

A Norovirus Protease Structure Provides Insights into Active and Substrate Binding Site Integrity†

Kentaro Nakamura,¹ Yuichi Someya,² Takashi Kumasaka,^{1,3*} Go Ueno,³
Masaki Yamamoto,³ Takao Sato,¹ Naokazu Takeda,²
Tatsuo Miyamura,² and Nobuo Tanaka^{1*}

Department of Life Science, Tokyo Institute of Technology, Yokohama 226-8501,¹ Department of Virology II, National Institute of Infectious Diseases, Shinjuku, Tokyo 162-8640,² and RIKEN Harima Institute/SPRING-8, Mikazuki, Hyogo 679-5148,³ Japan

Received 11 May 2005/Accepted 9 August 2005

Norovirus 3C-like proteases are crucial to proteolytic processing of norovirus polyproteins. We determined the crystal structure of the 3C-like protease from Chiba virus, a norovirus, at 2.8-Å resolution. An active site including Cys139 and His30 is present, as is a hydrogen bond network that stabilizes the active site conformation. In the oxyanion hole backbone, a structural difference was observed probably upon substrate binding. A peptide substrate/enzyme model shows that several interactions between the two components are critical for substrate binding and that the S1 and S2 sites appropriately accommodate the substrate P1 and P2 residues, respectively. Knowledge of the structure and a previous mutagenesis study allow us to correlate proteolysis and structure.

Noroviruses (NVs; formerly “Norwalk-like viruses”), which belong to the family *Caliciviridae*, are the major causative agents of nonbacterial acute gastroenteritis in humans (12, 22, 31). NVs have also been called “small round-structured viruses,” a name that describes the electron micrographic morphology of the fixed virion. NVs are serologically and genetically diverse—more than a thousand strains have been isolated worldwide (57). However, only a limited number of NV genomes have been sequenced (29, 32, 35, 49, 50), because appropriate cell culture systems and/or animal models supporting virus propagation have not been developed.

The NV genome, which consists of positive-sense, single-stranded RNA, contains three open reading frames (ORFs). ORF1 encodes a large polyprotein, the nonstructural protein, which is probably processed intracellularly into six proteins by the viral 3C-like protease (6). ORF2 encodes a capsid protein (the major structural protein), of which 180 molecules form a regular icosahedral virion (45). ORF3 encodes a small basic protein, which is probably the minor structural protein (20). The six NV ORF1 nonstructural proteins are homologous to picornaviral nonstructural proteins and are named accordingly: N-terminal protein, 2C-like nucleoside triphosphatase, 3A-like protein, 3B VPg (genome-linked viral protein), 3C-like protease, and 3D RNA-dependent RNA polymerase (3D^{pol}). NV 3B VPg binds to the 5′ end of the genome and to eukaryotic initiation factors such as eIF3 (15). The protein may function like the 5′-cap structure of eukaryotic mRNAs involved in translation initiation. NV 3D^{pol} is an RNA-dependent RNA

polymerase which is unique in that its polymerase activity is not poly(A) tail/primer dependent (19). 3C-like proteases are the key enzymes for ORF1 polyprotein processing (6, 37, 49, 50) and also cleave the poly(A)-binding protein, causing cellular translation inhibition (34). NV 3C-like proteases (NV 3C^{pro}) belong to the chymotrypsin-like protease family, in that they appear to have chymotrypsin-like folds. Picornaviral 3C proteases are also members of the chymotrypsin-like family (17). Probably, most of these chymotrypsin-like viral proteases have a cysteine, rather than a serine, as the active site nucleophile. By individual replacement of the charged residues and the putative active site cysteine of the 3C^{pro} from the norovirus Chiba virus (genogroup I strain) with alanines, His30 and Cys139 were identified as a catalytic dyad, which functions without active participation of an active site carboxyl moiety (51). An active site acidic residue seems nonessential for activity of this 3C^{pro}, although a Glu54-to-Gly mutation abolishes protease activity (26). The results of the mutagenesis study identified five additional residues (Arg8, Lys88, Arg89, Asp138, and His157) as indispensable for protease activity, but their precise roles were not ascertained (51). Characterization of the protease tertiary structure should clarify the functional roles of the residues identified as essential to activity by mutagenesis.

For the study reported herein, we determined the first crystal structure of an NV 3C^{pro} (at 2.8-Å resolution) using the Chiba virus 3C-like protease (CVP) (50) as the paradigm. The Chiba virus was isolated from a patient with gastroenteritis acquired during the 1987 oyster-associated outbreak in Chiba Prefecture, Japan (56). CVP, which is released from the ORF1 polyprotein by autoproteolysis, has 181 amino acid residues (19.4 kDa). The CVP amino acid sequence is homologous to those of picornavirus and coronavirus, as well as those of other noroviruses.

Inspection of the CVP structure clarifies the catalytic mech-

* Corresponding author. Mailing address: Department of Life Science, Tokyo Institute of Technology, Yokohama 226-8501, Japan. Phone and fax: 81-45-924-5707. E-mail for Takashi Kumasaka: tkumasak@bio.titech.ac.jp. E-mail for Nobuo Tanaka: ntanaka@bio.titech.ac.jp.

† Supplemental material for this article may be found at <http://jvi.asm.org/>.

TABLE 1. Data collection^a

Parameter	Result for crystal ^c :			
	Native	Hg-peak	Hg-edge	Hg-remote
Wavelength (Å)	1.0000	1.0050	1.0090	0.9800
Resolution (Å)	42.41–3.00 (3.11–3.00)	43.53–2.80 (2.90–2.80)	43.53–2.80 (2.90–2.80)	43.53–2.80 (2.90–2.80)
No. of reflections	122,454	148,926	147,395	147,765
No. unique	22,748	53,872	53,774	53,342
Completeness (%)	99.7 (99.9)	99.8 (100)	99.7 (100)	99.8 (100)
R_{merge}^b	0.089 (0.484)	0.088 (0.438)	0.088 (0.429)	0.089 (0.435)
$I/\Sigma(I)$	11.6 (3.6)	8.8 (2.5)	8.9 (2.5)	8.8 (2.5)

^a Regarding crystal data, space group was $P6_1$ and unit cell dimensions were as follows: $a = b = 128.89$ and $c = 118.08$ Å.

^b $R_{\text{merge}} = \sum_i \sum_{hkl} |I_i - \langle I \rangle| / \sum_i \sum_{hkl} I_i$; I_i and $\langle I \rangle$ are the observed intensity and the average intensity obtained from multiple measurements, respectively.

^c Numbers in parentheses indicate the highest-resolution shell.

anism and the interpretation of a mutagenesis study (51). We compared the CVP structure with the structures of proteases from poliovirus (PV) (40), human rhinovirus (HRV) (38), hepatitis A virus (HAV) (1, 7), coronavirus (2, 3), and severe acute respiratory syndrome coronavirus (58), as well as the HRV 2A protease structure (42). We expect that drug development for NV-associated gastroenteritis and other diseases caused by the viruses which have the proteases of this group will be facilitated by a global study of the available protease crystal structures.

MATERIALS AND METHODS

Protein purification and crystallization. Recombinant intein-fusion CVP was expressed in *Escherichia coli* BL21(DE3)-CodonPlus-RIL using the plasmid pTYB11[CVP], which contains a Chiba virus protease gene insert (NCBI/GenBank accession number of Chiba virus, AB042808). The gene had been excised from pUC[HisPro] (50). The expressed fusion protein, found in the cell lysate supernatant, was purified over chitin beads. Intein was removed using the IMPACT-CN system (New England Biolabs). The protease, which has a precise sequence, was further purified by gel filtration over a HiLoad 16/60 Superdex (75 µg; Amersham Pharmacia Biotech) column, equilibrated with 20 mM Tris-Cl, pH 8.2, 200 mM NaCl, 5 mM dithiothreitol. Enzymatic activity was spectrophotometrically assayed using the commercially available substrate H-Glu-Ala-Leu-Phe-Gln-pNA (Bachem). The protein solution was concentrated to 20 mg/ml. Although crystals were obtained under various conditions, the best crystals were grown at 20°C using a hanging drop vapor diffusion system consisting of a 3-µl protein solution and 3 µl crystallization buffer drop, equilibrated against crystallization buffer (1 ml). The buffer that produced the best crystals contained 20 mM HEPES, pH 7.5, 0.9 M sodium potassium tartrate ($\text{KNaC}_4\text{H}_4\text{O}_6$). Crystals grew within 2 to 3 weeks and typically had dimensions of approximately $0.3 \times 0.3 \times 0.4$ mm³.

Diffraction data collection and processing. For multiwavelength anomalous dispersion (MAD) data, mercury derivatives were prepared by soaking native crystals in 1 mM mercury chloride crystallization buffer for 24 h at 20°C or by cocrystallization of the protein with a mercury derivative in a solution similar to that used to prepare the underivatized crystal.

X-ray diffraction data were collected at SPring-8 (Hyogo, Japan). All data were collected using crystals flash cooled at 100 K, which were prepared by rapidly transferring the crystal into a cryoprotectant containing up to 30% (vol/vol) glycerol. MAD data were collected at four wavelengths at beamline BL26B2, which was equipped with a Jupiter charge-coupled device detector (Rigaku MSC). MAD data were collected for the f'' maximum, for the f'' minimum, and for remote wavelengths below and above the Hg L_{III} edge. All data were processed using the program CrystalClear. Subsequent calculations were performed using the CCP4 suite of programs (13).

Structure determination. The Hg MAD data were treated as four Hg derivatives, each with an anomalous scattering contribution. All data sets were scaled to a common level using the CCP4 program SCALEIT. Six of the eight mercury atom positions and the initial phases were identified using SOLVE. Next, the minor sites were identified using the initial phases. Then, the phases were refined using the positions of the eight Hg atoms and the program MLPHARE (13, 53, 54). Additional density modification was carried out in several steps using the

CCP4 program DM (14). After the noncrystallographic symmetries operator was determined from the eight Hg positions, solvent flattening, histogram matching, and three twofold averaging were performed. Density modification, as carried out originally by the program DM, was repeated, and the figure of merit was forced to decrease 0.4-fold. These operations were repeated three times. The electron density map showed almost the entire polypeptide chain for each of the four molecules of the asymmetric unit. The initial model was built using XtalView (39). The model was refined at 2.8-Å resolution using Refmac5 and CNS. The crystallographic R factor converged to a value of 0.208, with an associated R_{free} of 0.240 (10, 41). Data quality and refinement statistics are given in Tables 1 and 2, respectively. The quality of the structural model and its agreement with the structure factors were checked using the programs PROCHECK, PROMO-TIF, and SFCHECK.

Modeling study with substrate. An oligopeptide, corresponding to the third domain of the turkey ovomucoid inhibitor (OMKTY3) residues 14 to 21, which are the P5-P3' residues, was docked onto the CVP substrate binding site. Docking was performed by the CCP4 program lsqkab so that the alignment between the oligopeptide and CVP residues of the active site triad, the oxyanion hole, and the β -strand CVP residues 158 to 161 mimicked that of the oligopeptide and the corresponding *Streptomyces griseus* protease (SGPB) residues. Then, the OMKTY3 side chains were replaced with the corresponding side chains of the peptide Glu-Thr-Thr-Leu-Glu-Gly-Gly-Asp, which is the P5-P3' cleavage sequence for 3CD. Finally, the substrate conformation was first manually adjusted

TABLE 2. Refinement parameters

Parameter	Peak	
	Native	Hg
Resolution range (Å)	8.00–3.20	8.00–2.80
$R_{\text{work}}, R_{\text{free}}^a$	0.194 (0.272), 0.225 (0.321)	0.206 (0.299), 0.240 (0.342)
Ramachandran plot ^b		
% Allowed regions	100	100
% Disallowed regions	0	0
No. of atoms or molecules		
Protein atoms	5,212	5,212
Tartrate atoms	20	20
Water molecules	37	127
Hg atoms		8
Avg B (Å ²)	67.7	64.4
Matthews coefficient/solvent %	3.83/67.59	3.77/67.15

^a $R = \sum_{hkl} \|F_{\text{obs}}(h, k, l) - k|F_{\text{calc}}(h, k, l)|\| / \sum_{hkl} F_{\text{obs}}(h, k, l)$; R_{work} was calculated using a set of reflections where 10% of the total reflections had been randomly omitted from the refinement and used to calculate R_{free} .

^b Percentage of residues in the “allowed” and “disallowed” region of the Ramachandran plot (PROCHECK).

and then energy minimized to remove atomic overlaps while retaining the antiparallel β -strand interaction.

Figures were made using PyMOL (16), DINO (<http://www.dino3d.org>), MSMS (18), MEAD (4), and Secseq.

Protein structure accession number. The coordinates and structure factors were deposited in the Protein Data Bank (entry 1WQS).

RESULTS AND DISCUSSION

Global and domain structures. There are four CVP monomers (designated A, B, C, and D) per asymmetric unit. Each monomer consists of an N-terminal domain and a C-terminal domain (Fig. 1A). The secondary structure composition is primarily and extensively β -strand. A secondary structure topological representation and an assignment are shown in Fig. 1B.

The N-terminal domain has two α -helices and seven β -strands (aI, bI, cI, dI, eI, fI, and gI). The β -strands form a twisted antiparallel β -sheet resembling an incomplete β -barrel. N-terminal domain antiparallel β -barrels exist in the viral chymotrypsin-like cysteine proteases—HAV 3C^{Pro} (1, 7), foot-and-mouth disease virus 3C^{Pro} (8), tobacco etch virus protease (43), HRV 3C^{Pro} (11, 38), and PV 3C^{Pro} (9, 24, 28, 36, 40)—and in the coronavirus 3C-like proteases (also known as main proteases, M^{Pro} [2, 3, 58]). However, the CVP N-terminal domain incomplete β -barrel is intermediate in structure between the N-terminal domain four β -strands of HRV 2A^{Pro} (42) and the corresponding β -barrels of other chymotrypsin-like proteases. The core of the CVP incomplete β -barrel contains the hydrophobic residues Phe12, Phe25, Phe39, Phe40, Phe58 and Phe60, Trp19, Ile32, Ile44, Ile47, and Ile49 (see Fig. S1 in the supplemental material). The active site residue His30 is found in the N-terminal domain.

The C-terminal domain is a six-stranded antiparallel β -barrel, formed by strands aII, bII, cII, dII, eII, and fII. The active site Cys139 is found in the C-terminal domain. The catalytic site formed is situated deep within a cleft between the N- and C-terminal domains.

When the four monomers of the asymmetric unit are examined, the following structural features are found noteworthy. The root mean square deviation (RMSD) for the positions of 71 core C $_{\alpha}$ atoms, with four equivalent atoms per asymmetric unit, is 0.33 (± 0.02) Å. The RMSDs given above and below—with the associated uncertainties reported as standard deviations—are the averages of the RMSDs for the coordinates of included C $_{\alpha}$ positions, whereas the RMSD for 162 equivalent C $_{\alpha}$ atoms is 0.64 (± 0.10) Å (cutoff, 2.0 Å). Excluding residues sequentially adjacent to the N and C termini, the largest conformational variations are found for (i) a flexible surface loop (residues 33 to 36; RMSD, 1.68 \pm 1.09 Å) following the dI strand, (ii) a hairpin loop (residues 107 to 113; RMSD, 1.33 \pm 0.62 Å), (iii) a long loop (residues 124 to 132; RMSD, 2.10 \pm 0.98 Å), (iv) a loop (residues 147 to 150; RMSD, 1.17 \pm 0.55 Å), and (v) a β -hairpin loop (residues 162 to 164; RMSD, 1.42 \pm 0.89 Å) (Fig. 2). These loops are all solvent accessible and probably flexible. The C-terminal residues, 174 to 181, could not be traced in the electron density map for any of the four crystallographic independent molecules, perhaps because of disorder.

Active site. For most chymotrypsin-like proteases, a catalytic triad, composed of Ser or Cys, His, and Asp or Glu, is responsible for proteolysis (17). However, no obligatory acidic resi-

due, other than Asp138, that affected CVP activity was identified by alanine scanning (51). Therefore, the general base catalyst His30 and the nucleophilic Cys139 were proposed as a catalytic dyad. Asp138 could not be a third active site residue, because, sequentially, it is next to Cys139. Its side chain cannot be adjacent to that of His30. Examination of the CVP structure confirms that His30 and Cys139 are catalytic residues and shows that Glu54 is in the position of the third active site residue (Fig. 1C; for the stereo view around Cys139, see Fig. S2 in the supplemental material). The active site is located at the center of a deep cleft between the N- and C-terminal domains, as are the active sites of other chymotrypsin-like proteases. His30 is part of a short α -helix that follows strand cI, and Glu54 is part of a loop connecting strands fI and gI. Cys139 is part of a hairpin turn between an α -helix (Pro136-Asp138) and strand dII.

His30, Glu54, and Cys139 are conserved in all NV 3C-like proteases (Fig. 3). Mutagenesis of His30 to other residues always caused the enzyme to lose activity, indicating that His30 is indispensable (51). A mutant Ser139 CVP retains activity, whereas introduction of a threonine, tyrosine, or methionine at position 139 abolishes activity (51). Cys139 and His30 seem to work the nucleophilic residue and the general acid-base catalyst residue, respectively. On the other hand, the replacement of Glu54 by an alanine does not significantly affect protease activity. Only two residues, Cys139 and His30, are essential for CVP activity, and although it is not essential, Glu54 seems to fulfill the additional role in the activity. The situation for trypsin is different. In that case, if the active site aspartic acid is replaced by an asparagine, activity is severely diminished (52). The pK_a of a cysteine thiol is usually about 8.3, whereas that of a serine hydroxyl is about 13. Therefore, a cysteine thiol will ionize more readily than will a serine hydroxyl. This difference in ionization strengths probably accounts for the difference in the requirement for a catalytic carboxyl moiety. In addition, it suggests the possibility that the His30 imidazolium might exist and the chemical state of the Cys139-His30 interaction might be a thiolate-imidazolium ion pair, which is also found in papain-like proteases (44). For HAV 3C^{Pro}, although the position of Asp84 is spatially equivalent to the third triad member, its side chain points away from the general base His44 (7). Additionally, for coronavirus M^{Pro}, the residue equivalent to Glu54 is Val84. Its aliphatic side chain obviously cannot participate in catalysis (2), and a water molecule occupies the position of the third catalytic residue of the conventional triad. A catalytic dyad must be operational in M^{Pro}. Therefore, in general, a carboxyl moiety may not be necessary for catalysis by viral chymotrypsin-like cysteine proteases. For some proteases, an asparagine residue occupies the position corresponding to the active site carboxyl moiety of chymotrypsin-like proteases and can only hydrogen bond with the active site histidine. Sárkány et al. (47) suggested that a catalytically competent thiolate-imidazolium ion pair, not an imidazole general base catalyst, functions in the HRV 2A^{Pro} catalytic mechanism. While, in the C139S mutant, His30 is likely to work as a general base, rather than an imidazolium ion, it is still uncertain whether His30 works as a general base or as an imidazolium ion in native CVP, which has Cys139 (51). Kinetic and mutagenesis studies are needed to confirm that a third member is

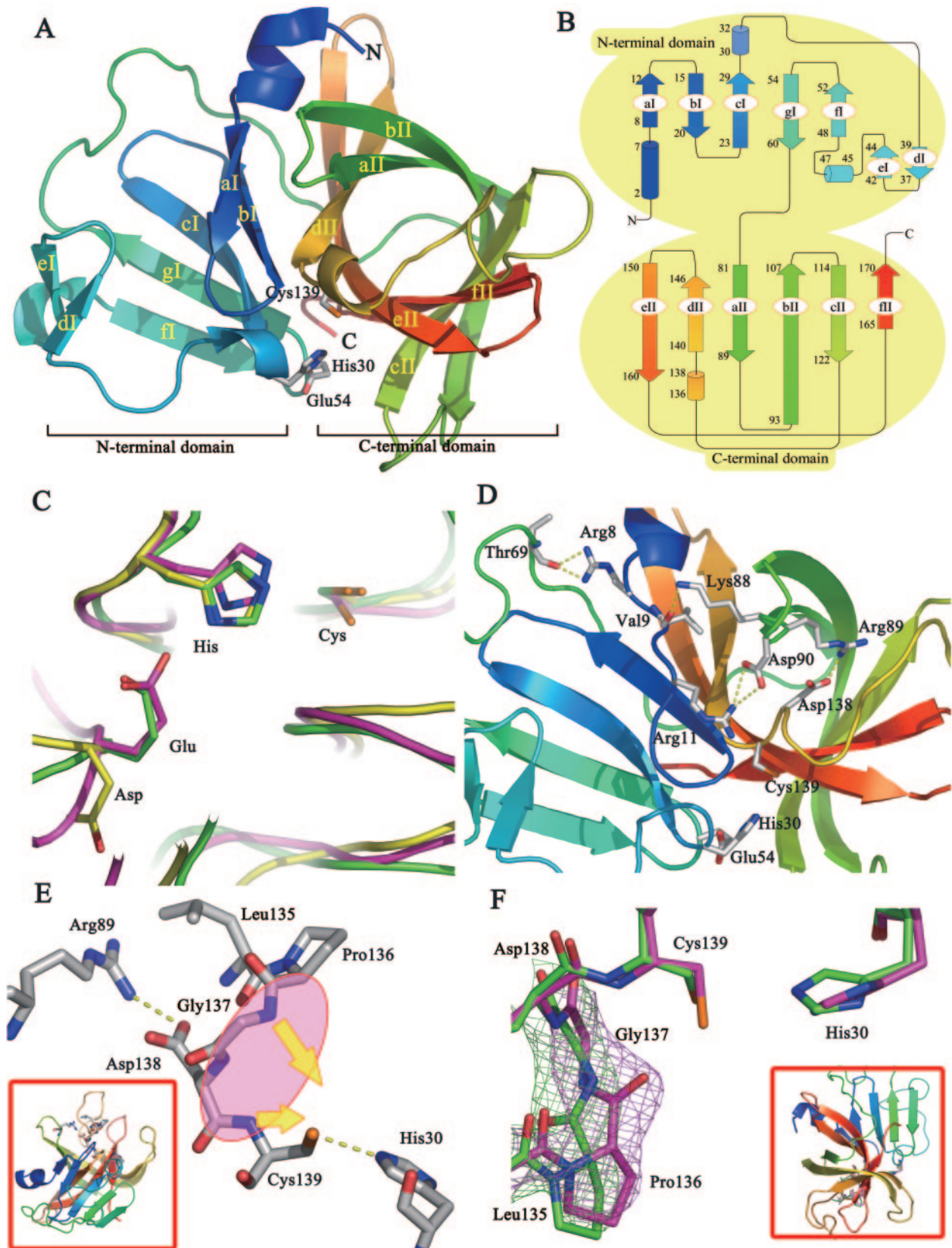


FIG. 1. The CVP crystal structure. (A) The CVP global structure. The N- and C-terminal domains are identified in the figure. Helices and β -strands are represented as ribbons and arrows, respectively. The three active site residues are labeled and represented as stick figures. (B) The CVP secondary structure topology. Residue numbers delineate the beginning and end of each of the secondary structure elements. (C) Superimposition of the CVP active site residues (green) onto the spatially equivalent HAV 3C^{Pro} residues (yellow) and onto the HRV 3C^{Pro} residues

not necessary for the activity and to confirm the possibility of a CVP thiolate-imidazolium ion pair.

The average distance between the sulfur of Cys139 and the N^{e2} of His30 is 3.4 Å in molecules B and C, which is that expected theoretically (23) and is similar to the distances measured for HRV 3C^{pro} (3.5 Å [38]) and for PV 3C^{pro} (3.4 Å [40]). The Cys sulfur-HisN^{e2} distance in molecule A (3.2 Å) is a little shorter than the distances found for the other molecules of the asymmetric unit. Tartrate was included in the crystallization buffer, and an electron density corresponding to tartrate is found between Cys139 and His30 of molecule A (see Fig. S4 in the supplemental material). It is possible that an interaction between molecule A and a tartrate mimics the substrate binding state. For molecule D, the Cys sulfur-HisN^{e2} distance (approximately 4 Å) is longer than average. The structures of coronavirus M^{pro} and a part of HAV 3C^{pro} are inactivated, and the distance of separation between their catalytic Cys sulfur and HisN^{e2} increased when their active site sulfurs were oxidized to either sulfinic acid or sulfonic acid during crystallization.

As noted above, although Glu54 is not essential for CVP activity, it is plausible that Glu54 decreases the range of motion for His30 if a negatively charged carboxylate and the positively charged imidazolium interact. The average distances between the Glu54 carboxyl oxygen and the His30N^{δ1} of molecules A, B, C, and D are 3.4 Å, 2.8 Å, 2.8 Å, and 4.1 Å, respectively; many of the values are within hydrogen bonding distance. Therefore, as indicated in the work of Someya et al. (51), CVP could be more active with Glu54 than without Glu54.

Examination of the CVP crystal structure shows that several hydrogen bonds help maintain the integrity of the active site and/or substrate binding site. These hydrogen bonds include one between the side chain nitrogen of Lys88 and the backbone carbonyl oxygen of Val9 and a second one between the guanidinium nitrogen of Arg8 and the backbone carbonyl oxygen of Thr 69 (Fig. 1D). These hydrogen bonds involve residues bridging the N- and C-terminal domains. Identification of these hydrogen bonds clarifies why Arg8 and Lys88 are conserved in NV 3C^{pro}s and why, when they are mutated to other residues in CVP, activity is lost (51). An interaction between the Asp90 side chain oxygens and an Arg11 guanidinium nitrogen helps orient the N- and C-terminal domains. Mutants with either the first five residues or the first eight residues deleted retain a low level of activity; whereas a mutant missing the first 11 residues is inactive (51). Therefore, the hydrogen bond made by the main chain oxygen of Val9 and the side chain amino group of Lys88 appears vital for activity. The integrity of the active site and/or of the substrate binding site and indeed the entire molecule may depend heavily on the hydrogen bonds described above.

Oxyanion hole. The amino acid sequence Gly-Xaa-Ser/Cys-Gly, which includes the active site nucleophile (underlined), is

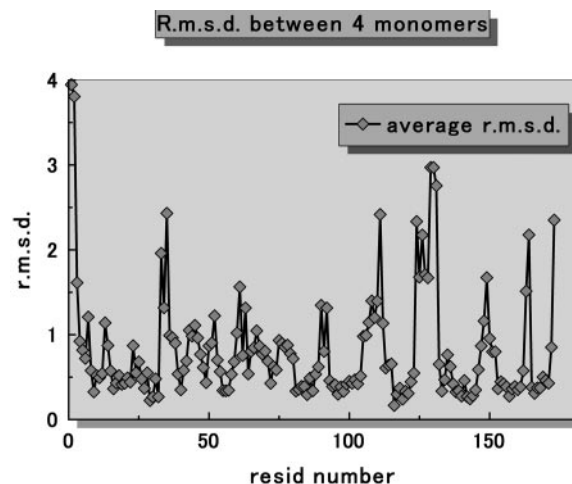


FIG. 2. Root mean square differences among the C_α atomic coordinates after superimposition of monomers A, B, C, and D.

highly conserved in chymotrypsin-like proteases (5). Aspartic acid occupies the Xaa position in most eukaryotic proteases, such as chymotrypsin and trypsin; in bacterial serine proteases; and in viral 2A cysteine proteases. However, for viral 3C or 3C-like cysteine proteases, the residue occupying the Xaa position varies, with Asp, Gln, Met, Tyr, or Trp often present. For CVP, the Gly-Asp-Cys-Gly motif forms the pocket of the oxyanion hole. The oxyanion hole helps to bind the substrate tightly and to stabilize the tetrahedral transition state by hydrogen bonding with the negatively charged P1-backbone carbonyl oxygen.

The CVP oxyanion hole is formed by a region preceding Cys139 and two backbone amides (those of Gly137 and Cys139), which point towards the P1-carbonyl oxygen. Additionally, a side chain Asp138 oxygen clearly forms a good hydrogen bond with the Arg89 side chain nitrogen (Fig. 1E) as reflected by the well-ordered corresponding regions of the electron density map. These two residues are conserved in all noroviruses, and no other residues tested could replace either Arg89 or Asp138 (51). Neither an R89K mutant nor a D138N mutant was active, although the R89K mutant had a positive charge at position 89 and the D138N should have been able to form a hydrogen bond between residues 89 and 138. Clearly, Arg89 and Asp138 are essential for activity and function by stabilizing the oxyanion hole.

In molecule A, the oxyanion hole is large enough to accommodate the main chain carbonyl oxygen of the P1 residue. However, the oxyanion holes of molecules B, C, and D are much narrower than that of molecule A. The volume differences are attributed to the conformations of Pro136 and Gly137 (Fig. 1F). The Pro136 carbonyl oxygen is displaced to the exterior of the molecule A oxyanion hole, whereas, for the

(purple). (D) The hydrogen bond network that stabilizes the active site or substrate binding site. Dashed yellow lines indicate hydrogen bonds; residues contributing to the network are labeled. (E) The oxyanion hole. The catalytic residues, Cys139 and His30, are shown. Dashed yellow lines indicate hydrogen bonds. The oxyanion hole is highlighted in pink. (F) Superimposition of molecule A and C oxyanion holes. The carbons of molecule A and C are colored green and purple, respectively. The $2F_o - F_c$ electron density maps of the Gly137 and Pro136 backbones are represented by colored mesh.



FIG. 3. Sequence alignment of norovirus 3C proteases. The initial alignment, followed by adjustments to improve agreement with the CVP structure, was performed using CLUSTAL X, version 1.81 (55), and Secseq (D. E. Brodersen, unpublished software; <http://xray.imsb.au.dk/~deb/secseq>). CVP β -strands and α -helices were introduced into the figure using PROMOTIF (27) and Secseq. The amino acid sequences were translated from the genomic polyprotein sequences that are available at NCBI/GenBank (accession numbers: Norwalk, NP_786949; Southampton, A37491, BS5, and AAC64602; Camberwell, AAD33960; Lordsdale, P54634; Hawaii, AAB97767).

other molecules, the oxygen faces inward. In the electron density maps, the different orientations could be clearly observed. Recall that a tartrate was found only in the active site of molecule A and mimics a substrate binding state. Therefore, in the absence of substrate, the oxyanion hole is stabilized by an electrostatic interaction between the carbonyl oxygen of Pro136 and the numerous peptide amides that line the hole, plus the aforementioned interaction involving Asp138/Arg89. Perhaps, substrate binding reduces the structural flexibility of the main chain of Pro136 and Gly137, so that the carbonyl oxygen of Pro136 rotates to the exterior of the oxyanion hole by almost 180 degrees, while the Gly137 backbone amide turns inward. The effect upon substrate binding in the oxyanion hole structure, as discussed in HRV 3C^{Pro}, has been deduced (38).

Substrate binding site. We built a model of an oligopeptide bound to CVP. This peptide, Glu-Thr-Thr-Leu-Glu*Gly-Gly-Asp, corresponds to P5-P3' of a substrate, and the asterisk marks the scissile bond. The notation Pn-P1'-Pn' for the residues of substrate or inhibitor is that of Schechter and Berger (48). (P1-P1' are the scissile bond residues.) We felt that a peptide/substrate model would increase our understanding of CVP substrate specificity. To model the peptide/CVP complex, we consulted various protease/inhibitor complex studies performed previously and used, as the starting point for the substrate conformation, the conformation of residues 14 to 21 of OMTKY3 when bound to SGPB (46) (see Materials and Methods for modeling details). For the peptide/CVP model, canonical substrate-binding interactions exist, including interactions between the enzyme and the substrate P4-P2 residues. In order to confirm the validity of this model, we determined the CVP-substrate complex crystal structure (H-Glu-Ala-Leu-

Phe-Gln-pNA [Bachem] was used as a substrate). As a result, although the resolution is low, the determined electron density maps indicate the main chain of the substrate and support this substrate binding model (data not shown; see Fig. S5 in the supplemental material). Figure 4C portrays the docked substrate, with the CVP oxyanion hole and eII strand visible. The substrate fills the CVP binding pocket of molecule A without any severe van der Waals clashes. However, for molecules B, C, and D, the P1 amino acid is too close to the Pro136 carbonyl oxygen to fit in the substrate binding pocket. The inability of molecules B, C, and D to accommodate the P1 residue is a reflection of the Pro136 carbonyl oxygen position, which, as noted previously, protrudes into the oxyanion hole. The oxyanion hole of molecule A in the substrate binding model is shown in Fig. 4B.

Examination of the peptide/CVP model suggests that substrate binding is stabilized by interactions involving hydrogen bonds formed by the backbone of residues P5-P2 and CVP residues 158 to 162 (Fig. 4A and C). The principal hydrogen bonds involve the backbone oxygen of P5 and the backbone nitrogen of Lys162, the backbone nitrogen of P3 and the backbone oxygen of Ala160, and the backbone oxygen of P3 and the backbone nitrogen of Ala160. The P3 and P4 backbone atoms lie in a tunnel formed by residues 159 to 162 and 108 to 110 of CVP (Fig. 4D). This tunnel has a wall built of Gln110 and Lys162 side chains. The wall seems to have the potential to open and close because the Gln110 and Lys162 side chains do not interact with other residues, which, in turn, provides freedom of movement. Furthermore, since the pockets for the side chains of P4 and P3 are large, the pockets can accommodate variously sized residues (e.g., Thr, Leu, Ala, Met, or Phe in the

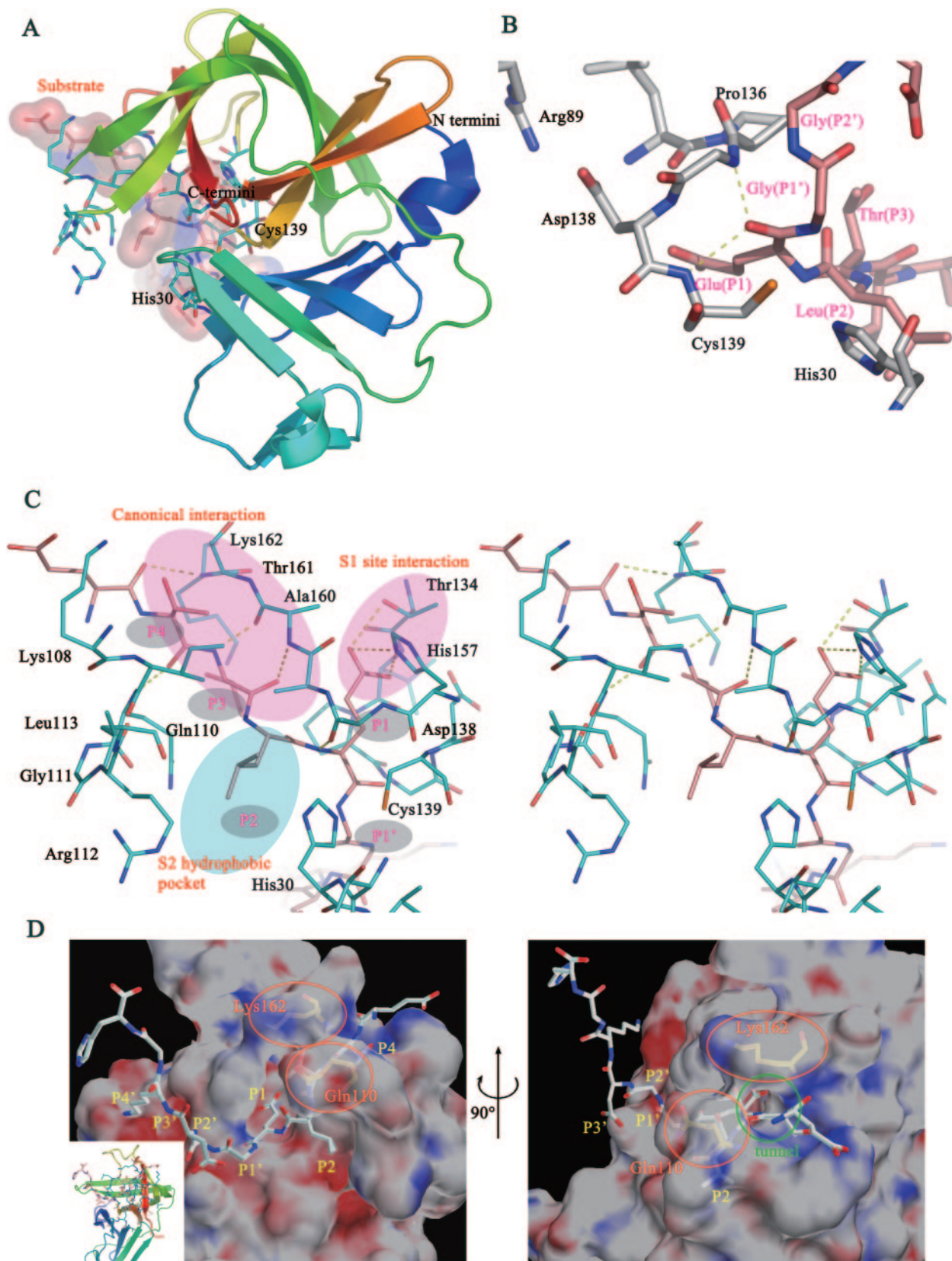


FIG. 4. (A) The CVP/substrate model. The substrate is represented by stick figures and is surrounded by a molecular surface representation. Secondary structures are colored as in Fig. 1. (B) The CVP oxyanion hole complexed with substrate. The substrate is predominantly colored pink, and CVP is predominantly colored white. (C) A stereo view of the model showing substrate bound to CVP. The substrate is predominantly colored pink, and CVP is predominantly colored blue. The antiparallel β -strand interactions and S1 site interactions are highlighted in pink. The S2 hydrophobic pocket is highlighted in sky blue. (D) The electrostatic potential of the CVP/substrate model. Positive and negative potentials are colored blue and red, respectively. "P" denotes substrate residues. An orange circle surrounds Lys162 and Gln110 of CVP. The carbon atoms of Lys162 and Gln110 are colored yellow. The tunnel of the substrate binding site is shown as a green circle.

P4 pocket and Thr, Ser, Gln, or His in the P3 pocket). In the model the allowed volume for the P1' residue is small. This observation is consistent with the fact that glycine or alanine is the P1' residue. The structure of the model explains the variety in P3 and P4 residues and why a glycine or an alanine is required at P1'.

The S1 and S2 sites. The residues at NV ORF1-polypeptide cleavage sites are glutamine or glutamic acid at the P1 site and glycine or alanine at the P1' site. No other residues occupy these sites (Fig. 3). His157, which is part of the S1 site, is important to substrate binding (Fig. 4C). Mutagenesis of His157 to any other tested residue severely reduces activity (51). For HRV 3C^{pro}, HAV 3C^{pro}, PV 3C^{pro}, and coronavirus M^{pro}, there are histidines at the S1 sites and their imidazoles interact with substrate P1 carboxamide side chains (1, 2, 3, 7, 38, 40, 58).

In the CVP/peptide model, the His157 imidazole is positioned to interact with the P1 glutamine side chain. His157 is centered in a hydrophobic pocket, which is composed of Ile85, Ile87, Leu97, Val99, Leu121, Pro136, Tyr143, Ala160, and Val167. Together, the Leu135 backbone, the Pro136 pyrrolidine ring, and the Ala160 methyl form the entrance to the S1 pocket. To ensure proper placement of the His157 imidazole, it is stabilized by a hydrogen bond with the buried Tyr143 hydroxyl, which has no other hydrogen-bonding partner. The average distance of His157N^{δ1}-Tyr143O^δ is 2.77 Å ± 0.12 Å (see Fig. S3 in the supplemental material). NV 3C^{pro} S1 pocket residues are highly conserved, and Tyr143 is always present. The S1 site histidines are similarly stabilized in other 3C or 3C-like proteases. For PV 3C^{pro}, the hydroxyl of Tyr138 hydrogen bonds to the S1 His161 imidazole, and for coronavirus M^{pro}, the analogous interaction is between the side chains of Tyr160 and His162. For HAV 3C^{pro}, Glu132 interacts through two water molecules with the S1 His191. Examination of the CVP/peptide model shows that the oxygen of the P1 glutamine side chain can hydrogen bond with the hydroxyl of Thr134, which is a residue that is conserved in all noroviruses (Fig. 4C). The distance between the two atoms is ~2.7 Å. A P1 residue may be stabilized by both His157 and Thr134 in the NV 3C^{pro}s. In addition, as mentioned above, NVs have glutamic acid or glutamine at the P1 site, but PV, HAV, HRV, and coronavirus have just a glutamine at the P1 site. In comparison of CVP structure with those 3C^{pro}s, two major differences were found. The first is the space size of the S1 site. The S1 site of CVP is larger than other viral 3C^{pro}s; therefore, the χ₁ of the P1 residue seems to be easy to rotate. The other point is Asp138 of CVP. Although other residues of the S1 site are similar to other 3C^{pro}s, CVP-Asp138, which is located facing His157, is different. Other viral 3C^{pro}s have Gln at the position. If χ₁ of the P1 residue rotates, Asp138 of CVP seems to be able to receive the side chain of the P1 glutamine and form a hydrogen bond further than Gln of other viral 3C^{pro}s. Additional structural and biochemical studies are required to elucidate this speculation.

A bulky hydrophobic amino acid, such as Leu, Met, and Phe, preferentially occupies the P2 positions of the polypeptide cleavage sites. The S2 site is also a hydrophobic pocket. It consists of Ile109, Val114, and Ala159 side chains and the Arg112 C^γ and C^δ atoms. The S2 pocket is large enough to accommodate a bulky P2 side chain (Fig. 4C).

Conclusions. Examination of the CVP crystal structure identified the following important structural features: (i) in the overall structure, CVP, which has a chymotrypsin-like fold, resembles other viral 3C, 3C-like, and 2A proteases; (ii) the active site is located in a deep cleft between the N- and C-terminal domains and includes His30, Cys139, and unessential Glu54; (iii) on the basis of the substrate binding model, substrate/enzyme binding involves antiparallel β-strand interactions and substrate P1 and P2 interaction with the enzyme S1 specificity site His157 and the hydrophobic S2 pocket, respectively; (iv) a hydrogen bond network, which is formed by several residues and was identified as essential by mutagenesis, contributes to the structural integrity of the active site and the domains; (v) there is a structural difference in backbone amides of the oxyanion hole probably upon substrate binding.

We believe that the CVP crystal structure is an appropriate structural model for structural studies involving other viral 3C or 3C-like proteases. A global comparison of viral 3C^{pro}s would be especially useful if drug development for nonbacterial acute gastroenteritis and other diseases associated with viruses expressing 3C^{pro} were the result.

ACKNOWLEDGMENT

This work was supported in part by The National Project on Protein Structural and Functional Analyses (Priority Research Program, Protein 3000 Project) to N.T. and T.K.

REFERENCES

- Allaire, M., M. M. Chernaia, B. A. Malcolm, and M. N. James. 1994. Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases. *Nature* **369**:72–76.
- Anand, K., G. J. Palm, J. R. Mesters, S. G. Siddell, J. Ziebuhr, and R. Hilgenfeld. 2002. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. *EMBO J.* **21**:3213–3224.
- Anand, K., J. Ziebuhr, P. Wadhvani, J. R. Mesters, and R. Hilgenfeld. 2003. Coronavirus main proteinase (3CL^{pro}) structure: basis for design of anti-SARS drugs. *Science* **300**:1763–1767.
- Bashford, D., and K. Gerwert. 1992. Electrostatic calculations of the pK_a values of ionizable groups in bacteriorhodopsin. *J. Mol. Biol.* **224**:473–486.
- Bazan, J. F., and R. J. Fletterick. 1988. Viral cysteine proteases are homologous to the trypsin-like family of serine proteases: structural and functional implications. *Proc. Natl. Acad. Sci. USA* **85**:7872–7876.
- Belliot, G., S. V. Sosnovtsev, T. Mitra, C. Hammer, M. Garfield, and K. Y. Green. 2003. In vitro proteolytic processing of the MD145 norovirus ORF1 nonstructural polyprotein yields stable precursors and products similar to those detected in calicivirus-infected cells. *J. Virol.* **77**:10957–10974.
- Bergmann, E. M., S. C. Mosimann, M. M. Chernaia, B. A. Malcolm, and M. N. James. 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.
- Birtley, J. R., S. R. Knox, A. M. Jaulent, P. Brick, R. J. Leatherbarrow, and S. Curry. 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.
- Blair, W. S., J. H. Nguyen, T. B. Parsley, and B. L. Semler. 1996. Mutations in the poliovirus 3CD proteinase S1-specificity pocket affect substrate recognition and RNA binding. *Virology* **218**:1–13.
- Brünger, A. T., P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, and G. L. Warren. 1998. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**:905–921.
- Cheah, K. C., L. E. Leong, and A. G. Porter. 1990. Site-directed mutagenesis suggests close functional relationship between a human rhinovirus 3C cysteine protease and cellular trypsin-like serine proteases. *J. Biol. Chem.* **265**:7180–7187.
- Clarke, I. N., P. R. Lambden, and E. O. Caul. 1998. Human enteric RNA viruses: caliciviruses and astroviruses, p. 511–535. *In* B. W. Mahy and L. Collier (ed.), *Topley and Wilson's microbiology and microbial infections*. Arnold, London, United Kingdom.
- Collaborative Computational Project, Number 4. 1994. The CCP4 suite:

- programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**:760–763.
14. **Cowtan, K. D.** 1996. Phase combination and cross validation in iterated density-modification calculations. *Acta Crystallogr. D Biol. Crystallogr.* **52**: 43–48.
 15. **Daughenbaugh, K. F., C. S. Fraser, J. W. Hershey, and M. E. Hardy.** 2003. The genome-linked protein VPg of the Norwalk virus binds eIF3, suggesting its role in translation initiation complex recruitment. *EMBO J.* **22**:2852–2859.
 16. **DeLano, W. L.** 2002. The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, Calif.
 17. **Dougherty, W. G., and B. L. Semler.** 1999. Expression of virus-encoded proteinases: functional and structural similarities with cellular enzymes. *Microbiol. Rev.* **57**:781–822.
 18. **Duncan, B. S., T. J. Macke, and A. J. Olson.** 1995. Biomolecular visualization using AVS. *J. Mol. Graph.* **13**:271–282.
 19. **Fukushi, S., S. Kojima, R. Takai, F. B. Hoshino, T. Oka, N. Takeda, K. Katayama, and T. Kageyama.** 2004. Poly(A)- and primer-independent RNA polymerase of norovirus. *J. Virol.* **78**:3889–3896.
 20. **Glass, P. J., L. J. White, J. M. Ball, I. Lepar-Goffart, M. E. Hardy, and M. K. Estes.** 2000. Norwalk virus open reading frame 3 encodes a minor structural protein. *J. Virol.* **74**:6581–6591.
 21. Reference deleted.
 22. **Green, K. Y., T. Ando, M. S. Balayan, T. Berke, I. N. Clarke, M. K. Estes, D. O. Matson, S. Nakata, J. D. Neill, M. J. Studdert, and H. J. Thiel.** 2000. Taxonomy of the calciviruses. *J. Infect. Dis.* **181**(Suppl. 2):S322–S330.
 23. **Hamilton, W. C., and J. A. Ibers.** 1968. Hydrogen bonding in solids, p. 167. W. A. Benjamin, Inc., New York, N.Y.
 24. **Hämmerle, T., C. U. Hellen, and E. Wimmer.** 1991. Site-directed mutagenesis of the putative catalytic triad of poliovirus 3C proteinase. *J. Biol. Chem.* **266**:5412–5416.
 25. Reference deleted.
 26. **Hardy, M. E., T. J. Crone, J. E. Brower, and K. Ettayebi.** 2002. Substrate specificity of the Norwalk virus 3C-like proteinase. *Virus Res.* **89**:29–39.
 27. **Hutchinson, E. G., and J. M. Thornton.** 1996. PROMOTIF—a program to identify and analyze structural motif in proteins. *Protein Sci.* **5**:212–220.
 28. **Ivanoff, L. A., T. Towatari, J. Ray, B. D. Korant, and S. R. Petteway.** 1986. Expression and site-specific mutagenesis of the poliovirus 3C protease in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **83**:5392–5396.
 29. **Jiang, X., M. Wang, D. Y. Graham, and M. K. Estes.** 1992. Expression, self-assembly, and antigenicity of the Norwalk virus capsid protein. *J. Virol.* **66**:6527–6532.
 30. Reference deleted.
 31. **Kapikian, A. Z.** 1994. Norwalk and Norwalk-like viruses, p. 471–518. In A. Z. Kapikian (ed.), *Viral infections of the gastrointestinal tract*. Marcel Dekker, Inc., New York, N.Y.
 32. **Katayama, K., H. Shirato-Horikoshi, S. Kojima, T. Kageyama, T. Oka, F. Hoshino, S. Fukushi, M. Shinohara, K. Uchida, Y. Suzuki, T. Gojobori, and N. Takeda.** 2002. Phylogenetic analysis of the complete genome of 18 Norwalk-like viruses. *Virology* **299**:225–239.
 33. Reference deleted.
 34. **Kuyumcu-Martinez, M., G. Belliot, S. V. Sosnovtsev, K. O. Chang, K. Y. Green, and R. E. Lloyd.** 2004. Calicivirus 3C-like proteinase inhibits cellular translation by cleavage of poly(A)-binding protein. *J. Virol.* **78**:8172–8182.
 35. **Lambden, P. R., E. O. Caul, C. R. Ashley, and I. N. Clarke.** 1993. Sequence and genome organization of a human small round-structured (Norwalk-like) virus. *Science* **259**:516–519.
 36. **Lawson, M. A., and B. L. Semler.** 1991. Poliovirus thiol proteinase 3C can utilize a serine nucleophile within the putative catalytic triad. *Proc. Natl. Acad. Sci. USA* **88**:9919–9923.
 37. **Liu, B., I. N. Clarke, and P. R. Lambden.** 1996. Polyprotein processing in Southampton virus: identification of 3C-like protease cleavage sites by in vitro mutagenesis. *J. Virol.* **70**:2605–2610.
 38. **Matthews, D. A., W. W. Smith, R. A. Ferre, B. Condon, G. Budahazi, W. Sisson, J. E. Villafranca, C. A. Janson, H. E. McElroy, and C. L. Gribbskov.** 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.
 39. **McRee, D. E.** 1999. XtalView/Xfit—a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**:156–165.
 40. **Mosimann, S. C., M. M. Cherney, S. Sia, S. Plotch, and M. N. James.** 1997. Refined X-ray crystallographic structure of the poliovirus 3C gene product. *J. Mol. Biol.* **273**:1032–1047.
 41. **Murshudov, G. N.** 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**:240–255.
 42. **Petersen, J. F., M. M. Cherney, H. D. Liebig, T. Skern, E. Kuechler, and M. N. James.** 1999. The structure of the 2A proteinase from a common cold virus: a proteinase responsible for the shut-off of host-cell protein synthesis. *EMBO J.* **18**:5463–5475.
 43. **Phan, J., A. Zdanov, A. G. Evdokimov, J. E. Tropea, H. K. Peters III, R. B. Kapust, M. Li, A. Wlodawer, and D. S. Waugh.** 2002. Structural basis for the substrate specificity of tobacco etch virus protease. *J. Biol. Chem.* **277**:50564–50572.
 44. **Polgár, L.** 1974. Mercaptide-imidazolium ion-pair: the reactive nucleophile in papain catalysis. *FEBS Lett.* **47**:15–18.
 45. **Prasad, B. V., M. E. Hardy, T. Dokland, J. Bella, M. G. Rossmann, and M. K. Estes.** 1999. X-ray crystallographic structure of the Norwalk virus capsid. *Science* **286**:287–290.
 46. **Read, R. J., M. Fujinaga, A. R. Sielecki, and M. N. James.** 1983. Structure of the complex of *Streptomyces griseus* protease B and the third domain of the turkey ovomucoid inhibitor at 1.8-Å resolution. *Biochemistry* **22**:4420–4433.
 47. **Sárkány, Z., T. Skern, and L. Polgár.** 2000. Characterization of the active site thiol group of rhinovirus 2A proteinase. *FEBS Lett.* **481**:289–292.
 48. **Schechter, I., and A. Berger.** 1967. On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**:157–162.
 49. **Seah, E. L., J. A. Marshall, and P. J. Wright.** 1999. Open reading frame 1 of the Norwalk-like virus Camberwell: completion of sequence and expression in mammalian cells. *J. Virol.* **73**:10531–10535.
 50. **Someya, Y., N. Takeda, and T. Miyamura.** 2000. Complete nucleotide sequence of the Chiba virus genome and functional expression of the 3C-like protease in *Escherichia coli*. *Virology* **278**:490–500.
 51. **Someya, Y., N. Takeda, and T. Miyamura.** 2002. Identification of active-site amino acid residues in the Chiba virus 3C-like protease. *J. Virol.* **76**:5949–5958.
 52. **Sprang, S., T. Standing, R. J. Fletterick, R. M. Stroud, J. Finer-Moore, N.-H. Xuong, R. Hamlin, W. J. Rutter, and C. S. Craik.** 1987. The three-dimensional structure of Asn102 mutant of trypsin: role of Asp102 in serine protease catalysis. *Science* **237**:905–909.
 53. **Terwilliger, T. C.** 1994. MAD phasing: treatment of dispersive differences as isomorphous replacement information. *Acta Crystallogr. D Biol. Crystallogr.* **50**:17–23.
 54. **Terwilliger, T. C., and J. Berendzen.** 1999. Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **55**:849–861.
 55. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**:4876–4882.
 56. **Utagawa, E. T., N. Takeda, S. Inouye, K. Kasuga, and S. Yamazaki.** 1994. 3'-terminal sequence of a small round structured virus (SRSV) in Japan. *Arch. Virol.* **135**:185–192.
 57. **Vinje, J., J. Green, D. C. Lewis, C. I. Gallimore, D. W. Brown, and M. P. Koopmans.** 2000. Genetic polymorphism across regions of the three open reading frames of “Norwalk-like viruses.” *Arch. Virol.* **145**:223–241.
 58. **Yang, H., M. Yang, Y. Ding, Y. Liu, Z. Lou, Z. Zhou, L. Sun, L. Mo, S. Ye, H. Pang, G. F. Gao, K. Anand, M. Bartlam, R. Hilgenfeld, and Z. Rao.** 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.