure of HAV 3C (PDB code 2CXV) aligns to previously published HAV 3C structures reasonably well. The root-mean-square deviations between 2CXV and molecules A and B in the triclinic crystal form (1HAV, HAV 3C C24S variant) are 0.99 Å and 1.31 Å, respectively. The alignment between 2CXV and the four molecules in the monoclinic crystal form (1QA7, HAV 3C C24S/F82A double variant in complex with iodoacetyl-Val-Phe-amide inhibitor) falls in a similar range (Table 3). Four regions consistently exhibit large variations among different HAV 3C structures: those around residues Lys21, Glu53, and Gln83, as well as residues from His145 to Val153 in the β-hairpin substructure. These sites coincide with the regions of the highest temperature factors in these structures, indicating relatively greater intrinsic flexibility in these regions of HAV 3C. Both Lys21 and Glu53 are situated in loop regions on the surface of the HAV 3C enzyme, Lys21 being in the part connecting strands A1 to B1 whereas Glu53 is in the loop connecting strands C1 to D1. These regions are involved in intermolecular contacts between the four molecules (molecules

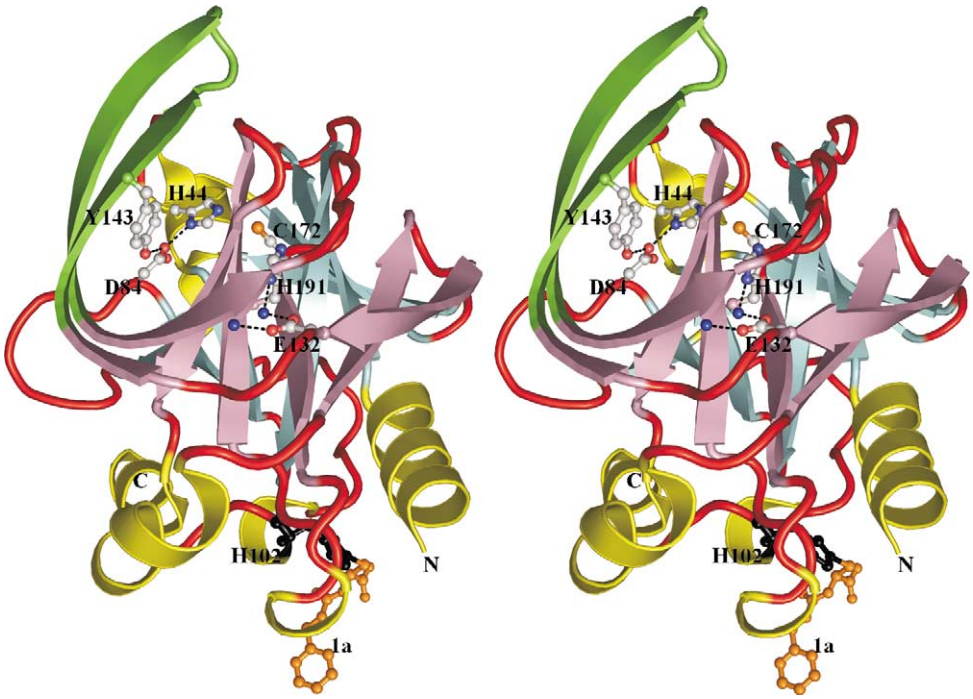



**Figure 6.** Structural overview of the HAV 3C–1a complex. A ribbon representation of the HAV 3C structure. The N and C-terminal β-barrels are colored magenta and cyan, respectively. Green depicts the β-hairpin (residues 138–158), yellow the N and C-terminal helices and red the connecting loops. His102 and the 1a adduct are shown in ball and stick with His102 colored black and 1a in orange. The catalytic triad, Cys172, His44 and Asp84 are also represented in ball and stick but colored by atom type (C atoms, grey; O, red; N, blue and S, orange). The hydrogen bonds between the carboxylate oxygen atoms of Asp84 and $N^{\delta 1}$ of His44 as well as $O^{\eta}$ of Tyr143 are shown as black broken lines. The two ordered water molecules near Glu132 are shown as spheres and colored dark blue, with their interactions with Glu132 and His191 illustrated in black broken lines.

A–B and C–D) in the asymmetric unit of the monoclinic crystal of HAV 3C (1QA7); hence, the structural divergence in these regions also reflects the differences in crystal contacts. The structural variations near Gln83 and in the β-hairpin region are similarly influenced by crystal packing in the monoclinic crystal form.[8] Interestingly, concomitant to those structural differences is the re-orientation of Asp84, the third member of the catalytic triad, with respect to the other two

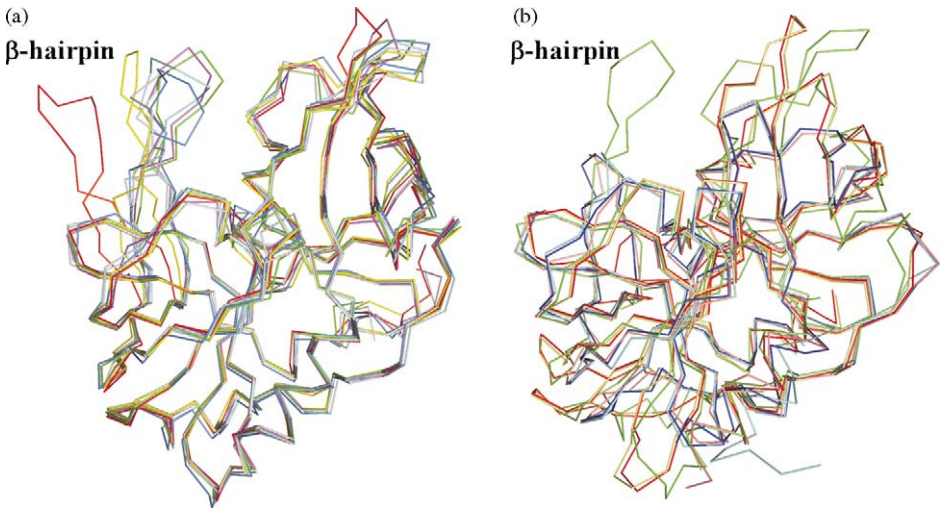



**Figure 7.** Structural alignments (a) among different HAV 3C structures and (b) among different picornaviral 3C structures. Color codes in (a): green for 2CXV; cyan and marine for molecules A and B in 1HAV, respectively; yellow, red, grey, magenta for molecules A, B, C, D in 1QA7, respectively. Color codes in (b): green for 2CXV; pink for 1CQQ (HRV2); cyan and blue for molecules A and B in 1L1N (PV), respectively; red and yellow for molecules A and B in 2BHG (FMDV), respectively.

**Table 3.** Statistics of the alignment of HAV3C crystal structures discussed in this study

| | 2A4O | HAV_A | HAV_B | 1QA7_A | 1QA7_B | 1QA7_C | 1QA7_D |
|---|---|---|---|---|---|---|---|
| 2CXV | 0.24/212[a] | 0.99/212 | 1.31/212 | 1.63/212 | 1.93/208[b] | 1.42/212 | 0.92/212 |
| 2A4O | | 1.01/212 | 1.35/212 | 1.69/212 | 2.04/208[b] | 1.50/212 | 0.92/212 |
| HAV_A | | | 1.08/214 | 1.38/213 | 1.81/213 | 1.27/213 | 0.87/216 |
| HAV_B | | | | 1.37/213 | 1.89/210[b] | 1.26/213 | 1.30/214 |
| 1QA7_A | | | | | 1.67/210[b] | 1.19/213 | 1.53/213 |
| 1QA7_B | | | | | | 1.93/209[b] | 1.73/213 |
| 1QA7_C | | | | | | | 1.46/213 |

[a] r.m.s.d. (Å) of $C^{\alpha}$ atoms/number of pairs.
[b] The positions of the C atoms of some residues in the B2/C2 loop (147–152) are too far away in the structures compared that the ALIGN program excluded them from r.m.s.d. calculation.

catalytic residues His44 and Cys172. The net result of the conformational change in Asp84 is the establishment of a functional catalytic triad in HAV 3C structure observed in this orthorhombic form for the first time.

### Interaction between HAV 3C and *N*-Cbz-L-serine-β-lactone(1a)

The contacts between HAV 3C and His102-bound 1a inhibitor involve three types of interactions: a covalent linkage between $N^{\varepsilon 2}$ of His102 and the β-carbon ($C^{\beta}$) of the inhibitor, multiple direct and water-bridged hydrogen-bonds, and abundant van der Waals interactions (Figure 8). The refined $C^{\beta}$–$N^{\varepsilon 2}$ bond shows good chemical geometry: $C^{\beta}$ is ~1.4 Å from $N^{\varepsilon 2}$ and is co-planar with the imidazole ring of His102. The torsion angle for $C^{\beta}$ of the inhibitor with the $N^{\varepsilon 2}$, $C^{\varepsilon 1}$, and $N^{\delta 1}$ atoms of His102 is 179.7° and that for $C^{\beta}$ with His102's $N^{\varepsilon 2}$, $C^{\delta 2}$, and $C^{\gamma}$ atoms is 179.1°. The bond angles $C^{\beta}$–$N^{\varepsilon 2}$–$C^{\varepsilon 1}$ and $C^{\beta}$–$N^{\varepsilon 2}$–$C^{\delta 2}$ are 123° and 128°, respectively. The carboxyl group of the inhibitor, resulting from the ring-opening event during alkylation (Figure 2(a)), forms several fairly strong hydrogen bonds with the residues in the vicinity: $O^{\gamma 1}$ of 1a hydrogen bonds to the main-chain N atom of Leu127 in the neighboring 3C molecule (3.0 Å), whereas $O^{\gamma 2}$ of 1a accepts two hydrogen bonds from the $N^{\varepsilon}$ and $N^{\eta 2}$ atoms of Arg97 in the inhibitor-bound 3C molecule (3.1 Å and 2.8 Å, respectively). Additionally, the N atom of 1a donates a hydrogen bond to the carbonyl oxygen of Pro127 of the neighboring molecule. The large number of van der Waals interactions involving atoms of 1a suggests that they provide the major stabilizing force for the 3C–1a complex. A number of surface residues in 3C are in appropriate van der Waals

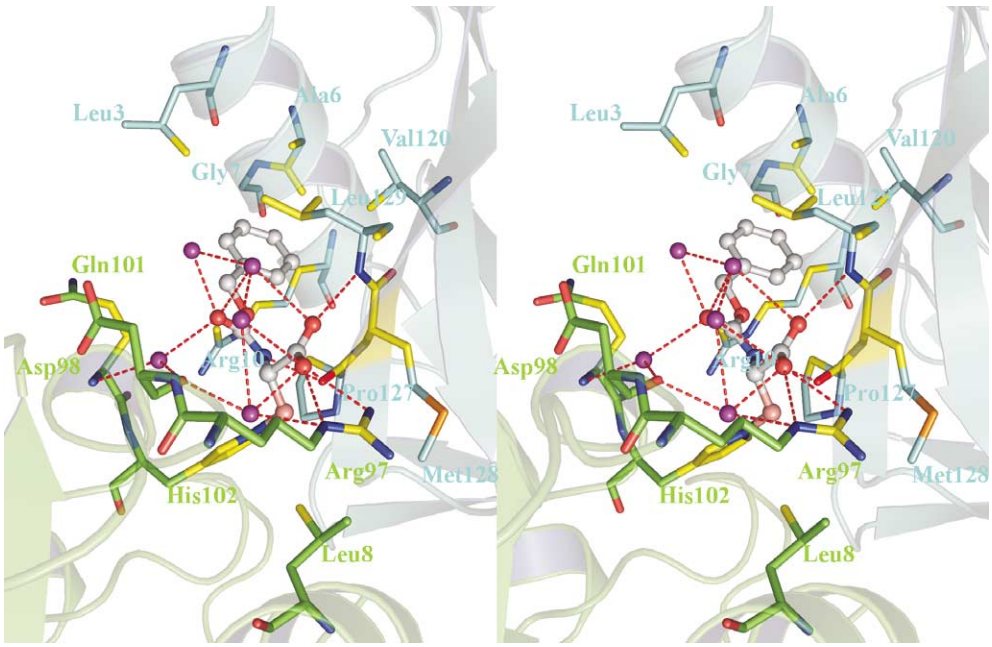



**Figure 8.** Interactions involving the 1a inhibitor at the dimer interface. The residues interacting with 1a are shown as sticks and colored by chain (molecule A is the His102 showing 3C and is colored green, molecule B is the neighboring molecule and colored cyan). The five ordered water molecules forming hydrogen bonds to 1a are represented as spheres and colored magenta. Atoms within 4 Å of the atoms of 1a are colored yellow. 1a is colored as in Figure 5. Additionally, shown are the hydrogen bonds between the inhibitor and the 1a-interacting atoms, which are colored red for oxygen and blue for nitrogen.

distance from 1a: Arg97, His102, Gln101, Leu8 from the 1a-binding 3C molecule and Leu3, Ala6, Gly7, Arg10, Pro127, Met128 and Leu129 in a neighboring molecule. In particular, the residues in the neighboring 3C molecule in the crystal form a well-defined hydrophobic pocket where the benzyl ring of the 1a inhibitor binds snugly (Figure 8).

The HAV 3C C24S variant crystallizes readily in the presence of 1a. This is in sharp contrast to the difficulty we have previously encountered while trying to crystallize the C24S variant alone. An obvious explanation is that the binding of 1a to His102 creates new protein–protein contacts that fortunately turned out to be favorable for crystal growth. Although intermolecular contacts were observed before in HAV 3C crystal structures (1HAV and 1QA7), the 3C–3C interfaces in those structures appear to be less specific and less defined than that observed in the new crystal form reported here, where two 3C–1a complexes interact in a "head-to tail" fashion. The dimer interface asymmetrically involves ten residues (Ser1, Glu4, Ile5, Leu8, Arg97, His102, Asn180, Gln181, Ser182, Gln184) from one 3C molecule and 16 residues (Leu3, Ala6, Gly7, Arg10, Asn60, Gly62, Gly63, Tyr65, Val120, Val123, Asn124, Gly125, Thr126, Pro127, Met128, Leu129) from the neighboring 3C molecule. Many of these residues have aliphatic side-chains. Burying the hydrophobic side-chains of a number of surface residues is likely to increase the entropy of the solvent, leading to a more ordered conformation of the protein, as well as lowering the overall free energy of the system. The enhancement to the stability of the 3C protein in solution is probably better for the crystallization of 3C protein. Furthermore, it is likely that ordered polymers readily develop from the head-to-tail mode of dimerization of 3C molecules, which can be conducive to crystal growth.

### Active site conformation in HAV 3C–1a complex

One of the major differences between the new crystal structure (PDB code 2CXV) and the previously published HAV 3C structures is that 2CXV provides the first observation of a properly formed catalytic triad in the active site of HAV 3C, in which the side-chain of Asp84 has undergone a significant conformational change to bring its side-chain carboxyl group into optimal hydrogen-bonding orientation and distance to the $N^{\delta 1}$ atom of His44, the general acid–base catalyst. Structural alignment shows that the positions of the side-chain atoms in Cys172 and His44 are only marginally different among structures obtained in different crystal forms (2CXV, orthorhombic; 1QA7, monoclinic; 1HAV, triclinic). In contrast, residues 82–84 exhibit large conformational differences among various HAV 3C models (Figure 9). The $\chi 1$ dihedral angles of Asp84 in 1QA7 or 1HAV are mostly either *gauche*(+) or *gauche*(−), whereas $\chi 1$ in 2CXV is in the *trans* conformation. This unexpected conformation of Asp84 in 2CXV is stabilized *via* a number of strong hydrogen bonds not present in other HAV 3C structures; those from the $O^{\eta}$ atom of Tyr143 and the $N^{\delta 1}$ atom of His44 to the $O^{\delta 2}$ atom of Asp84 seem to be the chief contributors. Additional stabilization comes from the van der Waals interactions formed between the side-chain atoms of Asp84 and those of Pro42, Ser43, His44, Phe82, Gln83, Val85, Tyr143, Leu155, Val157 and Val192 in the 2CXV structure. A neighboring residue, Gln83, shows locally the highest divergence among HAV 3C structures as evidenced by the relatively larger $C^{\alpha}$ r.m.s.d. values from structural alignments and greater variation in main-chain/side-chain dihedral angles. In 2CXV, Gln83 has a $\chi 1$ dihedral angle of $-57.3°$ thus close to the most energetically favorable *gauche*(+) conformation, which is further stabilized by a number of hydrogen bonds and van der Waals interactions. Although molecule B in the 1HAV structure has a similar value of $-53.5°$ for this $\chi 1$ torsion angle, the side-chain of Gln83 is rotated toward solvent thus forming a smaller number of interactions with other residues (Figure 9(b)). Similarly, Gln83 would have to flip by almost 180° from its position in 1HAV (molecule A) to adopt its conformation in 2CXV (Figure 9(a)). A rough comparison of the number of interactions involving residues 82–84 in various HAV 3C crystal structures could be taken to mean that the formation of a functional triad in HAV 3C is dependent on the establishment of many inter-residue interactions as a consequence of the conformational changes in residues 82–84. Moreover, it is likely that several residues participate in concert to achieve the conformational change in residues 82–84 in 2CXV, as intra-molecular interactions are usually "recycled" to avoid energetic penalties. For example, the $O^{\eta}$ atom in Tyr143 switched from accepting a hydrogen bond from the $N^{\delta 1}$ atom in His44 (molecule A in 1HAV) to donating a hydrogen bond to the $O^{\delta 2}$ atom in Asp84 (Figure 9(a)). Furthermore, the hydrogen bonds involving Val85 (N atom), Asp158 (N atom), Ser43 ($O^{\gamma}$ atom) are merely swapped between residues Gln83 and Asp84 in the 2CXV and 1HAV structures. Although it could be argued that the conformational change in Tyr143 that places its $O^{\eta}H$ group in hydrogen bonding position to the carboxyl group of Asp84 could be the cause instead of effect of the catalytic triad re-orientation, such hypothesis is at unfavorable odds with existing data. Three out of the six structures of HAV 3C each with a defective catalytic triad have the phenol ring of Tyr143 pointing in the same direction as observed in 2CXV (Figure 9(a)).

### The S1 substrate-binding pocket of various picornaviral 3C proteases

The S1 binding site in HAV 3C is a cleft that is partly exposed to solvent, a feature also shared by other picornaviral 3C structures. In HAV 3C, the S1 specificity pocket favors mostly hydrophobic interactions due to its chemical make-up; the walls of

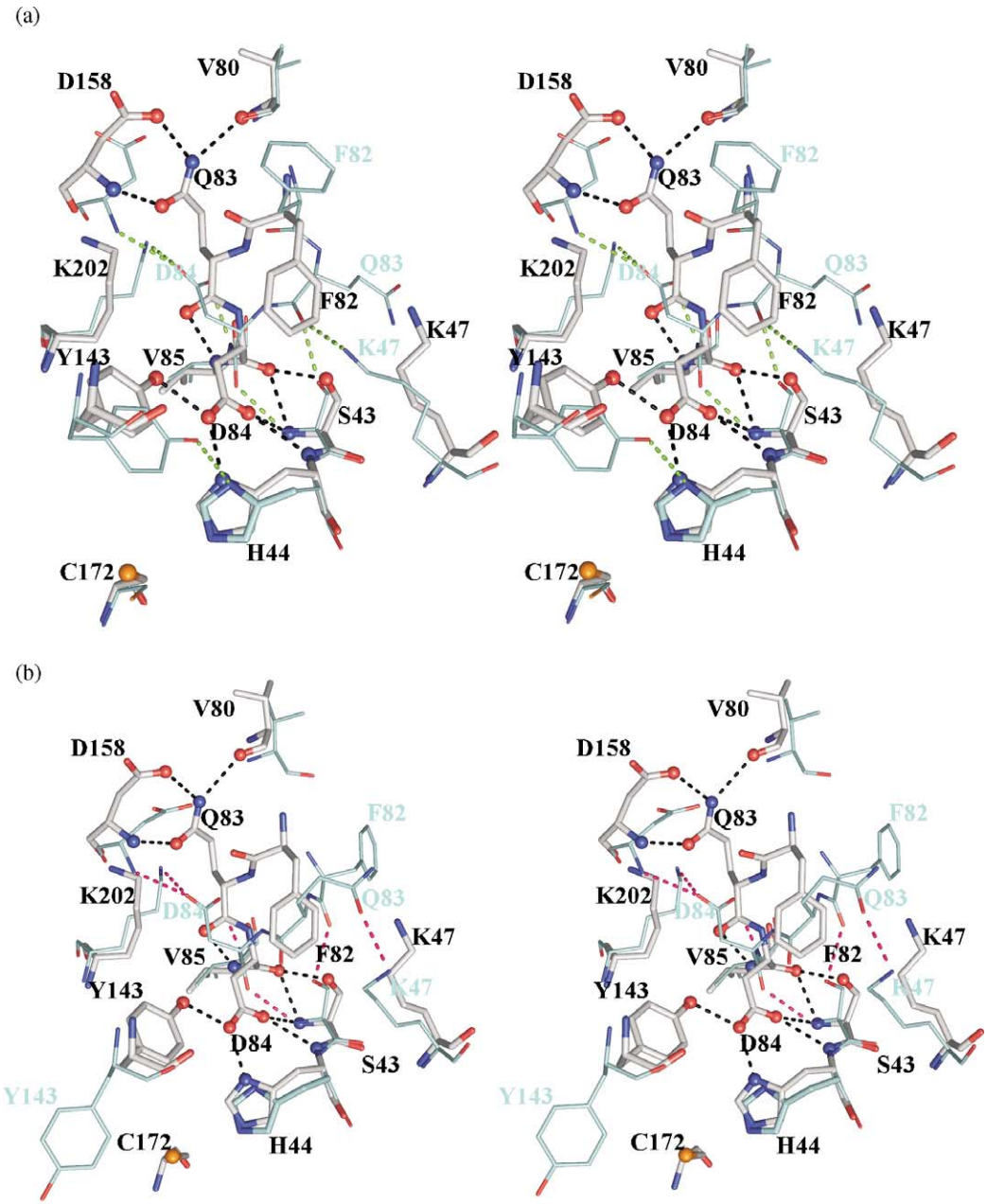



**Figure 9.** Comparison of the interactions involving residues 82–84 in the 2CXV and 1HAV structures. (a) 2CXV *versus* molecule A in 1HAV and (b) 2CXV *versus* molecule B in 1HAV. Residues in the 2CXV structure are shown in sticks with atoms forming hydrogen bonds shown in spheres. Structures are distinguished by their carbon atom colors: gray is 2CXV, cyan is 1HAV. For comparison, the corresponding residues in 1HAV are shown in lines, with the exception of Cys172 and His44, which are shown in sticks. The hydrogen bonds involving residues 82–84 are colored green and pink for molecules A and B in 1HAV, respectively. To label residues in drastically different conformations, black and cyan labels are used to represent 2CXV and 1HAV structures, respectively.

the binding groove consist of the side-chain atoms of Met171 and Leu199 and the main-chain atoms of Ala193, Gly194, Gly167, and Leu168, whereas His191, the P1 specificity provider, sits at the bottom of the S1 binding cleft.[7] Recognition of the natural substrate involves the formation of a hydrogen bond between the $N^{\varepsilon 2}$ atom of His191 and the carbonyl oxygen in the side-chain of the P1 glutamine. This arrangement restricts the imidazole ring of His191 to only a small range of possible tautomeric states, which are believed to be stabilized *via* a water-bridged interaction with a buried glutamate residue, Glu132.[7] In all previous HAV 3C structures, two ordered water molecules were observed forming hydrogen bonds to the side-chain atoms in both Glu132 and His191. This is clearly confirmed in our new crystal structure: the two buried water molecules form hydrogen bonds

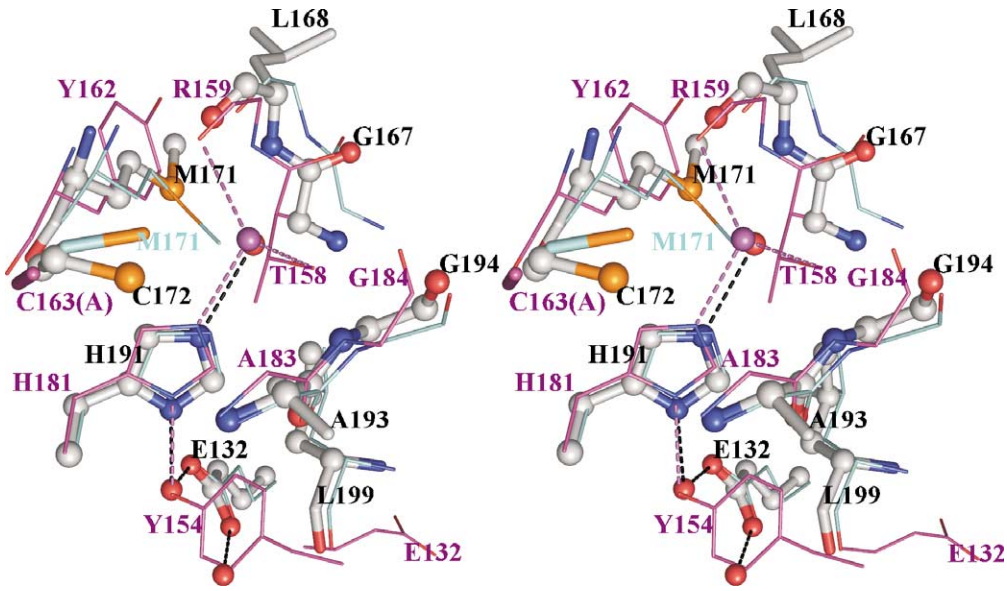



**Figure 10.** The P1 specificity pocket in the HAV 3C and FMDV 3C structures. The residues involved in forming the S1 site in HAV 3C (2CXV) are shown in ball and stick and colored similarly as in Figure 8. The two buried water molecules H-bonding to Glu132 and His191, as well as the water molecule in the S1 site occupying the proposed position of carboxamide oxygen in P1 glutamine, are shown as red spheres. Only molecule A in 1HAV is shown here in lines, with its Cys172 shown in sticks (carbon atoms colored cyan). The two ordered water molecules near Glu132 in 1HAV occupy very similar volumes in the superimposed structures of HAV 3C. The residues in the S1 site in FMDV 3C are shown analogously (carbon atoms in magenta) as those in 1HAV_A except for three differences: (1) there are no buried water molecules near His181; (2) the side-chain of Glu132 flips out toward the solvent; and (3) the side-chain oxygen in Thr158 is 3.58 Å away from the water molecule in the S1 site H-bonding to His181. The residue labels are colored according to the color of the C atoms in the corresponding structure.

to the $O^{\varepsilon 2}$ and $O^{\varepsilon 1}$ atoms of Glu132 at distances of 2.6 Å and 2.8 Å, respectively. One of the water molecules also donates a hydrogen bond to the $N^{\delta 1}$ atom of His191 (2.8 Å). As the new HAV 3C structure is of significantly higher resolution (1.4 Å *versus* 1.9 Å in 1QA7 and 2.0 Å in 1HAV), we were able to observe additional ordered water molecules in the S1 site, one of which is at a distance of 2.8 Å from the $N^{\varepsilon 2}$ atom in His191 and coplanar with the imidazole ring of His191 and the two buried water molecules interacting with Glu132 and His191 (Figure 10). We believe that the position of this water molecule resembles the binding site of the carboxamide oxygen in the P1 glutamine substrate. Strong support of this claim is that a similar water molecule was also found near His181, the equivalent residue to His191 in HAV 3C, in molecule A of the crystal structure of FMDV 3C (2BHG). The latter water molecule is 2.6 Å from the $N^{\varepsilon 2}$ atom in His181 and in the same plane as its imidazole ring. In the crystal structure of TGEV 3CL protease, two closely located water molecules (2.1 Å) were found in the S1 pocket of the enzyme, with one of them forming a hydrogen bond to the $N^{\varepsilon 2}$ atom of His162, the substrate specificity-determining residue.[32] It is interesting to note that although residue Glu132 is conserved in the 3C proteases between HAV and FMDV, the side-chain of that residue (Glu132) is not buried in FMDV 3C structure; instead, it flips 180° from the corresponding position taken by Glu132 in HAV 3C and becomes largely solvent exposed. Consequently, no buried water molecules were found in the vicinity of His181 in FMDV 3C. However, the side-chain of Tyr154, located on the β-strand topologically immediately following the β-hairpin region, interacts with the $N^{\delta 1}$ atom of His181 *via* its $O^{\eta}H$ group (2.9 Å). The $O^{\eta}$ atom of Tyr154, the imidazole ring of His181 and the ordered water molecule in the S1 site of FMDV 3C are all co-planar. Even more interestingly, although the 3C proteases in PV and HRV2 have valine instead of glutamic acid at the equivalent position of Glu132 in HAV 3C, the conservation of His191-interacting residue among them as well as with FMDV 3C is 100%! In PV and HRV2 numbering, Tyr138, the topological equivalent residue to Tyr154 in FMDV 3C, carries exactly the same function of interacting with His161 in the S1 site ($O^{\eta}$–$N^{\delta 1}$ distance of 2.5 and 2.8 Å in PV and HRV2, respectively) (Figures 10 and 11).

### Binding of iodo-acetyl-Val-Phe(amide) (IVF) inhibitor to HAV 3C–1a complex

To explore the possibility of using HAV 3C–1a crystals for the study of interactions between the protease and other active site-binding inhibitors, we soaked pre-formed HAV 3C–1a crystals in a solution containing IVF, an irreversible inhibitor previously used in the study of interactions between 3C and P′ residues.[8] IVF readily enters the vacant active site in crystallized 3C–1a complex

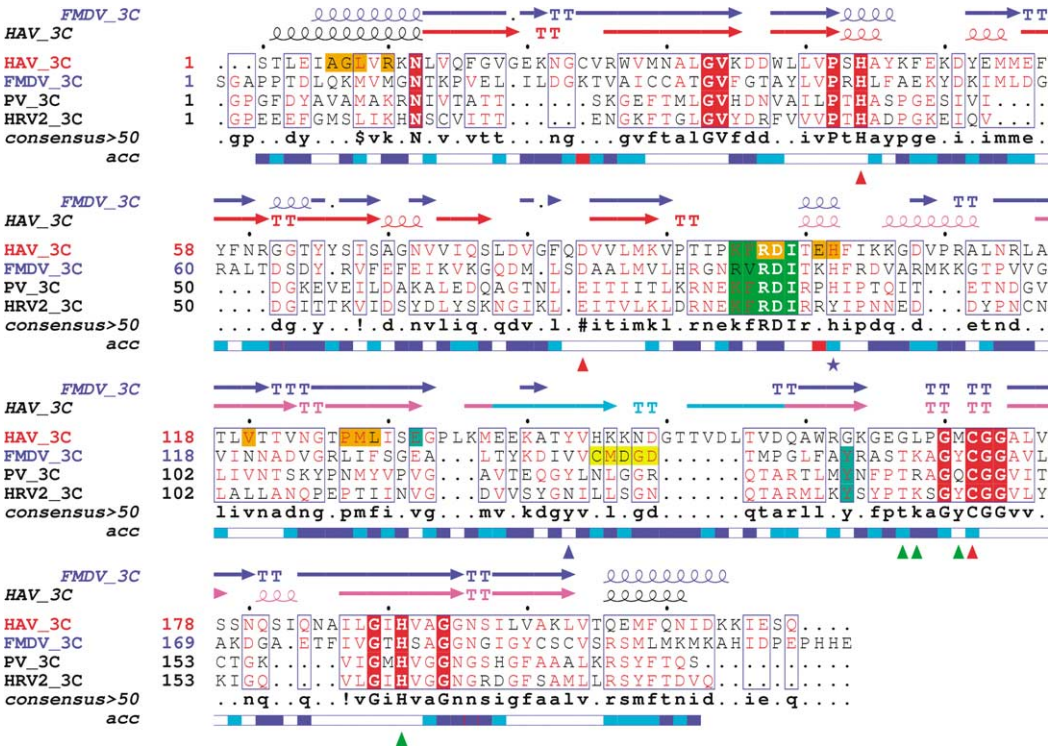



**Figure 11.** Sequence and structural alignment of known picornaviral 3C protease structures. The coloring scheme in HAV 3C secondary structure is as follows: the N and C-terminal helices are shown in black, the N-terminal β-barrel including the loops and helices connecting the β-strands of the barrel is colored red, the C-terminal β-barrel is similarly colored magenta, and the β-hairpin region (139–158) is colored cyan. In the sequence alignment, residues are colored according to level of conservation among the sequences compared, with red color indicating more conserved positions and red background indicating residues that are fully conserved; blue boxes imply contiguous regions with extensive sequence similarity; the RNA binding motifs are shaded in green background; the residues in a yellow background in FMDV 3C sequence are disordered in the crystal structure, whereas those in orange background in HAV 3C sequence are involved in the interaction with the 1a inhibitor; the tyrosine residues in FMDV (Tyr154), PV (Tyr146) and HRV2 3C (Tyr146) structures, the functional replacement of Glu132 in HAV 3C, are colored in the same marine color although they are not aligned together in the comparison of amino acid sequences. The solvent accessibility is calculated based on the new HAV 3C crystal structure and depicted in colored bars under the sequence alignment: the darker the blue color, the more accessible that residue is to the solvent, red areas indicate discrepancies between the PDB file and protein sequence used in the alignment. Under the solvent accessibility graph, the red triangles indicate the catalytic triad; the blue star indicates the position of His102, the surface residue modified by the 1a inhibitor; green triangles show the residues involved in forming the S1 site in HAV 3C including His191; blue triangle indicates the position of Tyr143, the residue participating in the re-organization of the catalytic triad in the new HAV 3C crystal form.

and the resulting triple complex 3C–1a–IVF retains the original crystal properties (Table 2). Direct refinement using protein coordinates of 2CXV as a model generated additional densities in the $2|F_o|-|F_c|$ and $|F_o|-|F_c|$ maps near His102 and Cys172 similar to those observed separately in the 2CXV and 1QA7 structures (Figure 12(a)). The final model (PDB code 2A4O) of 3C protease modified at both His102 (1a) and Cys172 (IVF) has $R_{working}$ and $R_{free}$ factors of 0.18 and 0.20, respectively. The Ramachandran plot indicates that the mainchain dihedral angles of Asp36 (phi, 51.3°; psi, −129.2°) and Asp84 (phi, 68.0°; psi, −68.7°) are in the generously allowed and disallowed regions, respectively. However, both of these residues are well-defined by unambiguous electron densities.

The overall r.m.s.d. value of the protein coordinates between 2A4O and 2CXV is 0.15 Å for 201 $C^{\alpha}$ pairs, whereas the r.m.s.d. values between 2A4O and previously published HAV 3C structures are very similar to those observed in 2CXV *versus* 1HAV and 2CXV *versus* 1QA7 comparisons (Table 3). The deviation in the coordinates of all atoms in the 1a inhibitor between 2A4O and 2CXV falls in the same range as the overall r.m.s.d., at 0.13 Å. Accordingly, the interactions involving the 1a inhibitor at the dimer interface are essentially the same as those observed in 2CXV (data not shown). Although the acetyl-Val-NH(CO) moiety of the IVF inhibitor retains its position relative to the active site residues of the protease in 2A4O and in 1QA7 structures, the benzene ring side-chain of the phenylalanine residue in IVF exists separately in two conformations (as defined by its χ1 angle) among five 3C–IVF structures (2A4O and the four molecules in 1QA7). Those Phe residues in molecules A, C and D in 1QA7 take on close to *gauche*(+) conformation at $C^{\beta}$, whereas those in molecule B and 2A4O are in

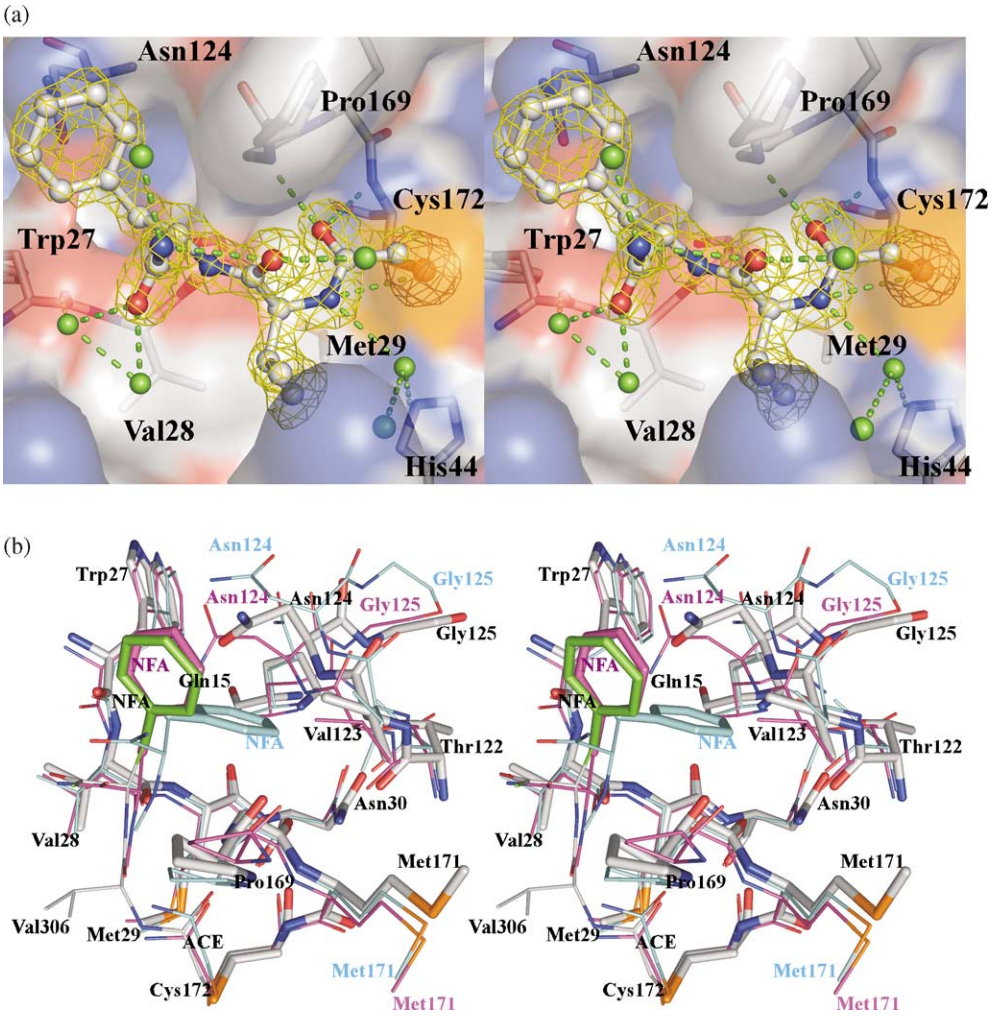



**Figure 12.** Interactions involving atoms of the IVF inhibitor in the active site of 2A4O. (a) The omit $|F_o|-|F_c|$ electron density map contoured at 3 sigma level with the final refined model of the IVF (iodoacetyl-Val-NFA) superimposed. NFA refers to phenylalanine-amide. The density of the inhibitor is continuous from the $2|F_o|-|F_c|$ density of the sulfur atom in Cys172, indicating a covalent linkage. The IVF inhibitor is represented in ball and stick and colored similarly as in Figure 8. The residues forming hydrogen bonds or van der Waals interactions with IVF inhibitor are shown in sticks. Also shown are several water molecules in green that interact with the covalently bound inhibitor *via* hydrogen bonds. (b) A structural alignment of the IVF inhibitors and S2′ sites in 2A4O and molecules A, B of the 1QA7 structure. The structures are distinguished by the coloring of carbon atoms: gray is for 2A4O, cyan for molecule A and magenta for molecule B. Residue labels (color codes same as carbon atoms) are provided where the three structures differ significantly.

a *trans* conformation. The benzene ring moiety of the IVF inhibitor in 2A4O did not bind inside the S2′ pocket in 2A4O; instead, it covers the entrance to the S2′ binding site (Figure 12(b)). Besides forming several non-specific hydrophobic interactions with the main-chain atoms of Gly170, Pro169 and Asn124, the center of the benzyl ring in IVF is ~3.4 Å from the $N^{\delta 2}$ atom of Asn124 with the $N^{\delta 2}$–H bond at an angle of ~25° from the 6-fold axis of the benzyl ring, an orientation commensurate with a fairly strong aromatic hydrogen bond observed and described by Perutz *et al.* in hemoglobin structures.[35,36] The position of the A2–B2 loop, wherein Asn124 resides, seems to determine the size of the S2′ binding pocket. In molecules A, C and D (1QA7), this loop is moved away from the opposing strands and further into solvent, leaving a cavity large enough to accommodate the phenyl group in IVF. In contrast, steric hindrance prevents the binding of the same group into the S2′ pocket in 2A4O and molecule B of 1QA7: the distances between the main-chain amide of Asn124 in 2A4O and the $C^{\varepsilon 1}$ and $C^{\varepsilon 2}$ atom in the benzene ring moiety of the IVF inhibitor in the superimposed structures of molecule A (1QA7) are only 2.5 Å and 1.7 Å, respectively. Previously, Bergmann *et al.* suggested that the S2′ pocket unique to HAV 3C is not necessarily designed to bind an aromatic residue at the P2′ position of the substrate.[8] Three lines of evidence seem to support this: (1) The lining of the walls of the binding pocket is decorated by the main-chain amides and carbonyl oxygen atoms from Val28, Met29, Asn30, Asn124, and Gly125. (2) The side-chains of Gln15 and Gln30 form

the hydrophilic "bottom" of the S2′ binding pocket. (3) In both molecule B (1QA7) and 2A4O, three ordered water molecules are observed binding inside the S2′ pocket, the positions of which are mutually exclusive with that of the benzene ring moiety of IVF in molecules A, C and D. One of these water molecules forms a fairly strong hydrogen bond to the carbonyl oxygen of Gly170 (2.8 Å), an interaction also observed in molecule B in 1QA7 (2.9 Å) as well as in 2CXV (2.7 Å).

## Discussion

### Functional catalytic triad in HAV 3C

Our observation that the binding of 1a to a surface residue, His102 of HAV 3C, stabilizes the enzyme in its active catalytic triad formation is analogous to the case of HRV2 3C in complex with AG7088 inhibitor.[9] In that study, the rigidity of the lactam ring at the P1 position of the inhibitor preferentially selects the HRV2 $3C^{Pro}$ species in solution that have the proper arrangement of the oxyanion hole, whereas the crystal structures of HRV2 3C with more than a dozen P1 Gln-containing inhibitors all displayed a defective oxyanion hole with the main-chain NH atoms in residues Val144 and His145 pointing away from the active site into solvent. Whether the binding of 1a to His102 in HAV 3C indirectly causes the formation of the functional catalytic triad in the active site or simply stabilizes existing 3C species in which a catalytic triad is already present in the active conformation remains an open question.

In PV and HRV, the equivalent residues of Gln83 are Leu70, whose side-chains occupy the same volume as that of Gln83 in 2CXV. This is expected because the side-chain of a leucine residue, being more hydrophobic than that of a glutamine, favors a more buried position. Consequently, the third members of the catalytic triads in PV and HRV2 3C proteases, Glu71, constitutively position themselves in such a way that their carboxyl oxygen atoms are in hydrogen-bonding distance from the $N^{\delta 1}$ of His40, the counterpart of His44 in HAV 3C. In contrast, the catalytic triads (Cys172, His44 and Asp84) in HAV 3C structures predominantly showed a defective assembly except for the new 3C–1a complex reported here. Perhaps this innate structural inactivity of the HAV enzyme by itself explains why HAV 3C is a relatively limited protease compared to other picornaviral 3C proteases in that it was not found to cleave many host cellular proteins, nor has it been unanimously confirmed that HAV 3C processes all the cleavage sites in the non-structural portion of the viral polyprotein, a task that depends almost exclusively on 3C or 3CD in entero- and rhinoviruses. In poliovirus, 3CD is the main viral protease and the 3D moiety of the 3CD proteases modulates the proteolytic activity of the 3C moiety.[37] Although HAV 3CD is a minor species compared to 3C during infection, it is possible that the protease function of HAV 3C is modulated *via* its interaction with other viral or host protein(s). A recent study suggests that the peptidyl substrate and viral RNA bind to HAV 3C co-operatively.[38] Perhaps the dimerization of HAV 3C resulting from the covalent modification at His102 by 1a causes structural changes in the protease in a fashion analogous to the effects of RNA or protein binding to 3C.

Since the major product of the first alkylation by 1a, the 3C protease modified at the $S^{\gamma}$ atom of Cys172, was never observed in the new $P2_12_12_1$ crystals, we came to the conclusion that this complex is so structurally distinct from the His102–1a 3C species that they are incompatible to be included in the same crystal. Neither should 3C (Cys172-1a) be able to form crystals similar to the 3C (His102–1a) crystals on its own, as the former species lacks the required 1a-mediated protein–protein contacts observed in the 3C (His102–1a) crystal structure. Although HAV $3C^{Pro}$ can be doubly modified (at Cys172 and His102) in solution by 1a inhibitor, such structures were not observed in our crystal structures despite a huge effort to soak pre-grown HAV 3C–1a crystals in fresh solutions of various β-lactone-derived inhibitors. It is possible that the structural re-organization in the active site resulting from the modification of His102 by 1a, however subtle, heightens the specificity requirement of the enzyme sufficiently such that 1a, failing to provide sufficient binding energy benefit, is kept out of the proteolytic active site by virtue of structural discrimination. Due to the high intrinsic flexibility of the β-hairpin substructure that forms part of the substrate specificity pocket in HAV 3C, this mechanism of protecting the protease against inactivation at the active site likely only works in an aggregated state of the protein such as that in a crystal. Protein–protein contacts leave the enzyme much less freedom to drastically readjust its local structure in such a way that the active site can again bind to molecules like 1a. This is probably why only solubilized 3C–1a complexes were susceptible to further modification at Cys172 by 1a, whereas soaking 3C–1a crystals in fresh solution of 1a for prolonged period of time did not lead to doubly reacted 3C proteins as confirmed by X-ray crystallography. The crystal structure of 3C (His102–1a/Cys172–IVF) reported here supports the above hypothesis in several aspects. (1) The entry and subsequent binding of a more specific peptidyl substrate in the active site is not affected in 3C–1a complex. (2) The thiol group of Cys172 is not compromised for its catalytic function as it readily forms a covalent linkage with IVF as a result of nucleophilic attack at the electron-deficient C atom in that inhibitor. (3) Asp84 and its neighboring residues Gln83 and Phe82 seem to exist in a low energy conformation in 2CXV, as they engage in more interactions with surrounding residues than were observed in 1HAV, i.e. the "ground state" of an active catalytic triad is also more stable. Such a property of the active site could be taken as a requirement of higher activation energy for

catalysis, which is provided upon specific substrate binding.

### Subtle but meaningful side-chain movement in the P1 substrate-binding pocket

The S1 site in picornaviral 3C structures is a quite conserved structure, partly because it is contiguous with the active sites of these enzymes. The active sites superimpose with remarkable fidelity among various viral 3C or 3C-like protease structures. Consequently, we did not expect to observe significant structural differences in the residues forming or adjacent to the S1 site, which is confirmed by their lower than average r.m.s.d. values in structural comparisons. However, there is one consistent difference: the new structure shows without ambiguity that the $C^{\varepsilon}$–$S^{\delta}$ bond of Met171 in the new $P2_12_12_1$ crystal form flipped 180° from its position present in all other HAV 3C crystal structures. While there are only marginal variations in the $\phi$, $\psi$, $\chi1$ and $\chi2$ angles in Met171 from structure to structure, the $\chi3$ angle swings from an average of $\sim -45°$ in 1QA7 or 1HAV to 178.6° in 2CXV. We believe this structural adjustment is significant in that it pre-emptively readies the S1 site for substrate binding. As mentioned above, the S1 site in HAV 3C is neither a deep nor a secluded pocket. Structural comparison of HAV 3C with HRV2 3C bound with the AG7088 inhibitor indicates that the correct recognition of the right P1 residue relies heavily on the formation of the hydrogen bond to be established between the $N^{\varepsilon 2}$ atom of His191 and the carbonyl oxygen in the side-chain of P1 glutamine. Molecular events must take place prior to or during the binding of substrate, so that the energy benefit resulting from specific binding is not negated by inter-atomic repulsion in the S1 site. The position of the imidazole ring in His191 seems rather rigid, as it is highly superimposable among the various picornaviral 3C structures. This subsequently defines the positions of the side-chain atoms in P1 glutamine for optimal interaction. It then follows that the positions of the $C^{\varepsilon}$ atoms of Met171 residues in previous HAV 3C structures are likely to be energetically unfavorable for substrate binding, as their $C^{\varepsilon}$ atoms may be too close to the side-chain carbonyl oxygen of the incoming P1 residue. Assuming the position of the water molecule H-bonding to His191 is close to that of a bound P1 side-chain carboxamide oxygen, the $C^{\varepsilon}$ atoms of the six structures in 1HAV and 1QA7 are at a distance of 2.9, 3.3, 3.1, 2.9, 2.4, 2.9 Å from this water molecule in superimposed structures.

### Future directions

It would initially seem counterintuitive to use 1a or its derivatives for inhibiting Hepatitis A or a related virus due to the "side effect" from the His–1a modification that stabilizes the enzyme in a conformation with a properly formed catalytic triad. However, several schools of thoughts might elevate the attractiveness of 1a as a lead for an antiviral drug. First, His102 is located in a small $3_{10}$ helix in the only region connecting the two β-barrels of the bipartite HAV 3C structure. This region of ~10 residues (amino acid residues 96–105) contains the RNA binding motif ($K^{95}F^{96}R^{97}D^{98}I^{99}$ in HAV numbering) that is highly conserved in many picornaviral 3C enzymes (Figure 11 and Table 4). The binding of 1a to His102 could potentially lead to the perturbation of the RNA binding activity of these 3C viral proteases, a process that has been shown to be critical for picornaviral RNA replication.[39–45] In particular, the side-chain of Arg97, a surface residue in HAV 3C presumably involved in interacting with RNA, is in well-defined density in the 3C–1a crystal structure due to the interactions between its side-chain atoms and those of the 1a inhibitor (Figure 5(c)). Moreover, a structure-assisted sequence alignment shows that most of the residues involved in the interaction between HAV 3C with 1a are also functionally conserved among 3C proteases in PV, HRV, and FMDV (Figure 11). Second, mutations in residues Thr154 and Lys156 of PV 3C were found to suppress the debilitating changes made into the 5′ non-translated region of the viral RNA genome.[46] Neither of their topologically equivalent residues in HAV 3C, Ser179 and Gln181, is far from the dimer interface in HAV 3C–1a crystals. Third, the stable arrangement of the functional catalytic triad in HAV 3C upon binding of 1a should facilitate the subsequent inhibition of the proteolytic activity by transition state structure-based inhibitors as in the case of HRV2 3C inhibition by AG7088. HAV 3C–1a crystals possess many nice qualities for crystallographic study (relatively strong-diffracting, easily grown, ready for further modification, high symmetry space group with one molecule per asymmetric unit), all of which certainly warrant further exploitation in the study of the interactions between HAV 3C and other active site ligands. That will be the immediate topic of forthcoming work.

**Table 4.** A list of the residues involved in the interactions between HAV 3C and the 1a inhibitor and their sequence conservation among several picornaviruses

| HAV | FMDV | PV | HRV2 | Interaction with 1a |
|---|---|---|---|---|
| His102 | His102 | His89 | Tyr89 | Covalent linkage, vdW |
| Ala6 | Gln9 | Val8 | Met8 | vdW |
| Gly7 | Lys10 | Ala9 | Ser9 | vdW (α-carbon) |
| Leu8 | Met11 | Met10 | Leu10 | vdW |
| Arg10 | Met13 | Lys12 | Lys12 | vdW (side-chain) |
| Arg97 | Arg97 | Arg84 | Arg84 | H-bond, vdW |
| Asp98 | Asp98 | Asp85 | Asp85 | H-bond |
| Glu101 | Lys101 | Pro88 | Arg88 | vdW (α-carbon) |
| Val120 | Val118 | Ile103 | Ala103 | vdW |
| Pro127 | Leu127 | Asn111 | Pro111 | vdW |
| Met128 | Ile128 | Met112 | Thr112 | vdW |
| Leu129 | Phe129 | Tyr113 | Ile113 | vdW |

vdW, van der Waals.

## Materials and Methods

### Production and assay of HAV 3C protease C24S, synthesis and testing of β-lactones and model compounds

Experimental procedures for enzyme production and assay, as well as the syntheses of compounds 1–5 (Table 1) have been published[16] and were followed in the present study. Enzyme inhibition and HMQC NMR experiments were also carried out according to the procedures described in the same report. Model compound 6 was prepared using the literature method.[22]

### Preparation and crystallization of the complex of HAV 3C with *N*-Cbz-L-Ser-β-lactone (1a)

HAV 3C protease variant C24S (100 μl of 0.12 μM solution in 20 mM potassium phosphate buffer (pH 5.4), 30 mM NaCl) was mixed with 3 μl of 100 μM *N*-Cbz-L-Ser-β-lactone in dimethyl sulfoxide (DMSO) to a final concentration of 3 mM and incubated for 1 h at room temperature. The mixture was washed in a concentrator unit with 1 mM 1a in water until the NaCl concentration was reduced to below 5 mM and protein concentration increased to 10–12 mg/ml. During the crystallization setup (hanging drops), small aliquots (2–3 μl) of the protein solution containing 2.5% (w/v) polyethylene glycol (PEG) 8000, 1.5% (v/v) glycerol and 10 mM Tris–HCl (pH 7.5) were equilibrated against 1 ml of the reservoir solution (5% PEG 8,000, 5% glycerol, 20 mM Tris–HCl (pH 7.5)). Before data collection, the crystals were rinsed with cryosolvent (30% glycerol, 5% PEG 8,000, 20 mM Tris–HCl (pH 7.5)) and flash-cooled in liquid nitrogen. To make the 3C–1a–IVF triple complex, pre-grown 3C–1a crystals were soaked in solution containing 5 mM IVF for at least 1 h before they were cryoprotected and flash-cooled for data collection.

### Assay of crystals for protease activity

A single crystal of the adduct of 1a with His102 of HAV 3C protease C24S was crushed and resuspended in assay buffer and analyzed for enzymatic activity using the fluorogenic substrate Dabcyl-GLRTQSFS-Edans under standard conditions.[16] The modified enzyme was able to hydrolyze the substrate at the glutamine residue, thereby leading to an increase in fluorescence (Figure 4). No increase in fluorescence was observed in the absence of the protein. Addition of an excess of the irreversible inhibitor *N*-iodoacetyl-Val-Phe-amide completely abolished the protease activity.

### Data collection and processing

X-ray diffraction data for the HAV 3C–1a complex and 3C–1a–IVF complex were collected at beamline 9.2 of the SPEAR2 ring at the Stanford Synchrotron Radiation Laboratory (SSRL) and Beamline 8.3.1 of the Advanced Light Source (ALS) at Berkeley Lawrence National Laboratory, respectively. The programs Denzo, Scalepack and the CCP4 suite were used to process and scale the datasets[47,48] (Table 2).

### Structure determination and refinement

The structure of the 3C–1a complex was solved by molecular replacement using program AMoRe[49] and molecule A from the previously solved structure of HAV 3C[7] (PDB ID code 1HAV) as a search probe. Rebuilding and refinement of the model utilized as much as possible the automated procedures implemented in program Arp/Warp.[50] The interactive molecular graphics programs O[51] and xFit[52] were used for visualization, fine-tuned model building and adjustment. Program Refmac5 was used in the refinement of the structures.[53] The structure of the 3C–1a–IVF complex was obtained by direct refinement using the 3C–1a coordinates as an initial model.

### Structure analysis and generation of Figures

The quality of the structures was assessed with the help of program PROCHECK.[54] Sequence and structural alignments were carried out using programs Multalin[55] and ALIGN,[56] respectively. Intra- and intermolecular contacts in various crystal structures were analyzed using the CCP4 suite of program CONTACT.[48] Protein sequences were downloaded from National Center for Biotechnology Information genome database†. Figures were prepared using ESPript[57] and Pymol‡.

### Protein Data Bank accession codes

The coordinates and associated structure factors have been deposited into the Protein Data Bank (PDB code 2CXV and 2A4O for HAV 3C–1a structure and HAV 3C–1a–IVF structure, respectively).

## Acknowledgements

We thank Bruce A. Malcolm (Schering Plough), David S. Wishart, and Scott Watson (Faculty of Pharmacy, University of Alberta) for assistance with enzyme preparations; and Carly Huitema and Lindsay D. Eltis (Departments of Microbiology and Biochemistry, University of British Columbia), Tara Sprules and Hanna I. Pettersson (Department of Chemistry, University of Alberta) for helpful discussions. X-ray diffraction data were collected at beamline 8.3.1 of the Advanced Light Source (ALS) at Lawrence Berkeley Lab, under an agreement with the Alberta Synchroton Institute (ASI). The ALS is operated by the Department of Energy and supported by the National Institute of Health. Beamline 8.3.1 was funded by the National Science Foundation, the University of California and Henry Wheeler. The ASI synchrotron access program is supported by grants from the Alberta Science and Research Authority (ASRA) and the Alberta Heritage Foundation for Medical Research (AHFMR). These investigations were supported by the Natural Sciences and Engineering Research Council of Canada, the Alberta Heritage Foundation for Medical Research and the CIHR Group Grant (to M.N.G.J.). J.Y. is a recipient of the Izaak Walton Killam postdoctoral fellowship award at University of Alberta.

† http://www.ncbi.nlm.nih.gov/entrez/
‡ http://www.pymol.org

## Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2005.09.074

## References


1. Raccaniello, V. R. (2001). Picornaviridae: the viruses and their replication. In *Fields Virology* (Knipe, D. M., Howley, P. M., Griffin, D. E., Martin, M. A, Lamb, R. A., Roizman, B. & Straus, S. E., eds) 4th edit., pp. 685–722, Lippincott Williams and Wilkins, Philadelphia, PA.
2. Pebody, R. G., Leino, T., Ruutu, P., Kinnunen, L., Davidkin, I., Nohynek, H. & Leinikki, P. (1998). Foodborne outbreaks of hepatitis A in a low endemic country: an emerging problem? *Epidemiol. Infect.* **120**, 55–59.
3. Schultheiss, T., Sommergruber, W., Kusov, Y. & Gauss-Muller, V. (1995). Cleavage specificity of purified recombinant hepatitis A virus 3C proteinase on natural substrates. *J. Virol.* **69**, 1727–1733.
4. Bergmann, E. M. & James, M. N. G. (2000). The 3C proteinases of picornaviruses and other positive-sense, single-stranded RNA viruses. In *Handbook of Experimental Pharmacology* (Von der Helm, K., Korant, B. & Cheronis, J. C., eds), vol. 140, pp. 117–143, Springer, Hidelberg.
5. Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R. & Hilgenfeld, R. (2003). Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science*, **300**, 1763–1767.
6. Allaire, M. & James, M. (1994). Deduction of the 3C proteinases' fold. *Nature Struct. Biol.* **1**, 505–506.
7. Bergmann, E. M., Mosimann, S. C., Chernaia, M. M., Malcolm, B. A. & James, M. N. (1997). The refined crystal structure of the 3C gene product from hepatitis A virus: specific proteinase activity and RNA recognition. *J. Virol.* **71**, 2436–2448.
8. Bergmann, E. M., Cherney, M. M., McKendrick, J., Frormann, S., Luo, C., Malcolm, B. A., Vederas, J. C. & James, M. N. (1999). Crystal structure of an inhibitor complex of the 3C proteinase from hepatitis A virus (HAV) and implications for the polyprotein processing in HAV. *Virology*, **265**, 153–163.
9. Dragovich, P. S., Webber, S. E., Babine, R. E., Fuhrman, S. A., Patick, A. K., Matthews, D. A. *et al.* (1998). Structure-based design, synthesis, and biological evaluation of irreversible human rhinovirus 3C protease inhibitors. 2. Peptide structure-activity studies. *J. Med. Chem.* **41**, 2819–2834.
10. Matthews, D. A., Smith, W. W., Ferre, R. A., Condon, B., Budahazi, G., Sisson, W. *et al.* (1994). Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein. *Cell*, **77**, 761–771.
11. Mosimann, S. C., Cherney, M. M., Sia, S., Plotch, S. & James, M. N. (1997). Refined X-ray crystallographic structure of the poliovirus 3C gene product. *J. Mol. Biol.* **273**, 1032–1047.
12. Birtley, J. R., Knox, S. R., Jaulent, A. M., Brick, P., Leatherbarrow, R. J. & Curry, S. (2005). Crystal structure of foot-and-mouth disease virus 3C protease. New insights into catalytic mechanism and cleavage specificity. *J. Biol. Chem.* **280**, 11520–11527.
13. Malcolm, B. A. (1995). The picornaviral 3C proteinases: cysteine nucleophiles in serine proteinase folds. *Protein Sci.* **4**, 1439–1445.
14. Ramtohul, Y. K., James, M. N. & Vederas, J. C. (2002). Synthesis and evaluation of keto-glutamine analogues as inhibitors of hepatitis A virus 3C proteinase. *J. Org. Chem.* **67**, 3169–3178.
15. Ramtohul, Y. K., Martin, N. I., Silkin, L., James, M. N. G. & Vederas, J. C. (2002). Synthesis of pseudoxazolones and their inhibition of the 3C cysteine proteinases from hepatitis A virus and human rhinovirus-14. *J. Chem. Soc. Perkin Trans.* **1**, 1351–1359.
16. Lall, M. S., Ramtohul, Y. K., James, M. N. & Vederas, J. C. (2002). Serine and threonine beta-lactones: a new class of hepatitis A virus 3C cysteine proteinase inhibitors. *J. Org. Chem.* **67**, 1536–1547.
17. Lall, M. S., Karvellas, C. & Vederas, J. C. (1999). beta-lactones as a new class of cysteine proteinase inhibitors: inhibition of hepatitis A virus 3C proteinase by N-Cbz-serine beta-lactone. *Org. Letters*, **1**, 803–806.
18. Allaire, M., Chernaia, M. M., Malcolm, B. A. & James, M. N. (1994). Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases. *Nature*, **369**, 72–76.
19. Arnold, L. D., Kalantar, T. H. & Vederas, J. C. (1985). Conversion of serine to stereochemically pure β-substituted α-amino acids *via* β-lactones. *J. Am. Chem. Soc.* **117**, 7105–7109.
20. Arnold, L. D., May, R. G. & Vederas, J. C. (1988). Synthesis of optically pure α-amino acids *via* salts of α-amino-β-propiolactone. *J. Am. Chem. Soc.* **110**, 2237–2241.
21. Ramer, S. E., Moore, R. N. & Vederas, J. C. (1986). Mechanism of formation of serine β-lactones by mitsunobu cyclization: synthesis and use of L-serine stereospecifically labelled with deuterium at C-3. *Can. J. Chem.* **64**, 706–713.
22. Ratemi, E. S. & Vederas, J. C. (1994). Reaction of trimethylsilylamines with N-Cbz-L-serine-β-lactone: a convenient route to optically pure a-amino-L-alanine derivatives. *Tetrahedron Letters*, **35**, 7605–7608.
23. Crane, S. N. & Corey, E. J. (2001). A novel enantioselective synthetic route to omuralide analogues with the potential for species selectivity in proteasome inhibition. *Org. Letters*, **3**, 1395–1397.
24. Groll, M., Ditzel, L., Lowe, J., Stock, D., Bochtler, M., Bartunik, H. D. & Huber, R. (1997). Structure of 20 S proteasome from yeast at 2.4 Å resolution. *Nature*, **386**, 463–471.
25. Hadvary, P., Sidler, W., Meister, W., Vetter, W. & Wolfer, H. (1991). The lipase inhibitor tetrahydrolipstatin binds covalently to the putative active site serine of pancreatic lipase. *J. Biol. Chem.* **266**, 2021–2027.
26. Kim, D. H. & Ryoo, J. (1995). (3*R*,4*S*)-3-Benzyl-4-bromomethyloxetan-2-one, a fast acting alternate substrate inhibitor of a-chymotrypsin. *Bioorg. Med. Chem. Letters*, **5**, 1287–1296.
27. Kisselev, A. F., Songyang, Z. & Goldberg, A. L. (2000). Why does threonine, and not serine, function as the active site nucleophile in proteasomes? *J. Biol. Chem.* **275**, 14831–14837.
28. McKendrick, J. E., Frormann, S., Luo, C., Semchuck, P., Vederas, J. C. & Malcolm, B. A. (1998). Rapid mass spectrometric determination of preferred irreversible proteinase inhibitors in combinatorial libraries. *Int. J. Mass Spectrom.* **176**, 113–124.

29. Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z. *et al*. (2003). The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl Acad. Sci. USA*, **100**, 13190–13195.
30. Love, R. A., Parge, H. E., Wickersham, J. A., Hostomsky, Z., Habuka, N., Moomaw, E. W. *et al*. (1996). The crystal structure of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site. *Cell*, **87**, 331–342.
31. Barrette-Ng, I. H., Ng, K. K., Mark, B. L., Van Aken, D., Cherney, M. M., Garen, C. *et al*. (2002). Structure of arterivirus nsp4. The smallest chymotrypsin-like proteinase with an alpha/beta C-terminal extension and alternate conformations of the oxyanion hole. *J. Biol. Chem.* **277**, 39960–39966.
32. Anand, K., Palm, G. J., Mesters, J. R., Siddell, S. G., Ziebuhr, J. & Hilgenfeld, R. (2002). Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. *EMBO J.* **21**, 3213–3224.
33. Murthy, H. M., Clum, S. & Padmanabhan, R. (1999). Dengue virus NS3 serine protease. Crystal structure and insights into interaction of the active site with substrates by molecular modeling and structural analysis of mutational effects. *J. Biol. Chem.* **274**, 5573–5580.
34. Phan, J., Zdanov, A., Evdokimov, A. G., Tropea, J. E., Peters, H. K., 3rd, Kapust, R. B., Li, M. *et al*. (2002). Structural basis for the substrate specificity of tobacco etch virus protease. *J. Biol. Chem.* **277**, 50564–50572.
35. Levitt, M. & Perutz, M. F. (1988). Aromatic rings act as hydrogen bond acceptors. *J. Mol. Biol.* **201**, 751–754.
36. Perutz, M. F., Fermi, G., Abraham, D. J., Poyaet, C. & Bursaux, E. (1986). Hemoglobin as a receptor of drugs and peptides: X-ray studies of the stereochemistry of binding. *J. Am. Chem. Soc.* **108**, 1064–1078.
37. Parsley, T. B., Cornell, C. T. & Semler, B. L. (1999). Modulation of the RNA binding and protein processing activities of poliovirus polypeptide 3CD by the viral RNA polymerase domain. *J. Biol. Chem.* **274**, 12867–12876.
38. Peters, H., Kusov, Y. Y., Meyer, S., Benie, A. J., Bauml, E., Wolff, M. *et al*. (2005). Hepatitis A virus proteinase 3C binding to viral RNA: correlation with substrate binding and enzyme dimerization. *Biochem. J.* **385**, 363–370.
39. Gerber, K., Wimmer, E. & Paul, A. V. (2001). Biochemical and genetic studies of the initiation of human rhinovirus 2 RNA replication: identification of a cis-replicating element in the coding sequence of 2A(pro). *J. Virol.* **75**, 10979–10990.
40. Gerber, K., Wimmer, E. & Paul, A. V. (2001). Biochemical and genetic studies of the initiation of human rhinovirus 2 RNA replication: purification and enzymatic analysis of the RNA- dependent RNA polymerase 3D(pol). *J. Virol.* **75**, 10969–10978.
41. Yang, Y., Rijnbrand, R., Watowich, S. & Lemon, S. M. (2004). Genetic evidence for an interaction between a picornaviral cis-acting RNA replication element and 3CD protein. *J. Biol. Chem.* **279**, 12659–12667.
42. Yin, J., Paul, A. V., Wimmer, E. & Rieder, E. (2003). Functional dissection of a poliovirus cis-acting replication element [PV-cre(2C)]: analysis of single- and dual-cre viral genomes and proteins that bind specifically to PV-cre RNA. *J. Virol.* **77**, 5152–5166.
43. Gamarnik, A. V. & Andino, R. (2000). Interactions of viral protein 3CD and poly(rC) binding protein with the 5′ untranslated region of the poliovirus genome. *J. Virol.* **74**, 2219–2226.
44. Du, Z., Yu, J., Ulyanov, N. B., Andino, R. & James, T. L. (2004). Solution structure of a consensus stem-loop D RNA domain that plays important roles in regulating translation and replication in enteroviruses and rhinoviruses. *Biochemistry*, **43**, 11959–11972.
45. Andino, R., Rieckhof, G. E., Achacoso, P. L. & Baltimore, D. (1993). Poliovirus RNA synthesis utilizes an RNP complex formed around the 5′- end of viral RNA. *EMBO J.* **12**, 3587–3598.
46. Andino, R., Rieckhof, G. E., Trono, D. & Baltimore, D. (1990). Substitutions in the protease (3Cpro) gene of poliovirus can suppress a mutation in the 5′ non-coding region. *J. Virol.* **64**, 607–612.
47. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
48. Collaborative Computational Project, No. 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
49. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallog. sect. D*, **50**, 157–163.
50. Perrakis, A., Sixma, T. K., Wilson, K. S. & Lamzin, V. S. (1997). wARP: improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models. *Acta Crystallog. sect. D*, **53**, 448–455.
51. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.
52. McRee, D. E. (1999). XtalView/Xfit—A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156–165.
53. Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallog. sect. D*, **55**, 247–255.
54. Laskowski, R. A., Moss, D. S. & Thornton, J. M. (1993). Main-chain bond lengths and bond angles in protein structures. *J. Mol. Biol.* **231**, 1049–1067.
55. Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucl. Acids Res.* **16**, 10881–10890.
56. Cohen, G. H. (1997). ALIGN: a program to superimpose protein coordinates, accounting for insertions and deletions. *J. Appl. Crystallog.* **30**, 1160–1161.
57. Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics*, **15**, 305–308.


***Edited by I. Wilson***

*(Received 11 July 2005; received in revised form 22 September 2005; accepted 23 September 2005)*
Available online 14 October 2005