Structural Basis of Substrate Specificity and Protease Inhibition in Norwalk Virus

Zana Muhaxhiri,^a Lisheng Deng,^b Sreejesh Shanker,^a Banumathi Sankaran,^c Mary K. Estes,^d Timothy Palzkill,^{b,d} Yongcheng Song,^b B. V. Venkataram Prasad^{a,d}

Verna and Marrs McLean Department of Biochemistry and Molecular Biology,^a Department of Pharmacology,^b and Department of Molecular Virology and Microbiology,^d Baylor College of Medicine, Houston, Texas, USA; Berkeley Center for Structural Biology, Lawrence Berkeley National Laboratory, Berkeley, California, USA^c

Norwalk virus (NV), the prototype human calicivirus, is the leading cause of nonbacterial acute gastroenteritis. The NV protease cleaves the polyprotein encoded by open reading frame 1 of the viral genome at five nonhomologous sites, releasing six non-structural proteins that are essential for viral replication. The structural details of how NV protease recognizes multiple substrates are unclear. In our X-ray structure of an NV protease construct, we observed that the C-terminal tail, representing the native substrate positions P5 to P1, is inserted into the active site cleft of the neighboring protease molecule, providing atomic details of how NV protease recognizes a substrate. The crystallographic structure of NV protease with the C-terminal tail redesigned to mimic P4 to P1 of another substrate site provided further structural details on how the active site accommodates sequence variations in the substrates. Based on these structural analyses, substrate-based aldehyde inhibitors were synthesized and screened for inhibition potency. Crystallographic structures of the protease in complex with each of the three most potent inhibitors were determined. These structures showed concerted conformational changes in the S4 and S2 pockets of the protease to accommodate variations in the P4 and P2 residues of the substrate/inhibitor, which could be a mechanism for how the NV protease recognizes multiple sites in the polyprotein with differential affinities during virus replication. These structures further indicate that the mechanism of inhibition by these inhibitors involves covalent bond formation with the side chain of the conserved cysteine in the active site by nucleophilic addition, and such substrate-based aldehydes could be effective protease inhibitors.

Noroviruses (NoVs) are the leading cause of nonbacterial acute, epidemic gastroenteritis in humans worldwide. They belong to the *Caliciviridae* family and are subdivided into five genogroups (GI to GV) (1, 2). Genogroup I contains the Norwalk virus (NV), a prototype human norovirus, along with Southampton and Chiba viruses. Currently, there is no cell culture system or animal model that supports efficient replication of human noroviruses, thus limiting a detailed knowledge of the molecular basis of virus propagation and consequently hindering development of antiviral therapies.

The NoV genome, which consists of an ~7.7-kb single-stranded, positive-sense RNA, has three open reading frames (ORFs) (3, 4). ORF1 encodes a precursor polyprotein that later undergoes cleavage by a viral protease (5, 6); ORF2 encodes VP1, which is the major capsid protein (7); and ORF3 encodes VP2, a minor capsid protein (8). The polyprotein is proteolytically processed by the viral protease into six individual nonstructural proteins essential for viral replication, which include p48 (9), p41 (10), p22 (6, 11), VPg (12), protease (13, 14), and RNA-dependent RNA polymerase (15, 16) (Fig. 1A). This processing occurs in a sequential manner as a mechanism to regulate viral genome expression and replication (5, 17). Correspondingly, viruses in the family *Picornaviridae* also encode a polyprotein precursor that is proteolytically processed by the viral 3C protease into nonstructural proteins. The calicivirus nonstructural proteins share sequence motif similarities with picornavirus nonstructural proteins, and in both cases the protease is necessary for viral replication, thus allowing functional and structural parallels to be drawn between the two families. The essential role of the protease in polyprotein processing and its requirement for successful viral

replication (6, 13, 17, 18) make it a prime target for antiviral drug design.

The NoV protease is a cysteine protease that adopts a chymotrypsin-like fold comprised of two domains, a twisted β -sheet domain and a β -barrel domain, separated by a groove where the active site is located (14, 19, 20). The active site consists of a catalytic triad, with cysteine (Cys139) as the nucleophile, histidine (His30) as the general base catalyst, and glutamic acid (Glu54) as the anion to orient the imidazole ring of His30 (14, 20). The proteolytic processing of the NoV polyprotein occurs at five sites, with either Q-G, E-G, or E-A as the cleavage junction (Fig. 1A) (5, 17, 21–24). These sites exhibit significant variations in amino acid composition in both the N-terminal (positions P5 to P2) and C-terminal (positions P2' to P5') sides flanking the scissile bond (positions P1 and P1') (Fig. 1B) (5, 17, 25). In addition, previous studies indicated that there is a preferential temporal order in the proteolytic processing of the polyprotein by the NoV protease (17). The Q-G sites are preferentially processed first, followed by the E-A and E-G sites. Mutational studies show that replacing E-G in site 5 with Q-G does not affect the proteolytic efficiency, suggesting that the composition of the neighboring amino acid residues contributes to substrate recognition by the protease during polyprotein processing (17). Although structures of some of the

Received 13 October 2012 Accepted 24 January 2013

Published ahead of print 30 January 2013

Address correspondence to B. V. Venkataram Prasad, vprasad@bcm.tmc.edu.

Copyright © 2013, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JVI.02869-12

FIG 1 (A) Norwalk virus genome organization showing the three ORFs. ORF1 encodes a polyprotein cleaved into six nonstructural proteins: p48, p41, p22, VPg, protease, and RNA-dependent RNA polymerase. The five dipeptide cleavage sites in the polyprotein are shown in red, with black arrows indicating the cleavage site. (B) Cleavage sites (P1-P1') with neighboring amino acid residues recognized by the NV protease during proteolytic processing of the polyprotein. (C) Surface representation of the NV protease with color-coded active site (red) and substrate binding pockets, with the S1 pocket in green, the S2 pocket in teal, and the S4 pocket in magenta. The dashed lines represent the cleft where the substrate binds.

NoV proteases have been determined (14, 26, 27), including the murine NoV protease in complex with a substrate peptide (26) and the protease of Southampton NoV in complex with a peptide mimic inhibitor (28), the structural basis of how NoV protease recognizes various nonhomologous substrate sites within the polyprotein with differential affinities is not understood.

We report here crystallographic analyses of the NV protease bound to two of its natural substrates that provide further understanding of the molecular and structural basis of substrate specificity. These structural analyses indicate that the protease undergoes conformational changes in the binding pockets upon substrate binding, predominantly in the S2 pocket, to accommodate variations in the P2 position of the substrate, suggesting a possible mechanism for how the NV protease recognizes multiple sites in the polyprotein with differential affinities. This is the first demonstration of conformational changes associated with substrate binding in this family of viral proteases. The structural information obtained from these studies allowed us to undertake a peptidomimetic approach to designing tripeptide aldehyde inhibitors that specifically target the substrate binding site in the protease. A series of tripeptide aldehyde inhibitors (29) were synthesized and screened via fluorescence resonance energy transfer (FRET) assay (14) to determine their inhibition constants, followed by crystallographic analyses of NV protease bound to the three inhibitors that exhibited the most potent inhibition. The structures clearly demonstrate covalent bond formation between the carbonyl group of the aldehyde and the thiol of the nucleophilic Cys139 residue. Analyses of protease interactions with inhibitors indicated that hydrophobic interactions involving the S1 and S2 pockets may play a role in modulating inhibitor efficiency. The results of these studies will be useful in the design of new and more potent and broad-spectrum inhibitors that can target proteases from different norovirus strains.

MATERIALS AND METHODS

Protein expression and purification. The native NV protease was cloned into the pET-46 Ek/LIC expression vector and purified over a Ni-nitrilotriacetic acid (Ni-NTA) column, followed by thrombin cleavage of the His₆ tag, as described by Zeitler et al. (14). Purified NV protease was concentrated, stored in 50 mM NaH₂PO₄ (pH 8), 50 mM NaCl, and 5 mM TCEP [tris(2-carboxyethyl)phosphine] buffer, and later used for cocrystallization with different inhibitors. The native NV protease is referred to as NV protease throughout this paper. The crystallographic studies describing the protease-substrate interactions were carried out with a VPg-Pro fusion construct. This construct was initially made with the expectation of obtaining the structure of VPg and possibly the cleavage site between VPg and the protease within the NV polyprotein bound in the protease active site. An NV VPg-Pro construct in which the protease was inactivated by mutation of the active site Cys139 to Ala was cloned into the pET-46 Ek/LIC expression vector. In addition, a thrombin cleavage site was engineered into the construct to later cleave the N-terminal His₆ tag with thrombin. The resulting plasmid was transformed into *Escherichia coli* BL21 (DE3) cells to express the His₆-tagged recombinant protein. The His₆-VPg-Pro^{C139A} mutant was purified via Ni-NTA column chromatography (Qiagen) followed by removal of the His₆ tag with thrombin. Further purification was achieved by size-exclusion chromatography (Superdex S75pg column; GE Healthcare). Purified VPg-Pro^{C139A} was concentrated and stored in buffer containing 50 mM Tris-HCl (pH 7.0), 50 mM NaCl, and 5 mM dithiothreitol (DTT).

Site-directed mutagenesis. A QuikChange site-directed mutagenesis system (Stratagene) was used to mutate the active site Cys139 to Ala according to the manufacturer's protocol. The forward primer used for the mutation was 5'-GGC ACT ATA CCA GGA GAC GCA GGG GCA CCA TAC GTC-3', and the reverse primer was 5'-GAC GTA TGG TGC CCC TGC GTC TCC TGG TAT AGT GCC-3'. This construct was labeled VPg-Pro^{C139A}. Furthermore, the same protocol was used to mutate the last 4 amino acids of the C-terminal tail of the protease (TALE) into the amino acids INFE, using the forward primer 5'-GCT GGA GAG GGC GAA ATA AAT TTT GAG TGA CCG GGC TTC TCC-3' and the reverse comple-

ment primer 5'-GGA GAA GCC CGG TCA CTC AAA ATT TAT TTC GCC CTC TCC AGC-3'. All positive clones underwent sequence analysis to confirm the desired mutations. The construct containing the native C-terminal tail is referred to as VPg-Pro^{C139A (TALE)}, and the construct with the four mutated amino acids in the C-terminal tail is referred to as VPg-Pro^{C139A (INFE)}.

Inhibition assays. The activity of the purified His₆-tagged NV protease was measured in the presence of different aldehyde inhibitors by use of a FRET assay (14). The details of the synthesis of these aldehyde inhibitors will be provided separately (our unpublished data). The kinetic assays were performed in 50 mM NaH₂PO₄ (pH 8), 50 mM NaCl, and 5 mM TCEP buffer. The NV protease (0.5 μM) was mixed with increasing concentrations of the inhibitor and incubated for 30 min at room temperature before the substrate (30 μM) was added to the reaction mix. The substrate was a fluorogenic peptide (EDANS-EPDFHLQGPEDLAK-Dab cyl) corresponding to the natural cleavage site between p48 and p41 in the NV polyprotein. When the fluorogenic peptide was cleaved by the protease, the donor fluorescence was no longer quenched, resulting in an increase in fluorescence. Once the substrate was added, the fluorescence was measured at excitation and emission wavelengths of 360 and 460 nm, respectively. Inhibition of the protease decreased the fluorescence intensity, allowing screening of prospective inhibitors. After the initial velocities were calculated, nonlinear regression analysis using GraphPad5 software was used to calculate the inhibition constant (K_i) for each inhibitor. Because the K_i values were similar to the concentration of enzyme used in the assay, the equation for tight-binding inhibitors was used to fit the inhibition curves and determine the K_i (30).

Crystallization. The VPg-Pro^{C139A (TALE)} construct was concentrated to ~7 mg/ml and crystallized by the hanging-drop vapor diffusion method at 20°C. Crystals were obtained in 1.6 to 1.8 M lithium sulfate and 0.1 M sodium cacodylate, pH 6.5, with 30% glycerol as an additive. The crystals were cryoprotected with a solution containing 20% glycerol or 20% ethylene glycol reconstituted with mother liquor and then flash frozen with liquid nitrogen. The VPg-Pro^{C139A (INFE)} construct was obtained under similar conditions to those for VPg-Pro^{C139A (TALE)}, concentrated to ~7 mg/ml, and crystallized by using a Mosquito automated nanoliter handling system (TTP LabTech) and a crystallizing solution containing 1.9 M lithium sulfate and 0.1 M sodium cacodylate, pH 6.8, with 25% glycerol as an additive. The crystals obtained were flash frozen with liquid nitrogen, using 20% glycerol as a cryoprotectant.

Inhibitor-protease cocrystals were obtained by adding a 10-fold molar excess of each of the inhibitors (initially dissolved in 100% dimethyl sulfoxide [DMSO]) to the NV protease (5 μM [final concentration]) in 50 mM NaH₂PO₄, 50 mM NaCl, and 5 mM TCEP buffer. The mixed sample was incubated for 1 h at 4°C and concentrated to ~9 mg/ml. The final DMSO concentration of the complex was 5%. All three inhibitor-protease cocrystals were obtained in 0.1 to 0.3 M potassium thiocyanate and 25 to 30% polyethylene glycol monomethyl ether 2,000 buffer, using the hanging-drop vapor diffusion method at 20°C. The crystals were cryoprotected with 25 to 30% glycerol and flash frozen.

Data collection. The data sets for VPg-Pro^{C139A (TALE)}, VPg-Pro^{C139A (INFE)}, syc-10-protease, and syc-08-protease complexes were collected on the SBC-CAT 19ID beamline at the Argonne National Laboratory Advanced Photon Source, Chicago, IL. For the syc-59-protease complex, the diffraction data were collected at the Advanced Light Source at Lawrence Berkeley National Laboratory. All data sets were processed using HKL2000 (31). VPg-Pro^{C139A (TALE)} crystallized in the space group P6₁22, with two molecules in the asymmetric unit; VPg-Pro^{C139A (INFE)} crystallized in the space group P6₃22, with one molecule in the asymmetric unit; syc-10-protease and syc-08-protease crystallized in the space group P6₅22, with one molecule in the asymmetric unit; and syc-59-protease crystallized in the P1 space group, with eight molecules in the asymmetric unit. The data collection statistics are summarized in Table 1.

Structure determination and refinement. The structures were determined by the molecular replacement (MR) method, with a previously

published unliganded NV protease structure (Protein Data Bank [PDB] entry 2FYQ) as a phasing model. The initial electron density map was obtained using PHASER (32) as implemented in the CCP4 suite (33) of programs. Following automated model building using ARP/WARP (34) and manual adjustment of the model by use of COOT (35), the structures were refined by iterative cycles of refinement, using Refmac (36) or PHENIX suite (37), with further model building based on $2F_o - F_c$ and $F_o - F_c$ maps. The ligand density in each case was confirmed by simulated annealing difference maps. Both translation, liberation, and screw (TLS) (38) and noncrystallographic symmetry (NCS) restraints were applied during the refinement cycles. The stereochemistry of all structures during the course of refinement was checked using COOT (35) and PROCHECK (39). The final refinement statistics are included in Table 1. Structural figures were prepared using PyMOL (40).

Protein structure accession numbers. Coordinates and structure factors of all structures discussed in this work have been deposited in the Protein Data Bank (www.pdb.org) under accession numbers 4IN2 and 4IN1 for substrate-bound protease structures and 4INZ, 4INH, and 4INQ for inhibitor-bound protease structures.

RESULTS

A VPg-Pro protein construct in which the protease was inactivated by mutating the nucleophilic cysteine (C139) to alanine was purified and crystallized. The structure was determined by MR as described in Materials and Methods. Although the protease part of VPg-Pro^{C139A (TALE)} was clearly represented in the electron density map, the density for the VPg protein was not visible, indicating that VPg is disordered in the crystal. Further analyses of different VPg constructs by a variety of biophysical methods suggested that VPg is predominantly an intrinsically disordered protein (unpublished data). Analysis of the crystal packing, with further validation using simulated annealing difference maps, revealed an interesting structural feature. The C-terminal tail of one protease molecule, i.e., residues 178 to 181, corresponding to positions P4 to P1 of the native substrate (Fig. 1B), was inserted into the active site of the second molecule, providing a clear visualization of protease interactions with its native substrate, TALE. This striking observation prompted us to undertake a redesign of the VPg-Pro construct, with the C-terminal tail of the protease mutated to mimic P4 to P1 residues INFE, corresponding to site 4 in the NV polyprotein (Fig. 1B). Crystallographic analysis of this redesigned construct showed a similar insertion of the C-terminal tail into the active site of the neighboring protease molecule, allowing us to gain further structural insight into how NV protease interacts with different substrates. In this section, we first discuss the structures of the protease with bound TALE and INFE substrates, followed by the design and synthesis of substrate-based aldehyde inhibitors and the crystallographic analysis of the three most potent inhibitors in complex with the protease.

Structure of the protease bound to the TALE substrate. The VPg-Pro^{C139A (TALE)} construct crystallized in the P6₁22 space group, with two molecules (A and B) in the crystallographic asymmetric unit. These crystals diffracted to a resolution of 2.4 Å. In the crystal structure, the C-terminal TALE residues of molecule A, representing P4 to P1 of one of the native substrates in the polyprotein (Fig. 1B), are completely inserted into the substrate binding pocket of molecule B (Fig. 2A). The TALE region binds to a groove between the two domains (β-barrel and β-twisted sheet), mainly through interactions with the β-barrel domain (Fig. 2B). It adopts an extended β-strand conformation in which the substrate backbone hydrogen bonds with the backbone carbonyl and amide groups of the eII β-strand of the protease, forming an antiparallel

TABLE 1 Data processing and refinement statistics

Parameter	Description or value ^a for protease in complex with:				
	TALE	INFE	syc-10	syc-59	syc-8
Space group	P6 ₁ 22	P6 ₅ 22	P6 ₅ 22	P1	P6 ₅ 22
Cell dimensions (Å)					
a	97.582	124.194	122.57	56.335	122.396
b	97.582	124.194	122.57	66.917	122.396
c	270.458	49.717	51.125	100.459	51.354
Angles (°)					
α	90	90	90	90	90
β	90	90	90	90	90
γ	120	120	120	75.93	120
Resolution (Å) (range)	50–2.4 (2.44–2.4)	40–2.05 (2.09–2.05)	50–1.7 (1.73–1.7)	50–1.7 (1.76–1.7)	50–1.5 (1.53–1.5)
Wavelength (Å)	0.97935	0.97921	0.97921	0.9765	0.97921
No. of reflections	432,122	128,128	297,568	352,891	583,485
No. of unique reflections	30,866	14,341	25,322	78,900	36,665
R _{merge} (%) ^b	8.7 (60)	10.5 (44)	7.7 (55.7)	6.0 (44.8)	6.9 (63)
I/σI	45.3 (6.0)	21.4 (5.34)	37.6 (5.96)	21.9 (3.5)	48.0 (5.81)
Completeness (%)	99.9 (100)	97.6 (96.8)	100 (99.8)	99.7 (99.2)	99.9 (99.9)
Redundancy	14 (14.4)	8.9 (9.2)	11.8 (11.9)	4.5 (4.4)	15.9 (14.9)
R _{work} (%) ^c	20.6	17.1	16.3	17.6	17.1
R _{free} (%) ^c	22.2	19.7	19.4	20.8	18.5
Bond length (Å)	0.007	0.007	0.017	0.010	0.020
Bond angle (°)	1.064	1.038	1.579	1.640	1.639

^a Values in parentheses are for the highest resolution.

^b $R_{\text{merge}} = \sum |I_{hkl} - \langle I_{hkl} \rangle| / \sum I_{hkl} \times 100$, where I_{hkl} is the intensity of an individual hkl reflection and $\langle I_{hkl} \rangle$ is the mean intensity for all measured values of this reflection. The summation is over all equivalent intensities.

^c $R_{\text{work}} = \sum ||F_o| - |F_c|| / \sum |F_o| \times 100$, where F_o represents the observed structure factor amplitudes and F_c represents the structure factor amplitudes calculated from the atomic model. R_{free} was calculated with the 5% of reflections excluded from the data set during refinement.

β-sheet (see Fig. 4A). These interactions allow correct positioning of the terminal residue (P1) of the substrate for proteolytic cleavage. Despite the mutation of the active site residue C139 to Ala, the terminal carboxyl group of P1 is properly aligned for nucleophilic attack by C139, reminiscent of the tetrahedral intermediate state. The backbone amide groups of the C139A mutant residue and G137 are correctly positioned, facing toward the active site, to form the oxyanion hole (Fig. 2D). In contrast, the C-terminal tail of molecule B is only partially inserted into the substrate binding pocket of molecule A (Fig. 2C). The electron density is discernible only for the P4 to P2 residues (TAL; residues 178 to 180) of molecule B and is further away from the active site of molecule A than that for the fully engaged TALE region as described above. The backbone and side chain conformations of the TAL region of molecule B vary considerably compared to those of the TALE region of molecule A.

Although the overall conformation of molecule A, with a partially inserted TAL region, is nearly identical to that of molecule B, with a fully inserted TALE region, there are some significant differences which appear to be caused by substrate binding. One striking difference between molecules A and B is in the conformation of the peptide between G137 and P136, which constitute the oxyanion hole. In molecule A (with a partially inserted TALE region), similar to the case in the previously determined native NV protease structure, the amide group of G137 points away from the active site, toward the solvent, and the carbonyl oxygen of P136 faces inward, forming a water-mediated hydrogen bond with Nε2 of the active site residue H30 (Fig. 2D). In contrast, in molecule B

(with a fully inserted TALE region), the G137-P136 peptide is flipped, with the backbone carbonyl group of P136 now pointing away from the active site and the backbone NH of G137 pointing toward the active site, forming a hydrogen bond with the carbonyl group of the terminal carboxyl group of the substrate (Fig. 2E). This carbonyl group of the substrate also engages in a hydrogen bond interaction with the backbone NH of the C139A mutant. The other oxygen of the carboxyl group makes a hydrogen bond with Nε2 of His30, further stabilizing binding of the substrate to the active site (Fig. 2E).

Structure of the protease bound to the INFE substrate. The VPg-Pro^{C139A} construct, in which the C-terminal residues were mutated to INFE, crystallized in the P6₅22 space group, with one molecule in the asymmetric unit, and crystals diffracted to a 1.9-Å resolution. In this structure, similar to the case for molecule B with the TALE substrate, the C-terminal INFE region inserts entirely into the substrate binding pocket of the neighboring symmetry-related molecule, making hydrogen bond interactions involving the backbone atoms of A158 and A160 (Fig. 3). Despite significant differences in amino acid composition compared to the TALE peptide, the INFE peptide adopts a very similar extended β-strand conformation, suggesting that this is the preferred conformation of the substrate bound to NV protease. The INFE region interacts with the carbonyl and amide groups of the eII β-strand of the protease, forming an antiparallel β-sheet with its carboxyl end positioned near the nucleophilic Cys139 residue, resembling a tetrahedral intermediate state. Similar to the case for molecule B with the TALE substrate, the G137-P136 peptide in the INFE structure

FIG 2 (A) Crystal packing of VPg-Pro^{C139A} in the P6₂22 space group, with two molecules in the asymmetric unit: molecules A (yellow) and B (blue). The active site (H30, C139, and E54) is represented by red sticks. The C-terminal tail of molecule A, representing native substrate positions P4 to P1 (yellow sticks), inserts fully into the active site of molecule B (shown in blue), whereas the C-terminal tail of molecule B (blue sticks) only partially inserts into the active site of molecule A (shown in yellow). The C-terminal tails of both molecules were built in the $F_o - F_c$ difference electron density map (3σ), shown in dashed boxes. (B) Surface representation of monomer B of the protease, with TALE (yellow sticks) fully inserted into the substrate binding pocket. (C) Surface representation of monomer A of the protease, with TAL (blue sticks) partially bound to the substrate binding pocket. (D) The last two residues of the TALE substrate (yellow sticks), LE, bound to the active site (red sticks). The TALE region forms hydrogen bond interactions (black dashed lines) with the active site's H30 residue (red) and a water molecule (blue sphere). The oxyanion hole is properly formed, and the amine groups of G137 and the C139A mutant residue (light blue) face the oxyanion hole, forming hydrogen bonds (black dashed lines) with the carbonyl oxygen of the substrate. (E) In the absence of the substrate, the oxygen of the P136 (light blue) carbonyl group points toward the oxyanion hole and interacts with a water molecule (blue), which subsequently interacts with the H30 residue of the active site (red). Consequently, the amide group of G137 points away from the oxyanion hole, resulting in improper oxyanion formation.

is flipped, indicating that the peptide flipping is induced by binding of the substrate to the active site of the protease.

Substrate interactions involving the S1 and S2 binding pockets. Comparative analysis of the protease structures with bound TALE and INFE substrates provided insight into how the amino acid variations are accommodated by the NV protease. The P4 to P1 residues in the substrate interact with specific pockets in the protease, termed S4 to S1, that line the substrate binding groove between the two domains. The majority of the interactions with both TALE and INFE substrates occur at the S1 and S2 pockets (Fig. 4). The S1 pocket, adjacent to the active site, is formed by residues in domain II, H157 and T134 in the loop region that connects β -strands cII and dII, and A160 of the eII β -strand. In both substrates, the P1 residue is Glu, which fits tightly within this pocket through a network of hydrogen bond interactions. Oe1 of the Glu-P1 side chain participates in hydrogen bond interactions with Ne2 of the imidazole ring of H157 and the hydroxyl group of T134, whereas Oe2 makes hydrogen bonds with two ordered water molecules. In the case of the INFE substrate, these water molecules participate in additional hydrogen bonds involving T134, Gln-P3, and the backbone carbonyl oxygen of Ala160. In the na-

tive protease structure, the hydroxyl group of the T134 side chain is hydrogen bonded to H157 (N δ 1) via a water molecule, whereas in the substrate-bound structures, this water molecule is replaced by Oe1 of Glu-P1. Tyr143 (OH) also participates indirectly in the stabilization of the P1 side chain by hydrogen bonding to H157 (N δ 1) such that the imidazole ring is properly oriented for hydrogen bond interactions with the Glu-P1 side chain. In addition to hydrogen bond interactions, Glu-P1 is further stabilized by several hydrophobic interactions in the S1 pocket, involving T134, I135, and P136, belonging to the dII loop, and A158 with A160 of the eII β -strand. These interactions allow the terminal carbonyl carbon of the substrate to be positioned appropriately relative to the nucleophilic Cys139 residue for catalysis.

In comparison with the native protease structure, the S1 pocket does not show significant differences upon substrate binding; however, the S2 pocket does undergo considerable conformational changes to accommodate variations in the P2 position of the substrates. In the NV polyprotein, the P2 position is occupied by hydrophobic residues such as Leu, Phe, and Pro. The S2 pocket is located in the loop between the bII and cII β -strands, defined by residues I109, Q110, R112, and V114 (Fig. 4). In the unbound

FIG 3 Crystal packing of VPg-Pro^{C139A} (INFE) construct. (A) The C-terminal tail (orange sticks) of the protease molecule (cartoon representation in orange), representing the natural substrate (P4 to P1 [INFE]), inserts into the active site of the symmetry-related molecule (shown in blue cartoon representation), and vice versa. The C-terminal tail was built in the $F_o - F_c$ difference electron density map (3σ). The close-up view of the inset shows INFE (orange sticks) bound to the substrate binding pocket of the protease (shown in surface representation).

state, the bII-cII loop is positioned pointing toward the active site, depicted as the closed conformation (Fig. 5A); with TALE substrate binding, the S2 pocket opens up slightly to accommodate the Leu side chain (Fig. 5B), and with binding of the INFE substrate, the S2 pocket opens even further to accommodate the bulkier Phe side chain (Fig. 5C). These alterations indicate the inherent

ability of the S2 pocket to undergo conformational changes in response to the binding of different substrates.

In the case of the TALE substrate, Leu-P2 is stabilized predominantly by the van der Waals contacts provided by I109 and Q110 of the S2 pocket (Fig. 4A). In comparison, with further opening of the S2 pocket upon binding of the INFE substrate, the bulkier Phe-P2 side chain makes additional hydrophobic interactions involving Val114 of the S2 pocket. Additionally, in contrast to the case with the TALE substrate, where the Q110 side chain is involved in hydrogen bond interactions with the backbone amide of Leu-P2, in the case of the INFE substrate, the Q110 side chain is reoriented to form hydrogen bonds with the backbone oxygen of Ile-P4 and water molecules (Fig. 4B). Involvement of I109 of the S2 pocket in the binding of both TALE and INFE is consistent with previous mutational studies indicating an essential role of I109 in substrate binding (41).

S3 and S4 binding pockets. Compared to the S1 and S2 pockets, the S3 pocket is not as well defined. In the NV polyprotein, residues occupying the P3 position in the five substrate sites vary considerably (His, Gln, Val, Asp, and Ala). Except for conserved backbone-to-backbone hydrogen bond interactions involving the P3 residue of the substrates and A160, the other interactions vary with the side chain composition of the P3 residue. In the TALE structure, besides the backbone hydrogen bond interaction of Ala-P3 with A160, there are no additional interactions. However, with the INFE substrate, Asn-P3, in addition to an intrasubstrate hydrogen bond interaction involving the Glu-P5 side chain, is involved in van der Waals interactions with Q110. The S4 pocket, formed by residues M107, R108, I109, T166, and V168, resembles the S2 pocket in its hydrophobic nature, hence the evident preference for hydrophobic residues in the P4 position (Phe, Ala, Ile, and Thr). The S4 pocket exhibits correlated changes with the alterations in the S2 pocket upon substrate binding, constricting when the S2 pocket is widened and widening when the S2 pocket is constricted (Fig. 5D to F). While Thr-P4 of the TALE substrate

FIG 4 (A) TALE substrate (yellow sticks) bound to the active site of the protease. The active site residues E54, H30, and C139 are indicated in red. TALE corresponds to the P4 to P1 positions of the substrate within the polyprotein and is bound in the S1 (green sticks) and S2 (teal sticks) pockets and the oxyanion hole (blue). TALE interacts with the eII β -strand backbone (gray) of the protease, forming an antiparallel β -sheet. Hydrogen bonds that are formed between the substrate and the protease are shown as dashed lines. (B) INFE substrate (orange sticks) interactions with the protease active site (red sticks) and the substrate binding pockets (S1 in green and S2 in teal). INFE forms an antiparallel β -sheet with the eII β -strand of the protease. Hydrogen bonds are represented as dashed lines.

FIG 5 Protease molecule in surface representation, with highlighted residues in color-coded pockets. (A) Unbound structure of protease, with the S2 pocket (teal, bII-cII loop) in the closed conformation. The active site is depicted in red, the S1 pocket in green, and the S2 pocket in teal. (B) TALE-bound (yellow sticks) protease structure with color-coded pockets showing a semiclosed conformation of the S2 pocket (teal). (C) INFE-bound (orange sticks) protease structure showing an open conformation of the S2 pocket accommodating a larger P2 residue (Phe, orange). (D) The S4 pocket is shown in magenta and has a wider conformation in the absence of a substrate. (E) The S4 pocket narrows when the TALE substrate (yellow sticks) binds the pocket. (F) The conformation of the S4 pocket is narrower when the INFE substrate binds the substrate binding pocket.

makes van der Waals contacts with I109 and V168, Ile-P4 of the INFE substrate makes hydrophobic interactions only with I109, a network of hydrogen bonds involving Q110, the main chain carbonyl oxygen of R108, and a water molecule.

Substrate-based aldehyde inhibitors. Based on the structures described above, we designed and synthesized several substrate-based compounds and examined their ability to inhibit NV protease activity, as was done for some picornavirus 3C proteases (29, 42). These peptide compounds, with a protective N-terminal benzylloxycarbonyl (CBZ) group and a C-terminal aldehyde as the reactive group to allow formation of a covalent bond with the sulfhydryl group of C139, were designed to mimic the substrate residues P3 to P1 of the NV polyprotein (Table 2). The inhibition activities of these peptide-aldehydes were tested *in vitro* by using a FRET-based assay as described previously (14). To understand the structural basis of the NV protease inhibition by these compounds, the three most potent inhibitors, syc-10 ($K_i = 0.12 \mu\text{M}$), syc-59 ($K_i = 0.56 \mu\text{M}$), and syc-8 ($K_i = 1.5 \mu\text{M}$), were cocrystallized with the NV protease for crystallographic analyses (Table 2 and Fig. 6). In all of the cocrystal structures, the simulated annealing $F_o - F_c$ omit map (at a contour level of $\sim 3\sigma$) unambiguously showed the density for the inhibitor.

Inhibitor binding and covalent bond formation. These cocrystal structures revealed that all three inhibitors exhibit similar binding modes, not only between themselves but also with the substrates. The peptide backbone of the inhibitors, as with the substrates, interacts with the eII β -strand of the protease, forming an antiparallel β -sheet. One striking difference, however, is that

unlike the case in the substrate-protease structures, in which substrate binding induces flipping of the P136-G137 peptide unit to confer the correct oxyanion hole conformation, in the inhibitor-bound structures this peptide unit remains in the native orienta-

TABLE 2 Chemical structures of NV protease tripeptide aldehyde inhibitors

Inhibitor	Structure	K_i (μM) ^a
syc-10		0.12
syc-59		0.56
syc-08		1.5

^a Determined for protease in the presence of the inhibitor.

FIG 6 Rates of substrate hydrolysis (30 μ M fluorogenic peptide EDANS-EP DFHLQGPEDLAK-DabcyI) by the NV protease (0.5 μ M) at different concentrations of the three inhibitors. The K_i values were determined by fitting the initial velocities to the Morrison tight-binding equation, using Graphpad Prism 5 software.

tion, as in the unbound protease structure. In the inhibitor-bound structures, the carbonyl oxygen of the aldehyde group, similar to the terminal carbonyl oxygen in the substrates, is oriented toward the active site, making hydrogen bonds with the side chain (N ϵ 2) of H30 and a water molecule that is hydrogen bonded to S159 (Fig. 7D). In all three inhibitor-bound structures, the electron density is consistent with formation of a covalent bond between the carbon atom of the aldehyde group of the inhibitor and the sulfur atom of the nucleophilic C139 residue in the active site (Fig. 7A). The mechanism by which these aldehyde inhibitors inactivate the protease activity is likely due to this covalent bond formation, whereas the efficiency of inhibition is likely driven by how these inhibitors interact with the protease.

Protease-inhibitor interactions. The inhibitors were designed based on the NV substrates, with similar P1 to P3 residues. How do these residues in the inhibitors interact with the protease, and how do these interactions contribute, at least in part, to the differential inhibitory effects? As expected, the P1 side chain of the inhibitors inserts into the S1 pocket. The syc-10 and syc-59 inhibitors contain a glutamine derivative, *N,N*-dimethyl glutamine, at the P1 position, whereas in syc-08, the P1 side chain is an α,β -

FIG 7 Peptide aldehyde inhibitors bound to the substrate binding pockets of the NV protease. The S1 pocket is shown in green, with the residues depicted as green sticks, and the S2 pocket is shown in cyan. Hydrogen bonds are shown as dashed lines. (A to C) syc-10 (yellow ball-and-stick figure), syc-59 (magenta ball-and-stick figure), and syc-08 (pink ball-and-stick figure) interacting with the active site and binding pockets S1 (green sticks) and S2 (cyan sticks). syc-10, syc-59, and syc-08 form covalent bonds with the C139 residue of the active site (red sticks). (D) In the syc-10–protease complex, the main chain amine group of G137 and the main chain oxygen of P136 (gray sticks) flip, disrupting oxyanion hole formation. The dashed lines represent hydrogen bond formation between the syc-10 inhibitor carbonyl group and the active site H30 residue (red sticks) and a water molecule (w). (E) Surface representation of the NV protease structure S4 pocket in complex with the syc-10 inhibitor (yellow ball-and-stick figure). The pocket is occupied by the CBZ-P4 protective groups of superimposed syc-10 (yellow ball-and-stick figure), syc-59 (magenta ball-and-stick figure), and syc-08 (pink ball-and-stick figure). The CBZ groups of syc-10 and syc-59 fill the hydrophobic S4 pocket, whereas the CBZ group of syc-08 (pink) faces the solvent. (F) Surface representation of the substrate binding pocket of the NV protease–syc-08 complex structure. Yellow sticks show the C-terminal tail of the symmetry-related protease molecule in close proximity to the S4 pocket and the syc-08 CBZ group.

diamino propionic aldehyde, which is shorter and lacks the terminal dimethyl groups. In all three inhibitor-bound structures, the P1 side chain is involved in hydrogen bond interactions with H157 and T134, similar to P1-Glu of the substrates (Fig. 7). In syc-10 and syc-59, the terminal dimethyl groups in the P1 side chain make additional stabilizing van der Waals contacts with the backbone residues A159 and A160. These hydrophobic interactions are absent in the case of syc-08, as the P1 side chain lacks the dimethyl groups. Both the syc-10 and syc-08 inhibitors have Phe at the P2 position, as in the INFE substrate, whereas syc-59 has the smaller hydrophobic residue Leu at this position, as in the TALE substrate. The P2-S2 interactions of these inhibitors, including the conformational changes, are very similar to those observed in the INFE- and TALE-bound structures with P2-Phe and P2-Leu, respectively. A noticeable difference, however, is seen in regard to various side chain conformations of Q110 and R112 in the S2 pocket, particularly in the structures with P2-Phe. In the syc-10-bound structure, Q110 exhibits dual conformations, with one of them making a hydrogen bond with the backbone amide of Phe-P2 similar to that observed in the TALE structure with P2-Leu, whereas the other conformation is similar to that observed for syc-08 and the INFE substrate with P2-Phe.

As in the substrate-bound structures, the P3 residue of the inhibitors is not involved in any substantial interactions with the protease. In regard to the S4 pocket, inhibitor-bound structures also exhibit correlated changes with the S2 pocket, as observed in the substrate-bound structures. The N-terminal CBZ group of the inhibitors, which can be considered equivalent to P4, interacts with the S4 pocket. These interactions, however, vary among the three inhibitors. While the CBZ groups of syc-10 and syc-59 are tightly tucked inside the S4 pocket, making substantial van der Waals contacts through A160, T161, T166, V168, M107, and I109, in syc-08 the CBZ group is mainly solvent exposed.

C-terminal tail interactions. In the TALE and INFE structures, the C-terminal tail of the protease molecule inserts into the substrate binding site of the neighboring molecule. With the inhibitors competing for the same site, a relevant question is what conformation of the C-terminal tail exists in the inhibitor-bound structures. The C-terminal tail in the syc-59-bound structure is mostly disordered from residue 174 onward, as in the native protease structure. While in the syc-10-bound structure the C-terminal tail is ordered further, up to residue 179, in the syc-08-bound structure, it is completely ordered up to the last residue, residue 181. In the latter structure, in which the S4 pocket is not occupied by the CBZ group, in contrast to the syc-10- and syc-59-bound structures, the C-terminal tail of the neighboring molecule is positioned close to the S4 pocket, with the hydrophobic residue L180 inserting into this pocket and making hydrophobic interactions with residues I109, T161, T166, and V168 (Fig. 7F). These observations suggest that the C-terminal residues have a natural tendency to interact with the substrate binding site and that the relatively more potent syc-10 and syc-59 inhibitors compete with the C-terminal tail more effectively than the weaker syc-08 inhibitor.

DISCUSSION

The members of the *Caliciviridae* and *Picornaviridae* families encode a polyprotein that is subsequently processed into various nonstructural proteins by the viral protease, which is also carried within the polyprotein (5, 6). The proteases in these viruses cleave their respective polyproteins at multiple sites. In the studies de-

scribed here, we have provided some structural insight into how the viral protease, despite compositional changes, is able to recognize multiple sites; whether the substrates adopt the same conformation; and whether the protease undergoes substrate-induced structural alterations. Based on crystallographic analyses of the substrate-bound NV protease structures together with previous studies on picornavirus protease inhibitors (29, 42, 43), we designed substrate-based aldehyde inhibitors and determined the structures of the three most potent inhibitors in complex with the NV protease to gain further understanding of the mechanism of inhibition.

Protease induces an extended β -strand conformation in the substrate. In both TALE- and INFE-bound NV protease structures, the C-terminal tail, representing P4 to P1 in two of the NV polyprotein sites, adopts a very similar extended β -conformation, such that the side chains of the P4 to P1 residues align optimally with the S4 to S1 pockets of the protease and the cleavage site is appropriately positioned near the active site triad. In the native NV protease structure, C-terminal residues 174 to 181 are disordered. The structure of the B molecule in the TALE-bound structure discussed above, i.e., the C-terminal tail, which is only partially inserted into the substrate binding pocket of the neighboring protease molecule A, exhibits a different conformation. These observations suggest that formation of the extended β -strand conformation is induced only upon binding to the protease and that the C-terminal residues have the propensity to adopt multiple conformations. Such flexibility, which is likely a common feature in the C-terminal regions of the constituent proteins of the polyprotein, may be necessary for allowing the protease to access the cleavage sites during polyprotein processing. Indeed, this flexibility may also have been a major contributing factor to the total lack of density for VPg in our crystal structures obtained using a VPg-Pro^{C139A} construct, in addition to the possibility that VPg is intrinsically disordered. The β -strand conformation of the substrate upon binding allows it to interact with the eII β -strand of the protease to form an antiparallel β -sheet, imparting a stabilizing effect on substrate binding. In the three structures of the protease-inhibitor complexes and in the crystal structures of other viral serine-like proteases, including those of murine norovirus (26), tobacco etch virus (TEV) in the *Picornaviridae* family (44), and coronavirus (45), the peptide backbone of the substrate or the inhibitors adopts a similar extended conformation.

Cleavage specificity-determining interactions are conserved across viral serine-like proteases. In proteases of caliciviruses, picornaviruses, and coronaviruses, the P1 position of the substrate is restricted to either Glu or Gln, which interacts with the S1 pocket of the protease. In our structural analyses, the P1 position in both substrates (TALE and INFE) is occupied by Glu. In contrast, in all previous structural analyses of protease-substrate complexes from other viral systems, including other noroviruses, picornaviruses, and coronaviruses, the substrates have Gln in the P1 position (46). Comparative analysis of these structures indicates that Glu and Gln make identical sets of interactions and that these interactions, involving His, Thr, and Tyr in the S1 pocket, are highly conserved. In all these proteases, the Glu or Gln forms hydrogen bond interactions with the side chains of His and Thr, and in most cases the side chain of the His residue is positioned appropriately by its interactions with a Tyr residue. Our inhibitor-bound protease structures and previous norovirus inhibitor-protease complex structures, each with different Gln derivatives in the

P1 position, show the same set of interactions in the S1 pocket. The critical nature of His, Thr, and Tyr residues in the S1 pocket in the case of noroviruses was confirmed by previous mutational analysis of the protease of Chiba virus, which showed that alanine substitution of these residues abolished the catalytic activity of the protease (47). These observations suggest the possibility of designing broad-spectrum substrate-based protease inhibitors with Gln or Glu derivatives in the P1 position for some of these viruses (48).

The terminal carboxylate group of the P1 residue triggers formation of an oxyanion hole. A striking observation from the structural analysis of the TALE- and INFE-bound NV proteases was the formation of an oxyanion hole triggered by the terminal carboxyl group of the P1 residue. In both of these structures, the peptide unit linking Pro136 and G137 is flipped compared to the native unbound structure, such that the amide group forms a hydrogen bond with the terminal carbonyl oxygen (anion) of the P1 residue, thus mimicking the stabilization of the tetrahedral intermediate during peptide hydrolysis. In the native protease structure, the same peptide unit is oriented such that its amide group (of G137) points in the opposite direction and is exposed to the solvent, while the carbonyl group (of Pro136) is positioned inward toward the active site, participating in the hydrogen bond interaction with the catalytic residue H30. The possibility that flipping is due to the C139A mutation (in the substrate-bound structures) can be ruled out because in molecule A of the TALE-bound crystal structure with a partially inserted C-terminal tail, despite the C139A mutation, the peptide orientation remains the same as in the native structure. Thus, the positioning of the anionic carbonyl oxygen in the oxyanion hole triggers the flipping of the peptide unit. Superposing the substrate-bound structure with the native protease structure clearly shows that this oxygen atom clashes with the carbonyl group of Pro136 in the native structure, without peptide flipping. When the C139A mutant residue is computationally reverted to cysteine, the sulfhydryl group is appropriately positioned for nucleophilic attack of the carbon atom of the terminal carboxylate group. Such a ligand-induced peptide flipping is also observed in Chiba virus protease (27), although in this case, the ligand is the tartrate molecule, which is suggested to behave like a substrate.

In all of our inhibitor-bound protease structures, however, the orientation of the peptide unit remains the same as in the native protease structure. As a result of the formation of the covalent bond between the carbon atom of the aldehyde and the sulfur atom of C139, and because of the oxygen atom of the aldehyde pointing toward the catalytic H30 residue, it is likely that there is not a sufficient negative charge to induce flipping of the peptide. With the availability of the unbound structure and with a structure of the B molecule with a partially inserted C-terminal tail serving as an internal control, we were able to carry out a systematic comparative analysis to infer the effect of substrate binding on the oxyanion hole conformation. Whether this observation is unique to noroviruses or a common feature in other viral serine-like proteases cannot readily be addressed, as such systematic analyses with unbound and corresponding substrate-bound structures are not available for other serine-like proteases. For some of these proteases, a proper oxyanion hole conformation with the amide group correctly pointing toward the substrate binding site is observed in the unbound structure itself; however, whether the crystallization conditions played a role in “peptide flipping” is unclear. In the native protease structure of foot-and-

mouth disease virus, for example, the oxyanion hole is properly configured, perhaps because of the presence of a sulfate ion in close proximity to the oxyanion hole.

The S2 pocket undergoes P2-dependent conformational changes. In contrast to the P1 residue, which is constrained to be either Glu or Gln, the P2 and P4 residues in the various cleavage sites of the NV polyprotein exhibit significant variations. Although we could not determine the structure of the NV protease with all five substrates, the two substrate-bound structures in comparison with the unbound native protease structure clearly show the inherent plasticity in the substrate binding sites to accommodate changes in the P2 and P4 positions. A particularly striking observation was the alteration in the S2 pocket, which exhibits three distinct states: a closed state as in the unbound structure, a slightly open state with a P2-Leu residue, and a more open state with a bulkier P2-Phe residue. These changes in the overall size of the S2 pocket are caused mainly by the inherent flexibility of the loop (bII-cII), which constitutes a major portion of this pocket. In addition, some of the residues in the S2 pocket, such as R112 and Q110, exhibit changes in their side chain orientations to further optimize interactions with the P2 residue. Similar changes in the S2 pocket are also observed in the inhibitor-bound structures with P2-Leu (syc-59) and P2-Phe (syc-10 and syc-8). Although the flexibility of the S2 pocket in association with substrate binding was previously suggested (47), our comparative analysis of the same protease with and without bound ligands provides concrete evidence for how the S2 pocket is modulated in response to variations in the P2 residue.

The S2 and S4 pockets exhibit coordinated changes. A novel observation from our structural analysis is that the substrate-induced changes in the S4 pocket are correlated with alterations in the S2 pocket. A structural basis for such coordinated changes is evident by considering that these two pockets share the bII-cII loop formed by residues 107 to 114. The movement of this loop in response to the binding of the P2 residue affects the size of the S4 pocket, widening when the S2 pocket constricts and vice versa. Furthermore, this movement also differentially affects which residue in the loop interacts with P2 or P4. Interestingly, in both substrate-bound structures, I109 makes hydrophobic contacts with both the P2 and P4 residues, providing a structural rationale for the previous observation that mutation of I109 affects the binding of both P2 and P4 residues (41). These concerted changes in the S2 and S4 pockets suggest that substrate recognition and binding affinity are synergistically affected by both the P2 and P4 residues. Accordingly, the correlation between the nature of the P2 and P4 residues is evident, to a certain extent, in the five protease binding sites in the NV polyprotein. The first two sites in the polyprotein contain P2-Leu and P4-Phe (Fig. 1). Based on the TALE-bound structure with P2-Leu, we expect that the semiopen S2 pocket and a wider S4 pocket would be appropriate to accommodate bulkier P4-Phe residues. Modeling of a Phe residue in place of Thr in the TALE-bound structure indicates that Phe fits snugly into the S4 pocket, without steric clashes, and makes substantially more hydrophobic interactions with the residues in the S4 pocket, which may partially account for the observed temporal sequence in NV polyprotein processing, with the first two sites cleaved earlier than the last three sites (17).

In contrast to the well-defined S1, S2, and S4 pockets, the S3 pocket is not as well defined. The side chain of the P3 residue in each of the substrates and the inhibitors is oriented mostly toward

the bulk solvent, clearly indicating that this residue contributes minimally toward binding specificity. This is consistent with the significant divergence observed at the P3 position in the five substrate sites in the NV polypeptide. However, the backbone atoms of the residues in the S3 pocket may contribute to substrate binding affinity because of their involvement in forming the antiparallel β -sheet with the substrate backbone.

What is the structural basis for the different efficiencies of the inhibitors? Comparative structural analysis of the three inhibitor-bound structures showed that while these inhibitors exhibit several similarities in the way that they interact with the protease, there are noticeable differences that may account for their differential potency. The *syc-10* inhibitor is the most potent of the three and makes the most contacts with the protease. First, the Gln derivative in the P1 position interacts extensively with the S1 pocket, making additional hydrophobic contacts due to the presence of a dimethyl group; second, P2-Phe is fully engaged in the S2 pocket; and third, the protective group, as a surrogate P4 residue, fits into the S4 pocket with several stabilizing hydrophobic interactions. In contrast, the second most potent inhibitor, *syc-59*, exhibits identical interactions with the S1 pocket due to the same Gln derivative in the P1 position, but the interactions in the S4 pocket are not as substantial because of the widening of the S4 pocket due to P2-Leu interactions in the S2 pocket. With the third inhibitor, *syc-08*, with the absence of the dimethyl group, the P1 residue makes fewer contacts, and although P2-Phe exhibits similar interactions to those with *syc-10*, the CBZ group is moved away from the S4 pocket and does not have any substantial interactions with the protease. These observations suggest that maximizing the interactions in the S1, S2, and S4 pockets has a direct impact on the potency of the inhibitors.

Previous studies on norovirus proteases have indicated the possibility of designing inhibitors with different reactive groups, such as bisulfite and ethyl ester (28, 48). In our studies, which included other reactive groups, including esters and ketones, the most potent inhibitors of NV protease were those with an aldehyde as the reactive group. Interestingly, in recent studies of NV protease, the K_i of an inhibitor with P2-Leu and bisulfite as the reactive group, which during the reaction was converted into an aldehyde, was similar to the K_i of $\sim 0.6 \mu\text{M}$ for *syc-59* with P2-Leu (48). We have shown that replacing P2-Leu with P2-Phe (*syc-10*) increases the potency 2-fold. Previous studies with the Southampton protease suggested that substrate-based inhibitors with ethyl ester (P2-Leu) as the reactive group can also be effective; however, the K_i for this inhibitor was not reported for comparison (28). All of these studies underscore the possibility of designing more potent inhibitors for norovirus protease with further optimization of the P1 to P4 residues and the reactive groups. It is also clear that further studies are required not only for designing more potent inhibitors but also to ensure their viability for humans in the context of virus infection. In that regard, the structural insights we have provided here, through a systematic analysis of protease-substrate and protease-inhibitor complexes, will be useful not only for noroviruses but also for other serine-like viral proteases.

Conclusions. Because of their critical involvement in replication processes, the viral serine-like cysteine proteases in virus families such as the *Caliciviridae*, *Picornaviridae*, and *Coronaviridae* families, all of which consist of significant human pathogens, are attractive targets for designing antiviral compounds. A common feature of these proteases is their ability to recognize multiple

substrates with varied sequences flanking the scissile bond. Here we provided a systematic comparative crystallographic analysis of the NV protease in complex with two native substrates and three substrate-based aldehyde inhibitors. From these studies, we have made several novel observations, including the following: (i) the C-terminal residues of the constituent proteins of the polypeptide may have a propensity to adopt multiple conformations, and the extended β -strand conformation is induced only upon binding to the protease; (ii) interactions that determine cleavage specificity are possibly conserved across viral serine-like proteases; (iii) the terminal carboxylate group of the P1 residue triggers formation of an oxyanion hole; and (iv) the S2 and S4 pockets exhibit concerted alterations in response to substrate binding. From our structural studies on protease-inhibitor complexes, we have shown that the mechanism of inhibition involves covalent bond formation with the sulfhydryl of the conserved cysteine in the active site by nucleophilic addition, and such inhibitors could be effective protease inhibitors. From our comparative analysis, we have suggested a possible structural basis for the differential potency of these peptidyl aldehyde inhibitors. In addition to furthering our understanding of norovirus protease-substrate interactions, our studies may also be applicable to other viral serine-like proteases and will be useful in the future design of virus-specific or broad-spectrum protease inhibitors as therapeutic agents.

ACKNOWLEDGMENTS

This work was supported by grants from the NIH (PO1 AI057788 to M.K.E., T.P., Y.S., and B.V.V.P. and P30DK5638 to M.K.E.), the Robert Welch Foundation (Q1279 to B.V.V.P.), and the Caroline Wiess Law Fund for Molecular Medicine (to Y.S.).

We acknowledge the use of the synchrotron beamlines at Advanced Light Source (5.0.1), Berkeley, CA, and Argonne National Laboratories (SBC-CAT 19ID), Chicago, IL, for diffraction data collection. We thank the staff of those institutions for excellent help.

The SBC-CAT 19ID beamline at Advanced Photon Source is supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under contract W-31-109-Eng-38. The Berkeley Center for Structural Biology is supported in part by the National Institutes of Health, National Institute of General Medical Sciences, and the Howard Hughes Medical Institute. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under contract no. DE-AC02-05CH11231.

REFERENCES

- Green KY. 2007. Caliciviridae: the noroviruses, p 949–979. In Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA, Roizman B, Straus SE (ed), *Fields virology*, 5th ed. Lippincott Williams & Wilkins, Philadelphia, PA.
- Zheng DP, Ando T, Fankhauser RL, Beard RS, Glass RI, Monroe SS. 2006. Norovirus classification and proposed strain nomenclature. *Virology* 346:312–323.
- Hardy ME, Estes MK. 1996. Completion of the Norwalk virus genome sequence. *Virus Genes* 12:287–290.
- Jiang X, Wang M, Wang K, Estes MK. 1993. Sequence and genomic organization of Norwalk virus. *Virology* 195:51–61.
- Belliot G, Sosnovtsev SV, Mitra T, Hammer C, Garfield M, Green KY. 2003. In vitro proteolytic processing of the MD145 norovirus ORF1 nonstructural polypeptide yields stable precursors and products similar to those detected in calicivirus-infected cells. *J. Virol.* 77:10957–10974.
- Blakeney SJ, Cahill A, Reilly PA. 2003. Processing of Norwalk virus nonstructural proteins by a 3C-like cysteine proteinase. *Virology* 308:216–224.
- Jiang X, Wang M, Graham DY, Estes MK. 1992. Expression, self-assembly, and antigenicity of the Norwalk virus capsid protein. *J. Virol.* 66:6527–6532.

8. Glass PJ, White LJ, Ball JM, Lepar-Goffart I, Hardy ME, Estes MK. 2000. Norwalk virus open reading frame 3 encodes a minor structural protein. *J. Virol.* 74:6581–6591.
9. Ettayebi K, Hardy ME. 2003. Norwalk virus nonstructural protein p48 forms a complex with the SNARE regulator VAP-A and prevents cell surface expression of vesicular stomatitis virus G protein. *J. Virol.* 77:11790–11797.
10. Pfister T, Wimmer E. 2001. Polypeptide p41 of a Norwalk-like virus is a nucleic acid-independent nucleoside triphosphatase. *J. Virol.* 75:1611–1619.
11. Sharp TM, Guix S, Katayama K, Crawford SE, Estes MK. 2010. Inhibition of cellular protein secretion by Norwalk virus nonstructural protein p22 requires a mimic of an endoplasmic reticulum export signal. *PLoS One* 5:e13130. doi:10.1371/journal.pone.0013130.
12. Daughenbaugh KF, Fraser CS, Hershey JW, Hardy ME. 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.
13. Someya Y, Takeda N, Miyamura T. 2005. Characterization of the norovirus 3C-like protease. *Virus Res.* 110:91–97.
14. Zeitler CE, Estes MK, Venkataram Prasad BV. 2006. X-ray crystallographic structure of the Norwalk virus protease at 1.5-Å resolution. *J. Virol.* 80:5050–5058.
15. Ng KK, Pendas-Franco N, Rojo J, Boga JA, Machin A, Alonso JM, Parra F. 2004. Crystal structure of Norwalk virus polymerase reveals the carboxyl terminus in the active site cleft. *J. Biol. Chem.* 279:16638–16645.
16. Zamyatkin DF, Parra F, Alonso JM, Harki DA, Peterson BR, Grochulski P, Ng KK. 2008. Structural insights into mechanisms of catalysis and inhibition in Norwalk virus polymerase. *J. Biol. Chem.* 283:7705–7712.
17. Hardy ME, Crone TJ, Brower JE, Ettayebi K. 2002. Substrate specificity of the Norwalk virus 3C-like proteinase. *Virus Res.* 89:29–39.
18. Sosnovtsev SV, Garfield M, Green KY. 2002. Processing map and essential cleavage sites of the nonstructural polyprotein encoded by ORF1 of the feline calicivirus genome. *J. Virol.* 76:7060–7072.
19. Bazan JF, Fletterick RJ. 1988. Viral cysteine proteases are homologous to the trypsin-like family of serine proteases: structural and functional implications. *Proc. Natl. Acad. Sci. U. S. A.* 85:7872–7876.
20. Boniotti B, Wirblich C, Sibilia M, Meyers G, Thiel HJ, Rossi C. 1994. Identification and characterization of a 3C-like protease from rabbit hemorrhagic disease virus, a calicivirus. *J. Virol.* 68:6487–6495.
21. Liu B, Clarke IN, Lambden PR. 1996. Polyprotein processing in Southampton virus: identification of 3C-like protease cleavage sites by *in vitro* mutagenesis. *J. Virol.* 70:2605–2610.
22. Liu BL, Viljoen GJ, Clarke IN, Lambden PR. 1999. Identification of further proteolytic cleavage sites in the Southampton calicivirus polyprotein by expression of the viral protease in *E. coli*. *J. Gen. Virol.* 80:291–296.
23. Seah EL, Marshall JA, Wright PJ. 1999. Open reading frame 1 of the Norwalk-like virus Camberwell: completion of sequence and expression in mammalian cells. *J. Virol.* 73:10531–10535.
24. Someya Y, Takeda N, Miyamura T. 2000. Complete nucleotide sequence of the Chiba virus genome and functional expression of the 3C-like protease in *Escherichia coli*. *Virology* 278:490–500.
25. Scheffler U, Rudolph W, Gebhardt J, Rohayem J. 2007. Differential cleavage of the norovirus polyprotein precursor by two active forms of the viral protease. *J. Gen. Virol.* 88:2013–2018.
26. Leen EN, Baeza G, Curry S. 2012. Structure of a murine norovirus NS6 protease-product complex revealed by adventitious crystallisation. *PLoS One* 7:e38723. doi:10.1371/journal.pone.0038723.
27. Nakamura K, Someya Y, Kumasaka T, Ueno G, Yamamoto M, Sato T, Takeda N, Miyamura T, Tanaka N. 2005. A norovirus protease structure provides insights into active and substrate binding site integrity. *J. Virol.* 79:13685–13693.
28. Hussey RJ, Coates L, Gill RS, Erskine PT, Coker SF, Mitchell E, Cooper JB, Wood S, Broadbridge R, Clarke IN, Lambden PR, Shoolingin-Jordan PM. 2011. A structural study of norovirus 3C protease specificity: binding of a designed active site-directed peptide inhibitor. *Biochemistry* 50:240–249.
29. Matthews DA, Dragovich PS, Webber SE, Fuhrman SA, Patick AK, Zalman LS, Hendrickson TF, Love RA, Prins TJ, Marakovits JT, Zhou R, Tikhe J, Ford CE, Meador JW, Ferre RA, Brown EL, Binford SL, Brothers MA, DeLisle DM, Worland ST. 1999. Structure-assisted design of mechanism-based irreversible inhibitors of human rhinovirus 3C protease with potent antiviral activity against multiple rhinovirus serotypes. *Proc. Natl. Acad. Sci. U. S. A.* 96:11000–11007.
30. Copeland RA. 2002. *Enzymes*, 2nd ed. Wiley-VCH, New York, NY.
31. Otwinowski Z. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276:307–326.
32. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007. Phaser crystallographic software. *J. Appl. Crystallogr.* 40:658–674.
33. Collaborative Computational Project. 1994. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 50:760–763.
34. Morris RJ, Perrakis A, Lamzin VS. 2003. ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol.* 374:229–244.
35. Emsley P, Cowtan K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60:2126–2132.
36. Murshudov GN, Vagin AA, Dodson EJ. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* 53:240–255.
37. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 66:213–221.
38. Painter J, Merritt EA. 2006. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D Biol. Crystallogr.* 62:439–450.
39. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. 1996. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* 8:477–486.
40. Schrodinger, LLC. 2011. The PyMOL molecular graphics system, 1.5.0.4. Schrodinger, LLC, Portland, OR.
41. Someya Y, Takeda N. 2009. Insights into the enzyme-substrate interaction in the norovirus 3C-like protease. *J. Biochem.* 146:509–521.
42. Webber SE, Okano K, Little TL, Reich SH, Xin Y, Fuhrman SA, Matthews DA, Love RA, Hendrickson TF, Patick AK, Meador JW, 3rd, Ferre RA, Brown EL, Ford CE, Binford SL, Worland ST. 1998. Tripeptide aldehyde inhibitors of human rhinovirus 3C protease: design, synthesis, biological evaluation, and cocrystal structure solution of P1 glutamine isosteric replacements. *J. Med. Chem.* 41:2786–2805.
43. Dragovich PS, Webber SE, Babine RE, Fuhrman SA, Patick AK, Matthews DA, Reich SH, Marakovits JT, Prins TJ, Zhou R, Tikhe J, Littlefield ES, Bleckman TM, Wallace MB, Little TL, Ford CE, Meador JW, 3rd, Ferre RA, Brown EL, Binford SL, DeLisle DM, Worland ST. 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.
44. Phan J, Zdanov A, Evdokimov AG, Tropea JE, Peters HK, 3rd, Kaput RB, Li M, Wlodawer A, Waugh DS. 2002. Structural basis for the substrate specificity of tobacco etch virus protease. *J. Biol. Chem.* 277:50564–50572.
45. Goetz DH, Choe Y, Hansell E, Chen YT, McDowell M, Jonsson CB, Roush WR, McKerrow J, Craik CS. 2007. Substrate specificity profiling and identification of a new class of inhibitor for the major protease of the SARS coronavirus. *Biochemistry* 46:8744–8752.
46. Perona JJ, Craik CS. 1995. Structural basis of substrate specificity in the serine proteases. *Protein Sci.* 4:337–360.
47. Someya Y, Takeda N. 2011. Functional consequences of mutational analysis of norovirus protease. *FEBS Lett.* 585:369–374.
48. Kim Y, Lovell S, Tiew KC, Mandadapu SR, Alliston KR, Battaile KP, Groutas WC, Chang KO. 2012. Broad-spectrum antivirals against 3C or 3C-like proteases of picornaviruses, noroviruses, and coronaviruses. *J. Virol.* 86:11754–11762.