BioMol Concepts, Vol. 3 (2012), pp. 41–56 • Copyright © by Walter de Gruyter • Berlin • Boston. DOI 10.1515/BMC.2011.052

Review

From head to toe of the norovirus 3C-like protease

Yuichi Someya

Department of Virology II, National Institute of Infectious Diseases, 4-7-1 Gakuen, Musashi-Murayama, Tokyo 208-0011, Japan

e-mail: someya@nih.go.jp

Abstract

Noroviruses are major causative agents of viral gastroenteritis in humans. Currently, there are no therapeutic medications to treat noroviral infections, nor are there effective vaccines against these pathogens. The viral 3C-like protease is solely responsible for the maturation of viral protein components. The crystal structures of the proteases were resolved at high atomic resolution. The protease was also explored by means of mutagenesis. These studies revealed the active-site amino acid residues and factors determining and affecting substrate specificity as well as the principle of architecting the protease molecule. The possible mechanism of proteolysis was also suggested. Consideration of the data accumulated thus far will be useful for development of therapeutic drugs targeting the viral protease.

Keywords: crystal structure; mutagenesis; norovirus; protease; proteolysis.

Introduction

Norovirus is a major causative agent of acute, non-bacterial gastroenteritis in humans (1–3). Growing numbers of various norovirus strains are being isolated worldwide. Norovirus belongs to the family Caliciviridae and is a positive-sense, single-stranded RNA virus. The RNA genome is ~7.7 kb in length with a poly(A) tail at the 3′-end (Figure 1A). A genome-linked viral protein, VPg, is likely to bind to the 5′-end in place of the cap structure (4). Human noroviruses are still non-cultivatable in cell culture and experimental animals, which has hampered the virological and molecular biological characterization of this group of viruses.

The genome encodes three open reading frames (ORFs) (Figure 1A). The ORF1 product is a polyprotein and is cleaved by its viral 3C-like protease activity into six non-structural proteins including 2C-like NTPase, 3B VPg, and 3D RNA-dependent RNA polymerase, in addition to the 3C-like protease (Figure 1B). The ORF2 and ORF3 products are a major and a minor structural protein (VP1 and VP2), respectively. The norovirus virion consists of 180 VP1 molecules, which includes the VPg-linked RNA genome and VP2 molecules (5). The P2 (protruding) domain of the VP1 proteins from various norovirus strains exhibits a wide variety of amino acid sequences (6, 7). Based on the sequence diversity of this region, human noroviruses are divided into two major genogroups, genogroups I (GI) and II (GII), and nearly 20 genotypes have been identified in the respective genogroups (8). This genetic diversity is correlated with a wide variety of antigenicity among noroviruses and potentially provides different patterns of binding to histo-blood group antigens (HBGAs) (9). Although HBGAs are thought to be a part of the norovirus receptor and evidence of their interaction with VP1 proteins has been provided by structural studies, the specific roles played by HBGAs in infection and pathogenesis are still unknown.

Norovirus 3C-like protease is a central enzyme that is solely responsible for the maturation of the ORF1 polyprotein (Figure 1B), and it is therefore a potent therapeutic target for the treatment of norovirus gastroenteritis (10, 11). The nomenclature of 3C-like protease is derived from that of the evolutionally related picornavirus 3C proteases. These proteases are members of the chymotrypsin-like serine protease superfamily, except that their active site is a Cys residue instead of a Ser residue, which accounts for their designation as serine-like cysteine proteases (12–16). Sequence homology has enabled the identification of various 3C-like proteases in other viruses, including coronavirus and potyvirus.

For approximately 10 years, this author has been engaged in structure-function studies of norovirus proteins, especially the 3C-like protease. This review provides a summary of the findings to date regarding the structural and functional aspects of the 3C-like protease of the norovirus.

Crystal structures of norovirus 3C-like proteases

Our group was the first to characterize the crystal structure of the 3C-like protease from the Chiba strain of norovirus (GI/4; genogroup I, genotype 4) (17). The overall structure resembles that of chymotrypsin (Figure 2A). The active site is located in a cleft between the N- and the C-terminal domains, both of which adopt a barrel-like fold comprised of several β-sheets. The crystal structure indicated that the active-site catalytic triad, Cys139-His30-Glu54, participates in catalysis, in a manner similar to that of the active-site catalytic triad of chymotrypsin, Ser195-His57-Asp102 (18).

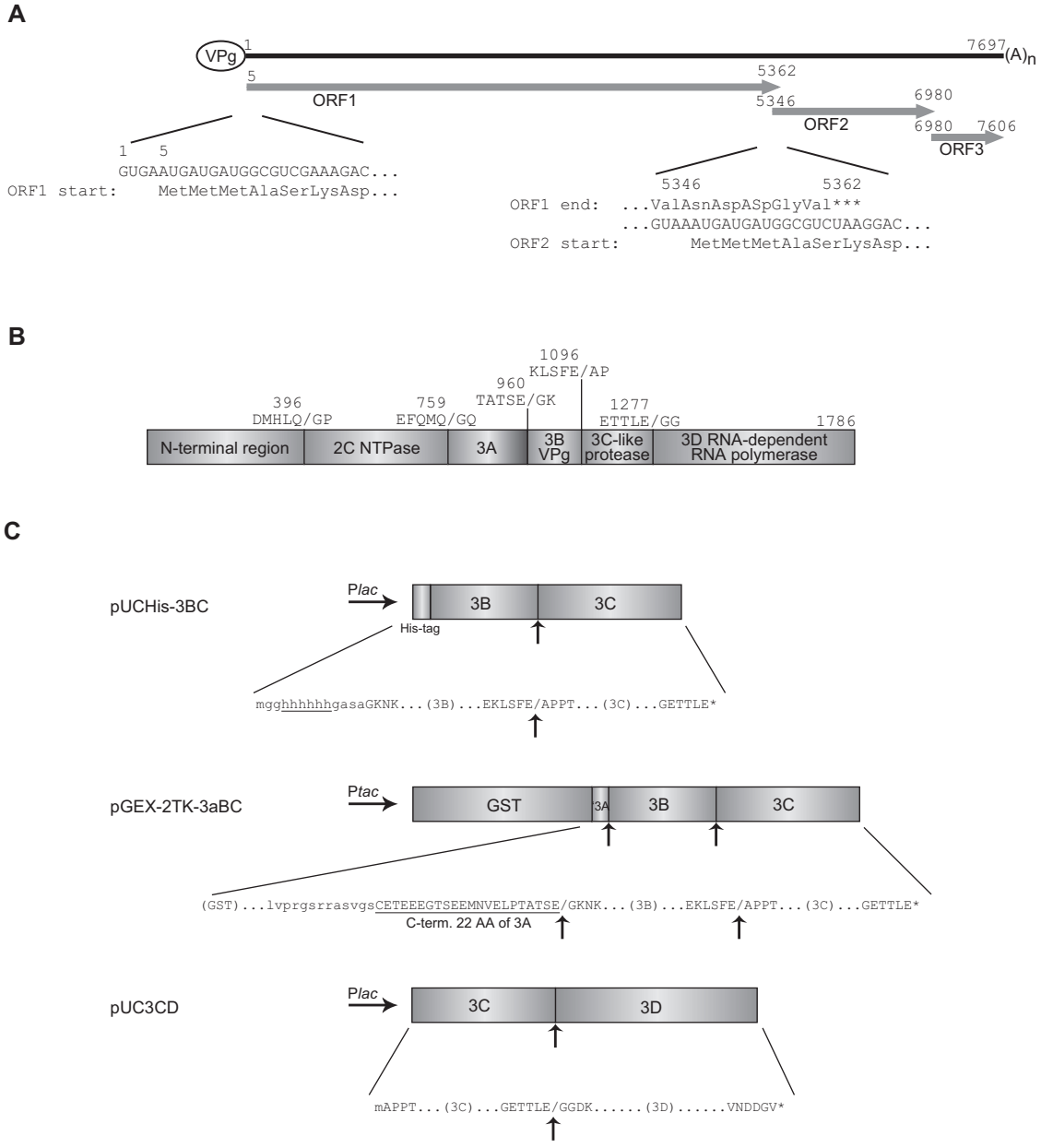

Figure 1 Schematic representation of the norovirus genome and plasmid constructs used for expression in *Escherichia coli* cells.
(A) Genomic organization of the norovirus Chiba strain (GenBank/EMBL/DDBJ ID: AB042808). The RNA genome is 7697 bases in length with its 3′-terminus polyadenylated. The viral protein, VPg, is thought to be covalently linked to its 5′-end. The first 25 nucleotides of the genome are highly homologous to the sequence around the start codon of ORF2. Parts of amino acid sequences of ORF1 and ORF2 are also shown. The end of ORF1 is partly overlapped with the start of ORF2. The numbers indicate the positions of the bases. (B) Proteolytic cleavage sites found in the ORF1 polyprotein. The numbers indicate the positions of the amino acids. Slashes indicate the cleavage sites recognized by the 3C-like protease. (C) Bacterial expression plasmids frequently used for detection of the proteolytic activity. pUCHis-3BC was used for mutagenesis studies of charged amino acid residues. pGEX-2TK-3aBC was used for the saturation mutagenesis of Glu54, Ala-scanning mutagenesis of neutral residues, and the C-terminal truncation study. pUC3CD was used for examination of the N-terminal truncation. The upward-pointing arrows indicate scissile bonds.

The 3C-like proteases from the Norwalk strain (GI/1) (19) and the Southampton strain (GI/2) (20) were also crystallized and their structures are available (Table 1). It is of note that the structures of these two proteins were obtained at higher resolution than that of the Chiba protease. Moreover, the Southampton protease was crystallized such that a peptide inhibitor was covalently bound at the active site (20); the crystal structure then enabled characterization of how substrates and products interact with protease molecules at the binding site.

Comparison of structures from these three strains (Chiba, Norwalk, and Southampton) revealed that a loop connecting

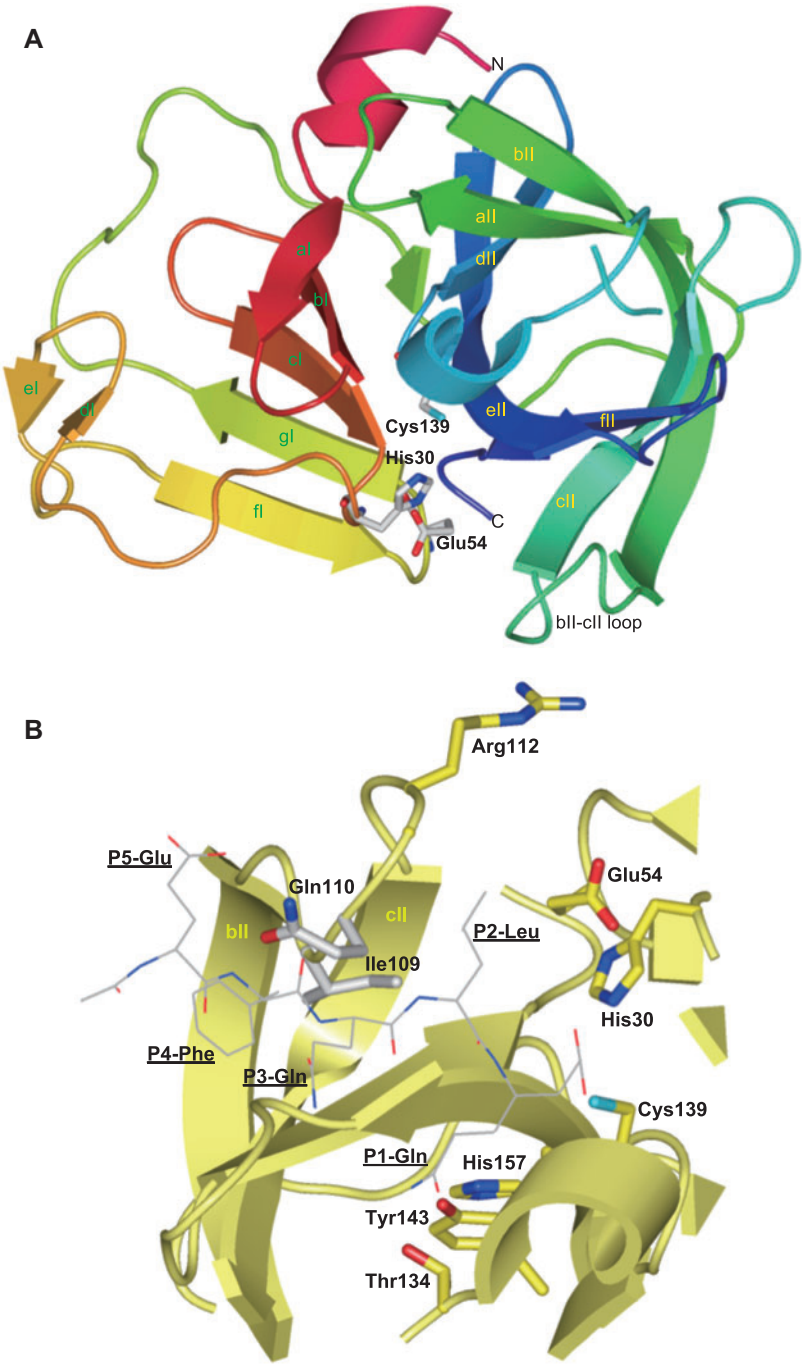

Figure 2 Crystal structure of the 3C-like protease from the norovirus Chiba strain.
(A) Overall structure (PDB ID 1WQS). A catalytic triad, Cys139-His30-Glu54, in the active site is situated in a cleft between the N- and C-terminal domains. Side chains of these three residues are represented as sticks. The names for each β-sheet are indicated. The bII-cII loop stands for a loop region connecting the bII and cII β-sheets. (B) Model of inhibitor binding. A peptide inhibitor that was included in the crystal structure of the Southampton protease (PDB ID 2IPH) was incorporated into the active site of the Chiba protease using Waals software (Altif Laboratories, Tokyo, Japan). The inhibitor is represented as a wire frame. His157, along with Thr134 and Tyr143, is involved in recognition of the side chain at the P1 position. Ile109, Gln110, and Arg112 are located in the bII-cII loop and were targets of mutagenesis as described in the main text. Ile109 is likely to affect the recognition of both P2 and P4 residues.

the bII and cII β-sheets is structurally highly flexible. This bII-cII loop is close to the active site, and is likely to play a role in facilitating substrate binding (Figure 2). The crystal structures of norovirus proteases suggest that the loop undergoes a dynamic conformational change between open and closed states during catalysis. The corresponding region of picornavirus 3C proteases, a β-ribbon, is suggested to be flexible and involved in substrate binding (34, 37).

Table 1 Data for three-dimensional structures of the 3C(-like) proteases from noroviruses and related viruses.

Virus	PDB ID	Resolution (Å)	Active site residues	Relevant protease mutation, substrate, inhibitor, etc.	References
Calicivirus					
Norovirus					
Chiba strain	1WQS	2.8	Cys139-His30-Glu54		(17)
Norwalk strain	2FYQ	1.5	Cys139-His30-Glu54		(19)
	2FYR	2.2		4-(2-Aminoethyl)-benzenesulfonyl fluoride	(19)
Southampton strain	2IPH	1.7	Cys139-His30-Glu54	Michael acceptor peptidyl inhibitor	(20)
Picornavirus					
Poliovirus	1L1N	2.1	Cys147-His40-Glu71		(21)
Poliovirus – 3CD precursor	2IJD	3.4	Cys147-His40-Glu71	C147A active site mutant	(22)
Rhinovirus serotype 14	N/A	2.3	Cys147-His40-Glu71		(23)
	2B0F	N/A (NMR)		Substrate-based ethylpropionate inhibitor	(24)
	2IN2	N/A (NMR)		Substrate-based ethylpropionate inhibitor	(24)
Rhinovirus serotype 2	1CQQ	1.85	Cys147-His40-Glu71	Ruprintrivir	(25)
	2XYA	2.4		2-Phenylquinolin-4-ol	(26)
Rhinovirus serotype 2 – 2A protease	2HRV	1.95	Cys106-His18-Asp35		(27)
Hepatitis A virus	N/A	2.3	Cys172-His44-Asp84		(28)
	1HAV	2.0			(29)
	1QA7	1.9		Iodoacetyl-valyl-phenylalanyl-amide	(30)
	2CXV	1.4		*N*-Benzoyloxycarbonyl-L-serine-β-lactone, Iodoacetyl-valyl-phenylalanyl-amide	(31)
	2A4O	1.55		*N*-Benzoyloxycarbonyl-L-serine-β-lactone	(31)
	2HAL	1.35		*N*-[(Benzoyloxy)carbonyl]-L-alanine	(32)
	2H9H	1.39		*N*-[(Benzoyloxy)carbonyl]-L-alanine	(32)
	2H6M	1.4		*N*-[(Benzoyloxy)carbonyl]-L-alanine, (4*S*)-4-Amino-6-fluoro-*N*, *N*-dimethyl-5-oxohexanamide	(32)
Foot and mouth disease virus	2BHG	1.9	Cys163-His46-Asp84	C163A active site mutant	(33)
	2J92	2.2		C163A active site mutant	(34)
	2WV4	2.5		C163A active site mutant, VP1-2A peptide substrate	(35)
	2WV5	2.7		C163A active site mutant, VP1-2Am peptide substrate	(35)
Coxsackievirus B3	2VB0	2.4	Cys147-His40-Glu71	*S*,*S*-(2-Hydroxyethyl)thiocysteine	N/A
	2ZTX	1.72		*N*-Ethyl-*N*-phenyldithiocarbamate	(36)
	2ZTY	1.72		Form II	(36)
	2ZTZ	2.0		Form I	(36)
	2ZU1	1.38		C147A active site mutant	(36)
	2ZU3	1.75		Peptidomimetic inhibitor TG-0204998	(36)
Enterovirus 71	3OSY	3.0	Cys147-His40-Glu71		(37)
Potyvirus					
Tobacco etch virus	1LVB	2.2	Cys151-His46-Asp81	C151A active site mutant	(38)
	1LVM	1.8			(38)
	1Q31	2.7		C151A active site mutant	(39)

Table 1 (Continued)

Virus	PDB ID	Resolution (Å)	Active site residues	Relevant protease mutation, substrate, inhibitor, etc.	References
Tobacco vein mottling virus	3MMG	1.70	Cys151-His46-Asp81	C151A active site mutant, peptide substrate	(40)
Coronavirus					
Human coronavirus 229E	1P9S	2.54	Cys144-His41		(41)
	2ZU2	1.8		*N*-Ethyl-*N*-phenyldithiocarbamate	(36)
SARS coronavirus	1Q2W	1.86	Cys144-His41		N/A
	1UJ1	1.9		At pH 6.0	(42)
	1UK2	2.2		At pH 8.0	(42)
	1UK3	2.4		At pH 7.6	(42)
	1UK4	2.5		Substrate analog chloromethyl ketone inhibitor	(42)
	1WOF	2.0		Substrate analog chloromethyl ketone inhibitor	(43)
	2AMQ	2.3		Substrate analog chloromethyl ketone inhibitor	(43)
	2AMD	1.85		Substrate analog chloromethyl ketone inhibitor	(43)
	2D2D	2.7		Substrate analog chloromethyl ketone inhibitor	(43)
	1Z1I	2.8			(44)
	1Z1J	2.8		C145A active site mutant	(44)
	2ALV	1.9		Ruprintrivir analog	(45)
	2A5A	2.08			(46)
	2A5I	1.88		Substrate-like aza-peptide epoxide	(46)
	2A5K	2.3		Substrate-like aza-peptide epoxide	(47)
	2BX3	2.0		At pH 5.9	(47)
	2BX4	2.79		At pH 6.6	(47)
	2C3S	1.9			(48)
	2GZ7	1.86		2-[(2,4-Dichloro-5-methylphenyl)sulfonyl]-1,3-dinitro-5 -(trifluoromethyl)benzene	(49)
	2GZ8	1.97		*S*-[5-(Trifluoromethyl)-4H-1,2,4-triazol-3-yl] 5-(phenylethynyl)furan-2-carbothioate	(49)
	2GZ9	2.17			(49)
	2GT7	1.82			(50)
	2GT8	2.0			(50)
	2GTB	2.0		Substrate-like aza-peptide epoxide	(50)
	2H2Z	1.6			(51)
	2HOB	1.95		Substrate-based irreversible inhibitor	(51)
	2GX4	1.93		Non-covalently linked inhibitor	N/A
	2Z3C	1.79		Substrate-based irreversible inhibitor	(52)
	2Z3D	2.1		Substrate-based irreversible inhibitor	(52)
	2Z3E	2.32		Substrate-based irreversible inhibitor	(52)
	2OP9	1.8		Epoxyketone inhibitor	(53)
	2DUC	1.7			N/A
	2Z94	1.78		Toluene-3,4-dithiolato zinc	(54)
	2Z9G	1.86		Phenylmercuric acetate	(54)
	2Z9J	1.95		*N*-Ethyl-*N*-phenyldithiocarbamic acid zinc	(54)

Table 1 (Continued)

Virus	PDB ID	Resolution (Å)	Active site residues	Relevant protease mutation, substrate, inhibitor, etc.	References
	2Z9K	1.85		(Nitrilotriacetato-*N*,O) zinc(II) acetate	(54)
	2Z9L	2.1		Bis(L-aspartato-*N*,O) zinc(II) acetate	(54)
	2QIQ	1.9		Substrate-based irreversible inhibitor	(55)
	2PWX	2.5		G11A dimer interface mutant	(56)
	2Q6G	2.5		H41A active site mutant	(57)
	2QC2	2.7		N214A monomeric mutant	N/A
	2QCY	1.75		R298A monomeric mutant	(58)
	2V6N	1.98		1-(4-Dimethylaminobenzoyloxy)-benzotriazole	(59)
	2VJ1	2.25		1-(Benzoyloxy)-benzotriazole	(59)
	2YY4	2.2			N/A
	2ZU4	1.93	Cys145-His41	Peptidomimetic inhibitor TG-0204998	(36)
	2ZU5	1.65		Peptidomimetic inhibitor TG-0205486	(36)
	3D62	2.7		Halomethyl ketone inhibitor	(60)
	3E91	2.55		S284A, T285A, I286A activity-enhanced mutant	N/A
	3EA7	2.65		S284A, T285A, I286A activity-enhanced mutant	N/A
	3EA8	2.25		S284A, T285A, I286A activity-enhanced mutant	N/A
	3EA9	2.4		S284A, T285A, I286A, F291A activity-enhanced mutant	N/A
	3EAJ	2.7		S284A, T285A, I286A, F291A activity-enhanced mutant	N/A
	3F9E	2.5		S139A dimer interface mutant	(61)
	3F9F	2.3		F140A dimer interface mutant, at pH 6.0	(61)
	3F9G	2.6		F140A dimer interface mutant, at pH 6.5	(61)
	3F9H	2.9		F140A dimer interface mutant, at pH 7.6	(61)
	3FZD	2.35		N28A inactivated mutant	(62)
	3IWM	3.2		Octameric form	(63)
	3M3S	2.3		N214A monomeric mutant	N/A
	3M3T	2.9		R298A monomeric mutant	N/A
	3M3V	2.7		S284A, T285A, I286A activity-enhanced mutant	N/A
Transmissible gastroenteritis virus	1LVO	1.96	Cys144-His41		(64)
	1P9U	2.37		Substrate-based inhibitor	(41)
	2AMP	2.7		Substrate-analog chloromethyl ketone inhibitor	(43)
Infectious bronchitis virus	2Q6D	2.35	Cys143-His41		(57)
	2Q6F	2.0		Substrate-based irreversible inhibitor	(57)
Sobemovirus					
Sesbania mosaic virus	1ZYO	2.4	Ser284-His181-Asp216		(65)
Astrovirus					
Human astrovirus serotype 1	2W5E	2.0	Ser551-His461-Asp489		(66)
Arterivirus					
Equine arteritis virus	1MBM	2.0	Ser120-His39-Asp65		(67)

N/A, not available.

The configurations of the respective C-terminal tails from the Norwalk, Southampton, and Chiba proteases differ greatly. The Chiba protease lacks the 8 C-terminal residues, which suggests that the configuration of the C-terminus is not fixed. Although the corresponding regions are present in the Norwalk and Southampton proteases, the sequences of these last eight residues are in opposing directions.

Table 1 provides the crystal structural data for the 3C and 3C-like proteases from picornaviruses, potyviruses, and coronaviruses, in addition to the noroviruses that are available in the PDB database. It is clear that these proteases share a common architecture, supporting the notion that they all belong to the chymotrypsin-like protease superfamily. These similarities might facilitate the development of a class of medications for treating diseases evoked by any one of these viruses.

Mutational analysis of Cys139 and charged amino acid residues

Before the determination of the crystal structure of the Chiba protease, Ala-scanning mutagenesis was used to identify the active-site amino acid residues (68, 69) (Table 2). The norovirus 3C-like protease is closely related to the picornaviral 3C protease, both of which are classified as a member of the chymotrypsin-like protease family. This similarity, based on the amino acid sequence of the 3C-like protease, helped lead to the identification of the GXCG motif, which is a Cys version of the GDSG sequence present in chymotrypsin. The Ser195 residue in the GDSG sequence is an active-site nucleophile indispensable for proteolysis. In turn, it seemed reasonable to assume that the Cys139 residue in the GXCG motif of the norovirus proteases functioned as a nucleophile. Indeed, the Ala mutation of Cys139 was found to have completely lost all proteolytic activity, indicating the essential role of Cys139 in proteolysis (68). Cys139 could then be replaced with the Ser residue without apparently affecting the activity, which suggested that the 3C-like protease is a member of the serine protease family (69).

In chymotrypsin, the His57 and Asp102 residues function as a general base and a third-member carboxylate during the proteolytic reaction, respectively. It was initially expected that similar residues participated in proteolysis by the norovirus protease. Based on this hypothesis, all of the charged amino acid residues found in the 3C-like protease of the Chiba strain were subjected to Ala-scanning mutagenesis (69) (Table 2). Among 37 individual Ala mutant proteases, those with mutations at six positions (Arg8, His30, Lys88, Arg89, Asp138, His157) exhibited no proteolytic activity. It was thus estimated by mutational analysis that the His30 residue functioned as a general base. Evidence for this estimation was later provided by the structural study described above. As no acidic residue corresponding to Asp102 of chymotrypsin has been found in the norovirus 3C-like protease, it appears to exert its proteolytic activity using the catalytic dyad of Cys139 and His30. The catalytic dyad composed of Cys and His functions in a manner similar to the active site of typical cysteine proteases, such as papain. It is intriguing that the norovirus 3C-like protease appears to adopt a mechanism of proteolysis involving a thiolate-imidazolium ion pair, although its structure resembles that of typical serine proteases. It remains unknown if the mechanism underlying proteolysis by the norovirus protease includes a charge relay or a thiolate-imidazolium ion pair. The coronavirus 3C-like proteases are likely to use a mechanism involving a thiolate-imidazolium ion pair, because their active site has been clearly identified as a catalytic dyad composed of Cys and His.

The Asp138 residue was the only acidic residue for which the Ala mutation led to a loss of activity (69). Since Asp138 is the residue preceding a nucleophile, Cys139, on the primary sequence, it was not expected to be involved in catalysis. In fact, it was revealed by the crystal structure that a carboxylate of Asp138 pointed in the opposite direction to that of Cys139, and it formed a salt bridge with Arg89, one of six residues identified as critical by Ala-scanning mutagenesis (Figure 3) (17). The crystal structure also indicated that the side chains of Arg8 and Lys88 interacted with a carbonyl oxygen of Thr69 and Val9, respectively. It is thought that these interactions are necessary for the proteolytic activity and/or integrity of the protease structure (17). Simultaneous mutation or exchange of charges between two residues involved in the formation of a salt bridge occasionally reverses the deleterious effect of a single mutation at each position. When double mutations, such as R89A/D138A or R89D/D138R were introduced into the protease, the effects of the respective single mutations were not reversed by the additional mutation (unpublished observation). These results indicate that Arg and Asp are essential for the function of the protease at positions 89 and 138, respectively.

His157 is located in the S1 hydrophobic pocket, which interacts with the side chain of Glu or Gln found at position P1 of the substrate (Figure 2B) (17, 19, 20). His157 is essential for substrate recognition, and it cannot be replaced by any other amino acid residue (69).

Saturation mutagenesis of the Glu54 residue

The Ala mutation of the Glu54 residue did not show any loss of proteolytic activity (69). On the other hand, the crystal structure indicates that Glu54 is located close to a general base, His30, in the active site, suggesting that catalysis is carried out by a catalytic triad in which Glu54 functions as a third-member carboxylate along with Cys139 and His30 (Figure 2A) (17, 19, 20).

To explore the role of the Glu54 residue and to determine whether Glu54 is involved in the mechanism of catalysis, saturation mutagenesis was applied to this residue (70) (Table 2). The glutathione *S*-transferase (GST) fusion construct (pGEX-2TK-3aBC; see Figure 1C) of the 22 C-terminal amino acids of the 3A protein and the entire region of 3B VPg and 3C-like protease were used to monitor the effects of Glu54 mutations on proteolytic activity. In this construct, two cleavage sites, 3A/3B and 3B/3C, were included. The properties of 19 different mutants were categorized into five groups: Gly, Ala, Ser, Cys, Gln, and Asp – mutants that exhibited high activity

Table 2 Summary of mutagenesis studies.

Residues	Mutated to	Effect of mutation	Remarks	References
Trp6	Phe	Wild-type activity	Aromaticity is essential for proteolysis	(72)
	Ala, His, Leu	No activity, with decreased expression		
Arg8	Ala, Lys, Leu, Gln	No activity	Arg is indispensable	(69)
Trp19	Phe, His	Wild-type activity	Aromaticity is important for proteolysis	(72)
	Ala, Leu	No activity, with decreased expression		
Thr27	Ala, Ser, Val	Decreased level of expression	Not essential for proteolysis	(72)
		Activity was either not or only negligibly affected		
His30	Ala, Asp, Glu, Asn, Gln, Arg, Tyr	No activity	The general base of a catalytic triad	(69)
			His is indispensable	
Glu54	Ala, Cys, Asp, Gly, Gln, Ser	Wild-type activity	Third-member carboxylate of a catalytic triad	(70, 71)
	His, Met, Asn	Reduced activity	Not essential for catalysis	
	Trp	Severely reduced activity		
	Ile, Leu, Pro	Altered substrate specificity		
	Phe, Lys, Arg, Thr, Val, Tyr	No activity		
Leu86	Phe, Ile, Met	Wild-type activity	A large hydrophobic amino acid is preferred	(72)
	Ala, Val	Reduced activity, with decreased expression		
	Gln	No activity, with decreased expression		
Lys88	Lys	Reduced activity	A positively charged amino acid is required	(69)
	Leu	No activity		
	Gln	No expression		
Arg89	Lys, Leu	No activity	Arg89 interacts with Asp138	(69)
	Gln	No expression	Arg is indispensable	
Leu95	Phe, Ile, Met, Gln, Val	Wild-type activity	Not essential for proteolysis	(72)
	Ala	No activity		
Leu97	Phe, Ile, Met, Val	Wild-type activity	Hydrophobicity is important	(72)
	Ala, Gln	No activity		
Met101	Ala, Leu	Severely reduced activity	Met101 is not essential for proteolysis but is important for protein stability	(72)
	Phe, Ile, Gln, Val	No activity		
		All mutations led to decreased level of expression		
Gln117	Leu, Val	Wild-type activity	Side-chain volume is important A charge is not preferred	(72)
	Ala	Severely reduced activity, with decreased expression		
	Glu, His, Asn	No activity		
Leu121	Ile, Val	Wild-type activity	A large, aliphatic amino acid is preferred	(72)
	Ala, Phe, Gln	No activity		
	Met	No expression		
Thr134	Ser	Wild-type activity	A residue in the S1 pocket	(72)
	Ala, Val	No activity	A hydroxyl group is essential	
Asp138	Ala, Glu, Met, Asn	No activity	Asp138 interacts with Arg89	(69)
			Asp is indispensable	
Cys139	Ser	Slightly reduced activity	The nucleophile of a catalytic triad	(68)
	Ala, Met, Thr, Tyr	No activity	A sulfhydryl group is essential for proteolysis	
			A hydroxyl group of Ser can act as a substitute	

Table 2 (Continued)

Residues	Mutated to	Effect of mutation	Remarks	References
Tyr143	Phe	Wild-type activity	A residue in the S1 pocket	(72)
	Ala, His, Ser	No activiy	Tyr143 interacts with His157	
			Aromaticity is important	
Val144	Leu, Met, Thr	Wild-type activity	A smaller or hydrophilic amino acid is not preferred	(72)
	Ala, Gln	No activity with a decreased level of expression		
His157	Ala, Pro, Gln, Arg, Ser, Tyr	No activity	A residue in the S1 pocket	(69)
			His is indispensable	
Val167	Thr	Wild-type activity	A residue in the S1 pocket	(72)
	Leu	Altered substrate specificity	Side-chain volume is important	
	Ala, Met, Gln	No activity with a decreased level of expression		

comparable to that of the wild type; His, Met, and Asn – mutants that exhibited slightly less activity than the wild type, as indicated by residual amounts of the GST-3aB intermediate, as well as the production of three final products; Trp – mutant that retained severely impaired but still significant activity, resulting in the production of very small amounts of final products; Leu, Ile, and Pro – mutants that cleaved only the 3B/3C site, but not the 3A/3B site; and Arg, Lys, Val, Thr, Phe, and Tyr – mutants lacking activity. It was clearly shown that the presence of the carboxylate of the Glu54 residue was not essential for proteolytic activity. However, the turnover rate of the mutants tested (Ala54 and Gln54 mutants) was significantly decreased compared with that of the wild-type protease. It is likely that the presence of a carboxylate at position 54 is required for effective proteolysis (70).

The C139S nucleophile mutant of the protease exhibits high proteolytic activity comparable to that of the wild-type protease. It was thus of interest to determine whether or not the triplet Ser139-His30-Glu54 functions as a catalytic triad in a manner similar to that of the catalytic triad in typical serine proteases. When the Glu54 residue was mutagenized in the construct possessing the C139S mutation, only the E54D/C139S and E54Q/C139S mutants retained slightly decreased activity compared with that of the parental C139S mutant, while the other double mutants exhibited virtually no activity. These results suggested that capacity to form hydrogen bonds was required for the side chain at position 54 in the Ser139 nucleophile background (70).

The far-reaching implications of these results remain to be seen. It is compelling to consider that the norovirus 3C-like protease adopts a papain-like mechanism involving a thiolate-imidazolium ion pair in the wild-type Cys nucleophile background, whereas in the Ser nucleophile background, it uses a chymotrypsin-like mechanism involving a charge relay. Additional experimental evidence from future studies will help clarify these issues.

Mutation of the Glu54 residue affects substrate specificity

The Leu, Ile, or Pro mutation of Glu54 resulted in the production of the GST-3aB intermediates and the 3C-like protease, implying that mutant proteases mediated only cleavage at the 3B/3C site (Lys-Leu-Ser-Phe-Glu/Ala-Pro), but not at the 3A/3B site (Thr-Ala-Thr-Ser-Glu/Gly-Lys) (70). One possibility is that the mutant protease is inactivated after the cleavage of the 3B/3C site, eventually resulting in a failure of the cleavage of the 3A/3B site. Another possibility is that the mutant protease possesses altered (strict) substrate specificity, as the 3A/3B cleavage sequence is apparently more hydrophilic than the 3B/3C sequence.

Various mutations were introduced at the 3A/3B or 3B/3C site using the fusion construct including the E54L protease mutation (71). When the Leu residue at the 3B/3C site was changed to Ala, the E54L protease did not cleave either the 3A/3B or the 3B/3C site. The Phe-to-Ser mutation of the 3B/3C site also gave the same result. On the other hand,

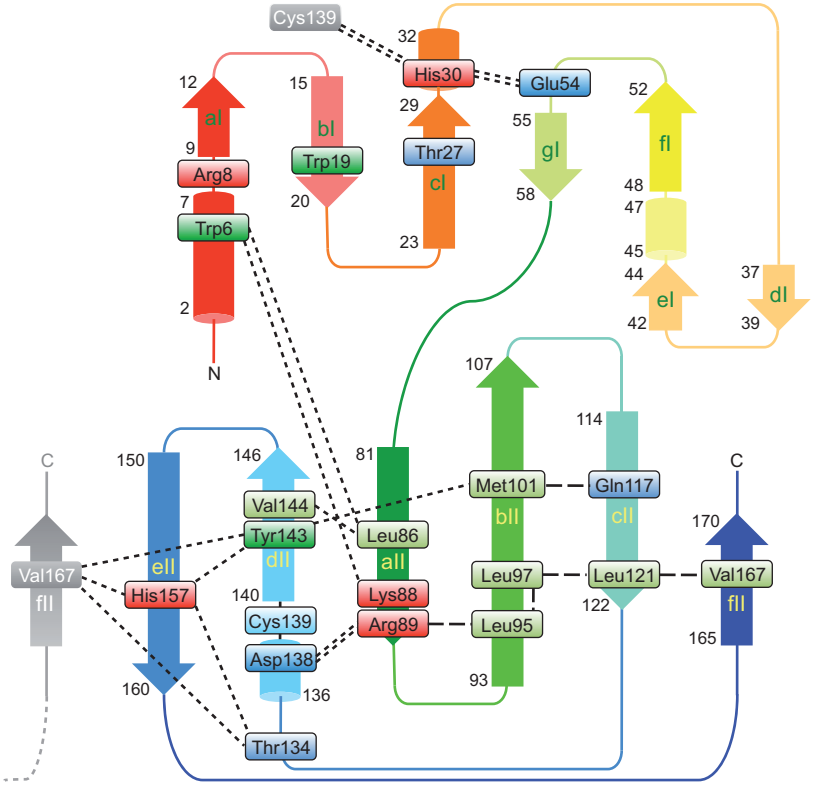

Figure 3 Map of amino acid residues important for proteolytic activity and/or stability.
Important residues identified in these mutagenesis studies are mapped on the topological view of the 3C-like protease. A dashed line between two residues indicates that the two residues are close to each other. A double dashed line indicates an ionic interaction. The triplet Cys139-His30-Glu54 constitutes a catalytic triad for proteolysis, although Glu54 is not essential. Arg89 was salt-bridged with Asp138. For clarity, the Cys139 residue and the fII β sheet are represented in two different locations.

when the Ala-Thr-Ser tripeptide at the 3A/3B site was changed to the Leu-Ser-Phe or Leu-Thr-Phe sequence, the E54L protease accomplished the cleavage at the 3A/3B site as well as at the 3B/3C site. These results indicate that the E54L mutant protease prefers a cleavage sequence that includes large hydrophobic amino acids at both the P4 and the P2 positions (71).

Careful consideration of the crystal structure of the protease clearly revealed that the Glu54 residue is located at the bottom of the S2 hydrophobic pocket that accommodates the side chain at the P2 position of the substrate (Figure 2B) (17, 71). As the Leu mutation of Glu54 leads to increased hydrophobicity of the S2 pocket, it is likely that the E54L protease requires a hydrophobic substrate to achieve a cleavage event.

Enzyme-substrate interaction as revealed by mutations at the 3A/3B cleavage site

When the S200F mutation of the 3A region was introduced into the GST-3aBC constructs including the wild-type protease, the protease moiety was isolated along with the GST-3a(S200F) peptide from the lysate of cultured bacterial cells (71). No other proteins accompanied the GST-3a protein from the completely wild-type GST-3aBC construct. The wild-type protease was not separated from the GST-3a(S200F) protein by gel filtration, indicating that this tight interaction was solely attributed to the S200F mutation in the 3A region. As the Ser200 residue of the 3A peptide corresponds to the P2 position in the cleavage sequence, it is possible that the C-terminus of 3A containing the S200F P2 mutation remains tightly bound at the active site in the protease, even after completion of the proteolytic reaction (71).

The tight interaction between the protease and the C-terminus of 3A occurred when the Tyr, Trp, Leu, or Met residue occupied position 200 of 3A, while this tight interaction was completely abolished when an Ala or His residue was introduced. These results suggest that hydrophobicity is responsible for the sustained interaction (71). As described above, the S2 pocket that accommodates the P2 position of the substrate is hydrophobic (17, 19, 20). It is likely that the hydrophobicity of both the S2 pocket and the P2 position synergistically strengthens the interaction.

On the other hand, when the Ala198 residue (P4 position) of 3A was additionally mutated to Gly or Val with the S200F mutation retained, the amounts of the protease co-isolated with GST-3a decreased significantly; moreover, the Ser or Leu mutation of Ala198 led to dissociation of the mutated C-terminus of 3A from the protease molecule. The binding sites for the P4 and P2 positions of the substrate are buried

in the protease. It is likely that the two binding sites mutually affect and sense each other when the proteolytic reaction proceeds (71).

As described above, the bII-cII loop connecting two β-sheets, bII and cII, is in the vicinity of the active site of the protease and possibly undergoes a dynamic conformational change during proteolysis (17, 19, 20). The S2 pocket is surrounded by the side chains of Ile109, Gln110, Arg112, and Val114 on the bII-cII loop (Figure 2B). Although the individual Ala single mutants at four positions had no effect on proteolytic activity, each of these Ala mutations significantly decreased the amounts of protease co-eluted with GST-3a in combination with the S200F mutation of 3A. In particular, the I109A mutation completely abolished the sustained interaction between the protease and the C-terminus of 3A, i.e., the product was released from the protease immediately after the completion of proteolysis. These results suggest that the observed tight binding between the protease and GST-3a is attributable to the hydrophobicity of both the Ile109 residue of the protease and the Phe200 residue of the mutated 3A (71).

According to the crystal structures of the proteases, it is reasonable to assume that the Ile109 residue is involved in the recognition of both the P2 and the P4 side chains, and therefore, mutual sensing of the P2 and P4 side chains by the protease probably occurs *via* the Ile109 residue (Figure 2B).

The tight interaction between the enzyme and the final product would be disadvantageous for the enzyme because the free enzyme is hardly regenerated, and therefore, such mutations at cleavage sites would never occur naturally. However, such artificial mutations may be used to help design inhibitors of these proteases.

It should be noted that the cleavage at the 3B/3C site precedes the cleavage at the 3A/3B site. If the two cleavage sites included in the GST-3aBC fusion proteins are equivalent, one would estimate that they should be exposed to the protease to the same degree. In fact, the GST-3aBC fusion containing the S200F mutation of 3A yielded the GST-3a peptide tightly associated with the 3C-like protease, but not the GST-3a bound to the 3BC intermediate. This would not be the case if the 3B/3C site were not cleaved in advance.

I want to mention an intriguing possibility on the bII-cII loop here. Staring at the amino acid sequence of this loop, it comes to our attention that its part including the Met107-Lys-Ile-Gln-Gly-Arg sequence resembles the cleavage sequences found in the norovirus polyprotein, especially the sequence between the N-terminal region and the 2C-like NTPase, Met-His-Leu-Gln/Gly-Pro. This raises the possibility that the bII-cII loop mimics the substrate and regulates the function of the protease. Although a drastic movement is needed along with a conformational change of the backbone peptide, the bII-cII loop might come to fit along the active site cleft as the Gln110 residue interacts with the His157 residue. This possibility cannot be completely denied as long as I see the crystal structure. The fact that the Met-Lys-Ile-Gln-Gly-Arg motif is highly conserved might also support this possibility. However, as there is no experimental evidence, it remains a matter of speculation.

Side chain-side chain interaction as revealed by Ala-scanning mutagenesis

At this stage of the research, it became important to consider whether neutral amino acids are involved in the mechanism of proteolysis, substrate binding, maintenance of the protein structure, and/or other aspects of the protease. To this end, the author's extensive mutagenesis study was expanded to include neutral amino acid residues, i.e., Pro, Val, Leu, Ile, Phe, Tyr, Trp, Ser, Thr, Cys, Met, Asn, and Gln (Table 2) (72). For that series, 106 individual Ala mutants were constructed. The results revealed that Ala mutations at 13 positions led to a loss of activity and/or a loss of stability of the protease. Ten of 13 residues (Leu86, Leu95, Leu97, Met101, Gln117, Leu121, Thr134, Tyr143, Val144, and Val167) were mapped on the C-terminal domain, and only 3 of 13 (Trp6, Trp19, and Thr27) on the N-terminal domain (Figure 3).

The crystal structure clearly indicated that the Thr134 residue, together with His157, is directly involved in binding of the P1 side chain of the substrate *via* the side-chain hydroxyl group (Figure 2B) (17, 19, 20). The Ser residue could be substituted for Thr134, but the Val residue could not, indicating an essential role of the hydroxyl group (72). The Tyr143 residue interacts with His157 by means of an aromatic interaction (17, 19, 20) because Tyr143 could be replaced with Phe (72). In addition to these residues, Leu97, Leu121, and Val67 are residues constituting the S1 specificity pocket (17). No other residues important for an enzyme-substrate interaction could be found in the mutagenesis study. These findings suggested that the substrate is primarily fixed in place by the protease due to its P1 position.

Interestingly, residues found in the C-terminal domain form a network in which side chains are close to each other. The network, most of which is based on hydrophobic interactions, is suggested to be important for the stability and integrity of the protease since mutations, especially those of Met101 and Gln117, often lead to decreased levels of expression (72).

As for the N-terminal domain, mutations at only three positions affected its activity and/or stability, which was in marked contrast to the C-terminal domain. The crystal structure indicates that Trp19 is located on top of an aromatic residue cluster found in the N-terminal domain (17). Since aromaticity at position 19 was found to be essential for achieving activity, an aromatic-aromatic interaction would be expected to be important, although the precise role of this interaction is unknown in this case (72). On the other hand, Trp6 is sandwiched between Leu86 and Lys88, thus representing the only interaction between two domains that was identified in this series (17). Crystallographic analysis revealed few interactions between the N- and the C-terminal domains. These findings suggested that limited interactions between the two domains played a part in the flexibility of the protease during proteolysis.

Effect of the truncation of the N- and C-termini

The question remained regarding which sequence of the protease is minimally required to achieve proteolytic activity.

Therefore, the N- and C-termini of the protease were truncated in separate studies to determine their respective effects on proteolytic activity.

The effects of the N-terminal truncation were assessed using the 3CD expression construct in which the first Ala residue of the protease was placed following the Met residue (Figure 1C) (69). The 3CD proteins were successfully cleaved into the 3C-like protease and 3D RNA polymerase in *Escherichia coli* cells. When the first five or eight amino acid residues of the protease were deleted, both of the truncated proteases were able to cleave the 3C/3D junction, although their respective levels of activity had markedly decreased. On the other hand, the deletion of the first 11 residues resulted in a complete loss of activity (69). These results are somewhat surprising because Trp6 and Arg8 are included within the first eight residues. Amino acid substitution of these residues was detrimental to the protease when the activity was measured using the GST-3aBC construct. The crystal structure shows that the N-terminus containing Trp6 and Arg8 forms an α-helix. One possible explanation for these findings was that substitution mutations of these residues alter the conformation of the N-terminus and thus prevent the proper formation of the α-helix. Therefore, substitution mutations might be more deleterious than deletion. It seems likely that the N-terminal containing Trp6 and Arg8 is not involved in proteolytic activity, and these two residues are responsible for the proper folding of the N-terminus; hence, the entire molecule of the protease is properly folded in this case. It should be noted that Trp6 and Arg8 are completely conserved in all norovirus strains, which suggests the crucial roles of these residues.

As noted above, the crystal analyses demonstrated that the C-terminal tails of each of the proteases from these three strains were not identical (17, 19, 20). These observations suggest that the C-terminus is not required for proteolytic activity. Mutagenesis study using the protease of the Chiba strain clearly indicated that 10 C-terminal residues could be truncated without affecting activity, and no specific interaction was formed with certain residues in the protease molecule (unpublished data).

Of 181 residues constituting the protease, neither the 8 residues from the N-terminus nor the 10 residues from the C-terminus are directly involved in achieving proteolysis. However, the N-terminus of the protease follows the C-terminus of 3B VPg *via* the 3B/3C junction, and similarly, the C-terminus of the protease is followed by the N-terminus of the 3D RNA polymerase *via* the 3C/3D junction. It is obvious from the crystal structures that the amino acid sequence corresponding to at least five residues from P4 to P1′ positions will be required for exerting the cleavage event when the junction as the substrate is fitted to its binding site in the protease *via* the antiparallel β-strand interaction.

Enzymatic characterization

To determine the biochemical and enzymatic properties of the norovirus protease, bacterially expressed protease molecules were purified and assayed *in vitro* using the purified fusion proteins, including the 3C/3D junction as the substrate (73). The purified protease carried out cleavage of the fusion proteins within a wide pH range (optimal pH, approx. 8.6) and without the requirement of divalent cations, such as Mg^{2+} and Ca^{2+}; however, cleavage was slightly inhibited in the presence of high concentrations of Na^{+} or K^{+}. In general, the proteolytic activity of the norovirus protease did appear to be quite low because not all of the substrate molecules were cleaved by the protease, not even after prolonged incubation (≥24 h). These results are in marked contrast with the observation that the fusion constructs were thoroughly cleaved inside *E. coli* cells. One possible explanation for this discrepancy is that intrinsic molecules in *E. coli* cells serve as modulators of proteolysis. As the fusion protein including the entire protease moiety with the C139A inactivating mutation was used as the substrate for the assay, it cannot be ruled out that this moiety might occlude the junction of another fusion molecule, eventually resulting in the accumulation of non-cleaved substrates.

Low proteolytic activity of the norovirus protease was also observed when synthetic, fluorogenic peptide substrates based on the 2C/3A junction were used (70). The wild-type protease showed a K_m value of 0.5 mM and a k_{cat} value of 0.6 s^{-1}, revealing that the enzyme had low affinity for the substrate and a slow turnover rate. It was a limitation of the assay that large amounts of fluorogenic peptides were used, which caused an unfavorable increase in the level of background fluorescence and resulted in a loss of linearity between the substrate concentration and the intensity of the generated fluorescence. The use of chromogenic peptide substrates would likely have yielded much more confirmable results than those of this series (17, 19, 20). However, the proteolytic activity of the norovirus protease did indeed appear to be low in this series.

Although it remains unknown how the norovirus protease accomplishes the cleavage of a viral polyprotein in infected cells, the junctions in a polyprotein that serve as substrate for the protease are in the vicinity of the protease itself, and thus it seems likely that the local concentration of the substrate is relatively high. Even if the protease exhibits weak proteolytic activity, it can take advantage of both substrate specificity and the adjacent localization of the substrate.

The protease was effectively inactivated by sulfhydryl reagents including *N*-ethylmaleimide, methyl methanethiosulfonate, *p*-chloromercuribenzoic acid, and the Hg^{2+} ion (73). These findings indicate that cysteine residues were involved in proteolysis. Although the protease from the Chiba strain has five Cys residues, only Cys139, a nucleophile in the active site, is conserved in all strains of norovirus. The crystal structure analysis revealed that four other residues (Cys77, Cys83, Cys154, and Cys169) serve as coordinates for heavy metal ions (17). As the sulfhydryl group of Cys139 is highly accessible to water, whereas the side chains of the other four residues are directed inward, it is highly possible that a target of sulfhydryl reagents is Cys139.

The heavy metal ions found in the crystal of the Chiba protease were possibly mercuric ions that were used to obtain heavy metal derivatives (17). Although the Norwalk or Southampton proteases had the same Cys cluster formed

inside the C-terminal domain, no metal ions were not found in their crsytals (19, 20). Therefore, such ions are unlikely to be needed for proteolytic activity. It is unknown whether the Cys cluster serves other particular function, although it will be interesting to explore the implication of its configuration.

Expert opinion

Noroviruses are distributed worldwide and cause the intense suffering of millions of people on a yearly basis. No therapeutic drugs or vaccines against norovirus gastroenteritis have been developed to date. The 3C-like protease is responsible for the maturation of the norovirus proteins, which are possibly involved in the replication of viral RNAs and the production of progeny viruses. Without the protease, noroviruses could not be propagated. A comprehensive understanding of all aspects of the norovirus protease and the concurrent development of medications to treat norovirus infections will provide a substantial contribution to human quality of life and health.

Norovirus protease is a member of the chymotrypsin-like serine protease superfamily. Chymotrypsin and its relatives are well-known; a wide variety of aspects of these proteases have already been reported in the literature (e.g., structural features, mechanisms of catalysis, enzymatic characteristics, inhibitor profiles, roles played in cells). Should the norovirus protease be compared with chymotrypsin-like serine proteases? The norovirus protease does have a conspicuous feature in that its nucleophile is a Cys residue, although its overall structure adopts a chymotrypsin-like fold.

The norovirus proteases were crystallized and the structures were solved by three different research groups, which provided various insights into active site residues, the principle of protein integrity, the mode of substrate binding, and possible structural changes during proteolysis. The norovirus protease has also been subjected to extensive and detailed mutagenesis studies, shedding light not only on the putative roles of individual amino acid residues but also on interactions between residues and interactions between enzymes and substrates.

The detailed structural and biochemical examination of norovirus proteases conducted thus far is expected to contribute substantially to the eventual development of therapeutic drugs for treating norovirus infections.

Outlook

The mechanism of catalysis exerted by the norovirus protease is still not resolved. The mechanism involving a charge relay system is plausible, considering that the protease is a member of the chymotrypsin-like serine protease family. This classification is supported by crystal structure analyses indicating that a triplet, Cys139-His30-Glu54, is properly positioned in the active site. However, mutagenesis study of Glu54 clearly indicated that Glu54 was not essential for proteolytic activity. It is possible that a dyad of Cys139 and His30 catalyzes a cleavage reaction by a mechanism involving a thiolate-imidazolium ion pair, as is the case with typical cysteine proteases, such as papain. Otherwise, the norovirus protease might utilize either mechanism, adopting one of the two depending on the context. This possibility is supported by the activity shown by the protease within a wide range of pH levels. Under acidic conditions, Cys139 is protonated such that proteolysis proceeds by a charge relay, whereas under alkaline conditions, Cys139 is easily deprotonated such that proteolysis is mediated by a thiolate-imidazolium ion pair. This hypothesis is based on the fact that the pK_a value of the sulfhydryl group of Cys (~ 8.3) is much smaller than that of the hydroxyl group of Ser (~13). Due to low levels of protease activity, it has been difficult to characterize the detailed enzymatic features of the protease with *in vitro* assays involving fluorogenic or chromogenic peptide substrates. More sensitive methods will be needed to elucidate the mechanism of catalysis and the enzyme kinetics.

The norovirus protease is initially expressed as a part of polyprotein. Then, each unit is excised from the polyprotein by the protease. There are five junctions in the polyprotein. It is not known whether the junctions are cleaved in an intermolecular or an intramolecular manner. Comparing the generation of a viral protease from a polyprotein by the autocatalysis of zymogens, Khan et al. (74) speculated that cleavage at the 3B/3C junction of the picornavirus polyprotein occurs intramolecularly. The question remains whether this is applicable to the norovirus polyprotein as well. Although the intramolecular event is an intriguing hypothesis, experimental evidence is still needed for verification.

Highlights

- Determination of the crystal structures confirmed that the norovirus 3C-like protease adopts a chymotrypsin-like fold.
- The crystal structures also indicated that the active site consisted of a catalytic triad, Cys139-His30-Glu54.
- Mutagenesis studies indicated that Cys139 and His30 played indispensable roles in proteolysis, but Glu54 was not essential for proteolysis.
- Mutations of Glu54 affected substrate specificity. In particular, the E54L, E54I, and E54P mutant proteases cleaved the 3B/3C junction, but not the 3A/3B junction.
- Together with Thr134 and Tyr143, His157 in the S1 specificity pocket is involved in the recognition of the P1 Glu/Gln residue of the substrate.
- Substrates in which large, hydrophobic amino acids occupy the P2 and P4 positions are highly preferred for cleavage.
- The Phe mutation of 3A Ser200 corresponding to the P2 residue blocked the release of the product containing the mutation from the active site of the wild-type protease, even after the completion of proteolysis.
- Most of the residues identified as important by Ala-scanning mutagenesis form a network in which side chains are close to each other, which suggests that this arrangement

is possibly important for the stability and integrity of the protease.
- The mechanism underlying proteolysis by the norovirus protease requires additional research.
- The means by which the junction is cleaved remains to be determined, and future studies will still need to determine whether the process is intermolecular or intramolecular.

Acknowledgments

This work was supported by partly by a Grant-in-Aid from the Ministry of Health, Labour and Welfare, Japan.

References

1. Glass RI, Parashar UD, Estes MK. Norovirus gastroenteritis. N Engl J Med 2009; 361: 1765–85.
2. Patel MM, Hall AJ, Vinjé J, Parashar UD. Noroviruses: a comprehensive review. J Clin Virol 2008; 44: 1–8.
3. Hansman GS, Jiang XJ, Green KY, editors. Caliciviruses: molecular and cellular virology, Norwich, UK: Caister Academic Press, 2010.
4. Daughenbaugh KF, Fraser CS, Hershey JW, Hardy ME. The genome-linked protein VPg of the Norwalk virus binds eIF3, suggesting its role in translation initiation complex recruitment. EMBO J 2003; 22: 2852–9.
5. Prasad BV, Hardy ME, Dokland T, Bella J, Rossmann MG, Estes MK. X-ray crystallographic structure of the Norwalk virus capsid. Science 1999; 286: 287–90.
6. Cao S, Lou Z, Tan M, Chen Y, Liu Y, Zhang Z, Zhang XC, Jiang X, Li X, Rao Z. Structural basis for the recognition of blood group trisaccharides by norovirus. J Virol 2007; 81: 5949–57.
7. Shirato H, Ogawa S, Ito H, Sato T, Kameyama A, Narimatsu H, Xiaofan Z, Miyamura T, Wakita T, Ishii K, Takeda, N. Noroviruses distinguish between type 1 and type 2 histo-blood group antigens for binding. J Virol 2008; 82: 10756–67.
8. Okada M, Ogawa T, Kaiho I, Shinozaki K. Genetic analysis of noroviruses in Chiba Prefecture, Japan, between 1999 and 2004. J Clin Microbiol 2005; 43: 4391–401.
9. Shirato H. Norovirus and histo-blood group antigens. Jpn J Infect Dis 2011; 64: 95–103.
10. Belliot G, Sosnovtsev SV, Mitra T, Hammer C, Garfield M, Green KY. In vitro proteolytic processing of the MD145 norovirus ORF1 non-structural polyprotein yields stable precursors and products similar to those detected in calicivirus-infected cells. J Virol 2003; 77: 10957–74.
11. Sosnovtsev SV, Belliot G, Chang KO, Prikhodko VG, Thackray LB, Wobus CE, Karst SM, Virgin HW, Green KY. Cleavage map and proteolytic processing of the murine norovirus non-structural polyprotein in infected cells. J Virol 2006; 80: 7816–31.
12. Bazan JF, Fletterick RJ. Viral cysteine proteases are homologous to the trypsin-like family of serine proteases: structural and functional implications. Proc Natl Acad Sci USA 1988; 85: 7872–6.
13. Dougherty WG, Semler BL. Expression of virus-encoded proteinases: functional and structural similarities with cellular enzymes. Microbiol Rev 1993; 57: 781–822.
14. Gorbalenya AE, Donchenko AP, Blinov VM, Koonin EV. Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases. FEBS Lett 1989; 243: 103–14.
15. Malcom BA. The picornaviral 3C proteinases: cysteine nucleophiles in serine proteinase fold. Protein Sci 1995; 4: 1439–45.
16. Seipelt J, Guarné A, Bergmann E, James M, Sommergruber W, Fita I, Skern T. The structures of picornaviral proteinases. Virus Res 1999; 62: 159–68.
17. Nakamura K, Someya Y, Kumasaka T, Ueno G, Yamamoto M, Sato T, Takeda N, Miyamura T, Tanaka N. A norovirus protease structure provides insights into active and substrate binding site integrity. J Virol 2005; 79: 13685–93.
18. Fersht AR, Sperling J. The charge relay system in chymotrypsin and chymotrypsinogen. J Mol Biol 1973; 74: 137–49.
19. Zeitler CE, Estes MK, Prasad BVV. X-ray crystallographic structure of the Norwalk virus protease at 1.5-Å resolution. J Virol 2006; 80: 5050–8.
20. Hussey RJ, Coates L, Gill RS, Erskine PT, Coker S-F, Mitchell E, Cooper JB, Wood S, Broadbridge R, Clarke IN, Lambden PR, Shoolingin-Jordan PM. A structural study of norovirus 3C protease specificity: binding of a designed active site-directed peptide inhibitor. Biochemistry 2011; 50: 240–9.
21. Mosimann SC, Cherney MM, Sia S, Plotch S, James MNG. Refined X-ray crystallographic structure of the poliovirus 3C gene product. J Mol Biol 1997; 273: 1032–47.
22. Marcotte LL, Wass AB, Gohara DW, Pathak HB, Arnold JJ, Filman DJ, Cameron CE, Hogle JM. Crystal structure of poliovirus 3CD protein: virally encoded protease and precursor to the RNA-dependent RNA polymerase. J Virol 2007; 81: 3583–96.
23. Matthews DA, Smith WW, Ferre RA, Condon B, Budahazi G, Sisson W, Villafranca JE, Janson CA, McElroy HE, Gribskov CL, Woprland S. Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein. Cell 1994; 77: 761–71.
24. Bjorndahl TC, Andrew LC, Semenchenko V, Wishart DS. NMR solution structures of the apo and peptide-inhibited human rhinovirus 3C protease (serotype 14): structural and dynamic comparison. Biochemistry 2007; 46: 1294512958.
25. 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. 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 USA 1999; 96: 11000–7.
26. Baxter A, Chambers M, Edfeldt F, Edman K, Freeman A, Johansson C, King S, Morley A, Petersen J, Rawlins P, Spadola L, Thong B, Van de Poël H, Williams N. Non-covalent inhibitors of rhinovirus 3C protease. Bioorg Med Chem Lett 2011; 21: 777–80.
27. Petersen JF, Cherney MM, Liebig HD, Skern T, Kuechler E, James MN. 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 1999; 18: 5463–75.
28. Allaire M, Chernaia MM, Malcolm BA, James MN. Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases. Nature 1994; 369: 72–6.
29. Bergmann EM, Mosimann SC, Chernaia MM, Malcolm BA, James MN. The refined crystal structure of the 3C gene product from hepatitis A virus: specific proteinase activity and RNA recognition. J Virol 1997; 71: 2436–48.
30. Bergmann EM, Cherney MM, Mckendrick J, Frormann S, Luo C, Malcolm BA, Vederas JC, James MN. Crystal structure of an inhibitor complex of the 3C proteinase from hepatitis A virus (HAV) and implications for the polyprotein processing in HAV. Virology 1999; 265: 153–63.

31. Yin J, Bergmann EM, Cherney MM, Lall MS, Jain RP, Vederas JC, James MN. Dual modes of modification of hepatitis A virus 3C protease by a serine-derived β-lactone: selective crystallization and formation of a functional catalytic triad in the active site. J Mol Biol 2005; 354: 854–71.
32. Yin J, Cherney MM, Bergmann EM, Zhang J, Huitema C, Pettersson H, Eltis LD, Vederas JC, James MN. An episulfide cation (thiiranium ring) trapped in the active site of HAV 3C proteinase inactivated by peptide-based ketone inhibitors. J Mol Biol 2006; 361: 673–86.
33. Birtley JR, Knox SR, Jaulent AM, Brick P, Leatherbarrow RJ, Curry S. Crystal structure of foot-and-mouth disease virus 3C protease. New insights into catalytic mechanism and cleavage specificity. J Biol Chem 2005; 280: 11520–7.
34. Sweeney TR, Roqué-Rosell N, Birtley JR, Leatherbarrow RJ, Curry S. Structural and mutagenic analysis of foot-and-mouth disease virus 3C protease reveals the role of the β-ribbon in proteolysis. J Virol 2007; 81: 115–24.
35. Zunszain PA, Knox SR, Sweeney TR, Yang J, Roqué-Rosell N, Belsham GJ, Leatherbarrow RJ, Curry S. Insights into cleavage specificity from the crystal structure of foot-and-mouth disease virus 3C protease complexed with a peptide substrate. J Mol Biol 2010; 395: 375–89.
36. Lee CC, Kuo CJ, Ko TP, Hsu MF, Tsui YC, Chang SC, Yang S, Chen SJ, Chen HC, Hsu MC, Shih SR, Liang PH, Wang AH. Structural basis of inhibition specificities of 3C and 3C-like proteases by zinc-coordinating and peptidomimetic compounds. J Biol Chem 2009; 284: 7646–55.
37. Cui S, Wang J, Fan T, Qin B, Guo L, Lei X, Wang J, Wang M, Jin Q. Crystal structure of human enterovirus 71 3C protease. J Mol Biol 2011; 408: 449–61.
38. Phan J, Zdanov A, Evdokimov AG, Tropea JE, Peters HK 3rd, Kapust RB, Li M, Wlodawer A, Waugh DS. Structural basis for the substrate specificity of tobacco etch virus protease. J Biol Chem 2002; 277: 50564–72.
39. Nunn CM, Jeeves M, Cliff MJ, Urquhart GT, George RR, Chao LH, Tscuchia Y, Djordjevic S. Crystal structure of tobacco etch virus protease shows the protein C terminus bound within the active site. J Mol Biol 2005; 350: 145–55.
40. Sun P, Austin BP, Tözsér J, Waugh DS. Structural determinants of tobacco vein mottling virus protease substrate specificity. Protein Sci 2010; 19: 2240–51.
41. Anand K, Ziebuhr J, Wadhwani P, Mesters JR, Hilgenfeld R. Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. Science 2003; 300: 1763–7.
42. Yang H, Yang M, Ding Y, Liu Y, Lou Z, Zhou Z, Sun L, Mo L, Ye S, Pang H, Gao GF, Anand K, Bartlam M, Hilgenfeld R, Rao Z. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. Proc Natl Acad Sci USA 2003; 100: 13190–5.
43. Yang H, Xie W, Xue X, Yang K, Ma J, Liang W, Zhao Q, Zhou Z, Pei D, Ziebuhr J, Hilgenfeld R, Yuen KY, Wong L, Gao G, Chen S, Chen Z, Ma D, Bartlam M, Rao Z. Design of wide-spectrum inhibitors targeting coronavirus main proteases. PloS Biol 2005; 3: e324.
44. Hsu MF, Kuo CJ, Chang KT, Chang HC, Chou CC, Ko TP, Shr HL, Chang GG, Wang AH, Liang PH. Mechanism of the maturation process of SARS-CoV 3CL protease. J Biol Chem 2005; 280: 31257–66.
45. Ghosh AK, Xi K, Ratia K, Santarsiero BD, Fu W, Harcourt BH, Rota PA, Baker SC, Johnson ME, Mesecar AD. Design and synthesis of peptidomimetic severe acute respiratory syndrome chymotrypsin-like protease inhibitors. J Med Chem 2005; 48: 6767–71.
46. Lee TW, Cherney MM, Huitema C, Liu J, James KE, Powers JC, Eltis LD, James MN. Crystal structures of the main peptidase from the SARS coronavirus inhibited by a substrate-like aza-peptide epoxide. J Mol Biol 2005; 353: 1137–51.
47. Tan J, Verschueren KH, Anand K, Shen J, Yang M, Xu Y, Rao Z, Bigalke J, Heisen B, Mesters JR, Chen K, Shen X, Jiang H, Hilgenfeld R. pH-dependent conformational flexibility of the SARS-CoV main proteinase (M^{pro}) dimer: molecular dynamics simulations and multiple X-ray structure analyses. J Mol Biol 2005; 354: 25–40.
48. Xu T, Ooi A, Lee HC, Wilmouth R, Liu DX, Lescar J. Structure of the SARS coronavirus main proteinase as an active C2 crystallographic dimer. Acta Crystallogr Sect F Struct Biol Cryst Commun 2005; 61: 964–6.
49. Lu IL, Mahindroo N, Liang PH, Peng YH, Kuo CJ, Tsai KC, Hsieh HP, Chao YS, Wu SY. Structure-based drug design and structural biology study of novel non-peptide inhibitors of severe acute respiratory syndrome coronavirus main protease. J Med Chem 2006; 49: 5154–61.
50. Lee TW, Cherney MM, Liu J, James KE, Powers JC, Eltis LD, James MNG. Crystal structures reveal an induced-fit binding of a substrate-like aza-peptide epoxide to SARS coronavirus main peptidase. J Mol Biol 2007; 366: 916–32.
51. Xue X, Yang H, Shen W, Zhao Q, Li J, Yang K, Chen C, Jin Y, Bartlam M, Rao Z. Production of authentic SARS-CoV M^{pro} with enhanced activity: application as a novel tag-cleavage endopeptidase for protein overproduction. J Mol Biol 2007; 366: 965–75.
52. Yin J, Niu C, Cherney MM, Zhang J, Huitema C, Eltis LD, Vederas JC, James MN. A mechanistic view of enzyme inhibition and peptide hydrolysis in the active site of the SARS-CoV 3C-like peptidase. J Mol Biol 2007; 371: 1060–74.
53. Goetz DH, Choe Y, Hansell E, Chen YT, McDowell M, Jonsson CB, Roush WR, McKerrow J, Craik CS. Substrate specificity profiling and identification of a new class of inhibitor for the major protease of the SARS coronavirus. Biochemistry 2007; 46: 8744–52.
54. Lee CC, Kuo CJ, Hsu MF, Liang PH, Fang JM, Shie JJ, Wang AH. Structural basis of mercury- and zinc-conjugated complexes as SARS-CoV 3C-like protease inhibitors. FEBS Lett 2007; 581: 5454–8.
55. Ghosh AK, Xi K, Grum-Tokars V, Xu X, Ratia K, Fu W, Houser KV, Baker SC, Johnson ME, Mesecar AD. Structure-based design, synthesis, and biological evaluation of peptidomimetic SARS-CoV 3CLpro inhibitors. Bioorg Med Chem Lett 2007; 17: 5876–80.
56. Chen S, Hu T, Zhang J, Chen J, Chen K, Ding J, Jiang H, Shen X. Mutation of Gly-11 on the dimer interface results in the complete crystallographic dimer dissociation of severe acute respiratory syndrome coronavirus 3C-like protease: crystal structure with molecular dynamics simulations. J Biol Chem 2008; 283: 554–64.
57. Xue X, Yu H, Yang H, Xue F, Wu Z, Shen W, Li J, Zhou Z, Ding Y, Zhao Q, Zhang XC, Liao M, Bartlam M, Rao Z. Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design. J Virol 2008; 82: 2515–27.
58. Shi J, Sivaraman J, Song J. Mechanism for controlling the dimer-monomer switch and coupling dimerization to catalysis of the severe acute respiratory syndrome coronavirus 3C-like protease. J Virol 2008; 82: 4620–9.
59. Verschueren KH, Pumpor K, Anemüller S, Chen S, Mesters JR, Hilgenfeld R. A structural view of the inactivation of the SARS coronavirus main proteinase by benzotriazole esters. Chem Biol 2008; 15: 597–606.

60. Bacha U, Barrila J, Gabelli SB, Kiso Y, Amzel LM, Freire E. Development of broad-spectrum halomethyl ketone inhibitors against coronavirus main protease 3CLpro. Bioorg Med Chem Lett 2008; 72: 34–49.
61. Hu T, Zhang Y, Li L, Wang K, Chen S, Chen J, Ding J, Jiang H, Shen X. Two adjacent mutations on the dimer interface of SARS coronavirus 3C-like protease cause different conformational changes in crystal structure. Virology 2009; 388: 324–34.
62. Barrila J, Gabelli SB, Bacha U, Amzel LM, Freire E. Mutation of Asn28 disrupts the dimerization and enzymatic activity of SARS $3CL^{pro}$. Biochemistry 2010; 49: 4308–17.
63. Zhang S, Zhong N, Xue F, Kang X, Ren X, Chen J, Jin C, Lou Z, Xia B. Three-dimensional domain swapping as a mechanism to lock the active conformation in a super-active octamer of SARS-CoV main protease. Protein Cell 2010; 1: 371–83.
64. Anand K, Palm GJ, Mesters JR, Siddell SG, Ziebuhr J, Hilgenfeld R. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra α-helical domain. EMBO J 2002; 21: 3213–24.
65. Gayathri P, Satheshkumar PS, Prasad K, Nair S, Savithri HS, Murthy MR. Crystal structure of the serine protease domain of Sesbania mosaic virus polyprotein and mutational analysis of residues forming the S1-binding pocket. Virology 2006; 346: 440–51.
66. Speroni S, Rohayem J, Nenci S, Bonivento D, Robel I, Barthel J, Luzhkov VB, Coutard B, Canard B, Mattevi A. Structural and biochemical analysis of human pathogenic astrovirus serine protease at 2.0 A resolution. J Mol Biol 2009; 387: 1137–52.
67. Barrette-Ng IH, Ng KK, Mark BL, Van Aken D, Cherney MM, Garen C, Kolodenko Y, Gorbalenya AE, Snijder EJ, James MN. Structure of arterivirus nsp4. The smallest chymotrypsin-like proteinase with an α/β C-terminal extension and alternate conformations of the oxyanion hole. J Biol Chem 2002; 277: 39960–6.
68. Someya Y, Takeda N, Miyamura T. Complete nucleotide sequence of the Chiba virus genome and functional expression of the 3C-like protease in Escherichia coli. Virology 2000; 278: 490–500.
69. Someya Y, Takeda N, Miyamura T. Identification of active-site amino acid residues in the Chiba virus 3C-like protease. J Virol 2002; 76: 5949–58.
70. Someya Y, Takeda N, Wakita T. Saturation mutagenesis reveals that Glu54 of norovirus 3C-like protease is not essential for the proteolytic activity. J Biochem 2008; 144: 771–80.
71. Someya Y, Takeda N. Insights into the enzyme–substrate interaction in the norovirus 3C-like protease. J Biochem 2009; 146: 509–21.
72. Someya Y, Takeda N. Functional consequences of mutational analysis of norovirus protease. FEBS Lett 2011; 585: 369–74.
73. Someya Y, Takeda N, Miyamura T. Characterization of the norovirus 3C-like protease. Virus Res 2005; 110: 91–6.
74. Khan AR, Khazanovich-Bernstein N, Bergmann EM, James MNG. Structural aspects of activation pathways of aspartic protease zymogens and viral 3C protease precursor. Proc Natl Acad Sci USA 1999; 96: 10968–75.

Received June 16, 2011; accepted November 2, 2011

Yuichi Someya received his PhD in Pharmacy from the Graduate School of Pharmaceutical Sciences, Chiba University, Japan in 1996. He went on to become a research assistant at the Institute of Scientific and Industrial Research, Osaka, Japan from 1996 to 1998. He was then a researcher in the Department of Virology II, at the National Institute of Infectious Diseases, Tokyo, Japan. From 2002 to the present he has been a senior researcher at the same department.