C-Terminal Arginine Cluster Is Essential for Receptor Binding of Norovirus Capsid Protein

Ming Tan,¹ Jarek Meller,² and Xi Jiang^{1*}

Division of Infectious Diseases¹ and Division of Biomedical Informatics,² Cincinnati Children's Hospital Medical Center, Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, Ohio

Received 1 February 2006/Accepted 16 May 2006

Noroviruses are the major viral pathogens of epidemic acute gastroenteritis affecting people worldwide. They have been found to recognize human histo-blood group antigens as receptors. The P domain of norovirus capsid protein was found to be responsible for binding to viral receptors, and the recombinant P protein forms P dimers and P particles in vitro. In this study, we demonstrate that a highly conserved arginine (R) cluster at the C terminus of the P domain is critical for receptor binding and P particle formation of the P proteins. Deletions of the R cluster abolished these functions. Replacement of the R cluster with histidines (another positively charged amino acid) resulted in low efficiency of receptor binding and P particle formation, while replacement with alanines led to loss of both functions completely. The R cluster also contains a highly conserved trypsin digestion site. A treatment of capsid protein or P domain mutants from both genogroup I (Norwalk virus) and genogroup II (VA387) noroviruses with trypsin resulted in a removal of the R cluster and the S domain, leaving a P polypeptide of 31.3 kDa (Norwalk virus) or 34.3 kDa (VA387), similar to the soluble P protein found in vivo. Our findings imply that the proteolytic process could be a necessary step for norovirus replication in the host.

Noroviruses are the major viral pathogen of epidemic acute gastroenteritis, affecting people in both developing and developed countries (4, 5). They are a group of small round structured viruses containing a single-stranded, positive-sense RNA genome of ~7.5 kb. Noroviruses are genetically highly diverse, with five genogroups and at least 25 genetic clusters identified so far (1, 7, 8). The study of noroviruses has been hampered by the lack of a cell culture or an animal model for human norovirus. Nevertheless, recent advances in understanding of norovirus-host interaction and the viral receptors has provided new insights into the host range and pathogenesis of norovirus disease (17, 31), opening new possibilities for the development of strategies to control and prevent norovirus infection.

Noroviruses have been shown to recognize human histo-blood group antigens (HBGAs) as receptors (14, 15, 18, 22, 24, 30, 31). Human HBGAs are complex carbohydrates present on the surfaces of red blood cells and, more importantly, on the mucosal epithelium and in body fluids with important biological functions (23, 28). The recognition of HBGAs by noroviruses is highly strain specific, and eight distinct receptor-binding patterns have been identified so far (12–15, 19, 24, 33). The linkage of Norwalk virus (NV) binding to HBGAs with clinical infection has also been demonstrated in human volunteer studies (16, 18, 22). By analogy, it is anticipated that other noroviruses with distinct receptor-binding patterns may have their corresponding host ranges, although direct evidence remains lacking.

Noroviruses contain a protein capsid that is composed of a single major structural protein, the capsid protein (26, 27).

When the capsid protein is expressed in insect cells, it spontaneously forms virus-like particles (VLPs) that are morphologically and antigenically indistinguishable from the authentic virus. Therefore, VLPs serve as valuable models for studying norovirus immunology, epidemiology, and pathogenesis (14, 15, 21). The crystal structure of the recombinant Norwalk virus capsid shows that norovirus capsid is composed of 180 capsid protein monomers that organize into a T=3 icosahedron (25). The capsid protein consists of two major domains, the shell (S) and the protrusion (P) domains, which are linked by a short hinge. The S domain forms the interior shell of the capsid, while the P domain builds up arch-like structures extending from the shell. When the S domain alone was expressed in insect cells, it formed smooth, thin-layer particles with a smaller size than VLPs (2, 29). The P domain can be further divided into P1 and P2 subdomains that correspond to the leg and the top of the arch-like capsomer, respectively (25). Since the P2 subdomain is located at the most exterior surface of the capsid and contains the most variable sequence, it is believed that this subdomain is important for host immune response and receptor interactions. In accordance with this hypothesis, a receptor-binding site has been identified in this region by computer modeling followed by site-directed mutagenesis analysis (3, 20, 30).

The crystal structure of the Norwalk virus capsid also revealed several putative intermolecular interaction sites in each capsid protein (25) which may be responsible for the capsid assembly. Most of these sites reside in the P domain, suggesting a strong intermolecular interaction between the P domains. Indeed, when a P domain was expressed in bacteria, it formed spontaneously a P dimer and a P particle that is composed of 12 P dimers (29, 32). While the P dimer retained only weak binding affinity to HBGA receptors, the P particle displays an enhanced binding activity which is comparable or even higher

* Corresponding author. Mailing address: Division of Infectious Diseases, Cincinnati Children's Hospital Medical Center, 3333 Burnet Ave., Cincinnati, OH 45229-3039. Phone: (513) 636-0119. Fax: (513) 636-7655. E-mail: jason.jiang@cchmc.org.

TABLE 1. Primers used to generate constructs for expression of recombinant P proteins

Name	Sequence (5' to 3')	Sense	Enzyme	Construct generated
VA387				
P422	GCACGGATCCTTCTTGGTGCCACCCACAGTT	+	BamHI	P/H
P524	GCACGGATCCTCAAGAACTAAACCATTCACC	+	BamHI	P
P560	GCGTGGATCCAAACCATTCACCGTCCCGATC	+	BamHI	P-N-2
P612	ACTCAGCGGCCGCTAATGCACGTCTGCGCCCCGCTCC	–	NotI	PC-1
P613	ACTCAGCGGCCGCTAAACGTCTGCGCCCCGCTCCATT	–	NotI	PC-2
P614	ACTCAGCGGCCGCTAATCTGCGCCCCGCTCCATTTC	–	NotI	PC-3
P615	ACTCAGCGGCCGCTAAGCGCCCCGCTCCATTTCCT	–	NotI	PC-4
P561	ACTCAGCGGCCGCTAACCCTGCTCCATTTCCTGGG	–	NotI	PC-5
P651	GGAAATGGAGCGGGGCGCGCACGTGCATTATAAGCGGCCG	+	NotI	P-RARAL
P652	GCGGCCGCTTATAATGCACGTGCGCGCCCCGCTCCATTTC	–	NotI	P-RARAL
P653	GGAAATGGAGCGGGGCGCAAGACGTGCATTATAAGCG	+	NotI	P-ARRAL
P654	CGCTTATAATGCACGTCTTGCCTCCGCTCCATTTC	–	NotI	P-ARRAL
P667	AGTAAGCGGCCGCTTAGGCAAGTGTGTAGAAGTGGTTG	–	NotI	PC-12
P731	GGAAATGGAGCGGGGCGACACCATGCATTATGCGATTGC	+		P-HHHAL-CDCRGDCFC
P732	GCAATCGCATAATGCATGGTGGTGCCCCGCTCCATTTC	–		P-HHHAL-CDCRGDCFC
Norwalk virus				
P493	GCACGGATCCTTTTGTAGTCCCTCCTACGGTG	+	BamHI	P/H
P668	GCATGCGGCCGCTTACAGACCAAGCCTACCTCTTGC	–	NotI	PC-2 NV

than that of VLP. Since both P dimer and P particle can be easily produced in a bacterial system and they both retain the binding function, they may serve as a new model for studying norovirus-receptor interactions.

Previous studies showed that as much as 50% of the norovirus antigen was secreted in the stool of norovirus-infected patients as a soluble protein of ~30 kDa (6, 9). This protein (P polypeptide) has subsequently been shown to include most of the P domain which was cleaved from the full-length capsid protein by trypsin (11). Sequence analysis of norovirus capsid protein reveals many potential proteolytic cleavage sites across the entire capsid protein, including ~15 trypsin recognition sites within the P domain. However, trypsin digestion of Norwalk virus capsid protein resulted in only partial digestion of the capsid protein, in which the S domain was completely degraded while the P domain remained basically intact as an ~32-kDa protein (11). Partial digestion of Norwalk virus capsid protein was also observed when it was expressed in insect cells using a baculovirus system (21). These observations raise the possibility that the soluble P polypeptide could play a role in the replication, host immune response, and pathogenesis of noroviruses.

In this study, we further characterized the structure/function relationship of norovirus P domain and demonstrated that a highly conserved arginine (R) cluster at the C terminus of the P domain is critical for receptor binding and P particle formation of the P protein. Notably, the R cluster is also sensitive to trypsin digestion. A treatment of a capsid protein with trypsin resulted in a P polypeptide of 31 to 34 kDa that lost its ability to bind receptors and form the P particle. Our findings implied that the proteolytic process could be an important step in norovirus replication in the host. The resulting inactive P polypeptide could be involved in the replication, host immunity, and pathogenesis of noroviruses.

MATERIALS AND METHODS

Construction of norovirus P domain mutants. Different P domain mutants of VA387 (genogroup II) and Norwalk virus (genogroup I) used in this study were

constructed by cloning the corresponding coding regions of the P domain cDNA sequences into the pGEX-4T-1 (GST gene fusion system; Amersham Biosciences, Piscataway, NJ) at the BamHI and NotI sites for expression in bacteria. The P mutants with N- or C-terminal-linked short peptides were prepared using primers (Table 1) with the peptide-encoding sequences at their 5' or 3' end, as described in a previous study (32). The P mutants P-RARAL and P-ARRAL of VA387 were constructed with the QuickChange site-directed mutagenesis kit (Stratagene, La Jolla, CA) using the plasmids of P mutants as templates. The construct P-HHHAL-CDCRGDCFC was made with the same QuickChange site-directed mutagenesis kit, using the plasmid P-CDCRGDCFC (32) as a template. The mutation sites and cloning conjunctions of all expression constructs were confirmed by sequencing. All primers used in this study are listed in Table 1.

Expression and purification of the recombinant P proteins in *Escherichia coli*. After sequence confirmation, the constructs containing different norovirus P cDNAs were expressed in *E. coli* (strain BL21) at room temperature overnight following an induction with 0.1 to 0.5 mM isopropyl- β -D-thiogalactopyranoside. Purification of the recombinant glutathione S-transferase (GST)-P fusion protein from bacteria was performed using Glutathione Sepharose 4 Fast Flow (Amersham Biosciences, Piscataway, NJ) according to the manufacturer's instruction, as described previously (29, 32). The fusion proteins were eluted by glutathione (Amersham Biosciences, Piscataway, NJ), and the P proteins were released from GST by thrombin (Amersham Biosciences, Piscataway, NJ) cleavage at room temperature for 16 h. Further purification was conducted by gel filtration using a size-exclusion column and/or anion exchange (see below). Protein concentrations were measured spectroscopically or by quantitative sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), depending on the purity of the proteins (29, 30).

Expression and purification of norovirus VLPs. The capsid proteins of Norwalk virus and VA387, representing the genogroup I and II noroviruses, respectively, were expressed in *Spodoptera frugiperda* (sf9) insect cells using the Bac-to-Bac baculovirus expression system (Invitrogen, Carlsbad, CA) as described previously (14). The VLPs from the infected insect cells were purified by centrifugation through a sucrose gradient containing 10% to 50% sucrose. The peak fractions containing VLPs were pooled and stored at -70°C in $1\times$ phosphate-buffered saline (PBS), pH 7.4.

Protease digestions. Modified trypsin (sequencing grade) was purchased from Promega (San Luis Obispo, CA). Stock solutions were prepared in a concentration of 0.1 mg/ml in trypsin resuspension buffer (50 mM acetic acid). Protease digestions were performed in $1\times$ PBS (pH 7.4) at a protease:protein ratio of 1:50 to 1:20 (wt/wt) at 37°C for the indicated times. Alternatively, proteins were digested by trypsin-EDTA solution (5 mg/ml; Invitrogen, Carlsbad, CA) at the indicated concentration and time.

Gel filtration. To further purify the capsid proteins and to determine if the P proteins form a dimer or a P particle (29, 32), the affinity column-purified proteins were loaded on a Superdex 200 size-exclusion column (HiLoad 16/60;

FIG. 1. Sequence alignment of the N and the C termini of norovirus P domains. (A) Shown is a schematic linear structure of norovirus capsid protein with indications of the major domains and the two conserved trypsin cleavage sites. (B to D) Shown are some norovirus representatives of genogroups I (B), II (C), III (D), and IV (E). Boldface letters indicate the trypsin cleavage sites and the conserved arginine (R) cluster. Arrows indicate actual trypsin cleavage sites. Numbers show the position of the last amino acid of the corresponding capsid proteins. Roman numerals indicate the genogroups.

Amersham Biosciences, Piscataway, NJ) powered by an AKTA FPLC System (model 920; Amersham Biosciences, Piscataway, NJ). The proteins of each peak were analyzed by SDS-PAGE and/or Western blot analysis. The molecular weights of proteins in each fraction were calibrated with the Gel Filtration Calibration kit (Amersham Biosciences, Piscataway, NJ). Gel filtrations were run using 1× PBS, pH 7.4, except where otherwise indicated.

Anion-exchange chromatography. A Q Sepharose high-performance anion exchanger (Amersham Biosciences, Piscataway, NJ) was packed in a Flex column with a flow adapter (Kontes, Vineland, NJ) powered by an AKTA FPLC System (model 920; Amersham Biosciences, Piscataway, NJ). The P proteins purified by an affinity column or gel filtration were adsorbed to the anion exchanger in a binding buffer (20 mM Tris-HCl, pH 8.0). The proteins were then eluted by an increasing salt gradient of the elution buffer (1 M NaCl, 20 mM Tris-HCl, pH 8.0). The proteins from each peak were analyzed by SDS-PAGE and/or Western blot analysis.

Assay of capsid proteins binding to HBGAs. The binding capabilities of wild-type and mutant capsid proteins to HBGAs were measured by saliva-binding assays as described previously (14, 15). The blood types of the saliva donors whose saliva samples were used in this study were determined in previous studies (14, 15).

Binding affinity of norovirus capsid protein to HBGA receptors. The binding affinity of the recombinant capsid protein was determined by an endpoint dilution method (32). In a standard saliva-binding assay (see above), the recombinant capsid proteins were diluted serially to reach a cutoff point of an optical density at 450 nm of ≥ 0.1 . For direct comparison, all recombinant P proteins were adjusted to a starting concentration of 1 mg/ml. Baculovirus-expressed VLPs of VA387 were included in each plate as an internal standard.

Mass spectrometry. The molecular weights of P proteins before and after trypsin cleavage were determined by electrospray ionization-mass spectrometry and/or matrix-assisted laser desorption ionization-mass spectrometry (MALDI-MS), which were performed by the core facility at the University of Cincinnati, Ohio. For electrospray ionization-mass spectrometry, P proteins were diluted in 50% acetonitrile-water, 0.1% formic acid buffer and analyzed through a Micro-mass Q-TOF II mass spectrometer (Waters Corporation, Milford, MA). The sample was directly infused through the instrument using a syringe pump. MALDI-MS was performed through a standard protein analysis protocol using Sinapinic acid as a matrix. The MALDI-MS instrument was a Bruker Daltonics Reflex IV (Billerica, MA).

RESULTS

A highly conserved arginine cluster at the C terminus of the P domain is important for receptor binding. Our previous studies have demonstrated that the P domain of norovirus can function independently to form a dimer and/or a subviral P particle and to bind to HBGA receptors in patterns similar to those of their parental VLPs (29, 32). We also found that a deletion at the C terminus of the norovirus capsid or P proteins significantly affected their abilities of receptor binding (29). Multiple-sequence alignment of norovirus capsid proteins revealed a highly conserved arginine (R) cluster at the C terminus of the P

TABLE 2. Summary of the P mutants with deletions and substitutions at the C terminus and their receptor-binding capabilities

Genogroup and P mutant ^a	P particle and/or dimer formation ^b	Receptor-binding capability for saliva type ^c :		
		A	B	O
VA387				
P (wild type)	P particle and dimer ^d	+++	+++	
PC-1	ND	++	++	
PC-2	Dimer	++	++	
PC-3	Dimer	—	—	
GST-PC-3	ND	+++	+++	
PC-4	Dimer	—	—	
GST-PC-4	ND	+++	+++	
PC-5	Dimer	—	—	
PC-12	ND	—	—	
GST-P-N-2-C-5	ND	+++	+++	
P-T ₃₃₈ A	Dimer	—	—	
GST-P-T ₃₃₈ A	ND	—	—	
GST-CNGRC-P-T ₃₃₈ A	ND	—	—	
CNGRC-P-T ₃₃₈ A	95% P particle	—	—	
P-RRAL	Dimer	—	—	
P-RARAL	Dimer	—	—	
P-ARRAL	Dimer	++	++	
P-HHHAAL-CDCRGDCFC	~5% P particle	++	++	
Norwalk virus				
P NV	Dimer	++		++
PC-2 NV	Dimer	—		—

^a P mutants PC-1 to PC-12 have deletions of the last 1 to 12 amino acids, respectively; mutant P-N-2-C-5 is PC-5 with an additional deletion of the first 2 amino acids of the P domain, and this mutant could not be released from the GST by thrombin.

^b Determined by gel filtration. ND, not determined.

^c A, B, and O represent A-, B-, and O-type saliva, respectively. The number of plus signs represents the relative binding capability.

^d P particle formation of this mutant is P protein concentration dependent (32).

FIG. 2. Loss of receptor-binding activity of the P mutants without the R cluster can be compensated for by a GST tag. (A and B) Shown are receptor-binding curves of the P mutants PC-4 (A) and H/P-T₃₃₈A (B) before (left panels) and after (right panels) thrombin cleavage. The letters O, A, B, and N represent saliva with blood types O, A, and B and saliva nonsecretor, respectively. OD, optical density.

domain (Fig. 1), with genogroup-specific motifs of “R(G)RL-GXRRX” for genogroup I and “RRRXQ/L” for genogroup II. To test if the positively charged R cluster plays a role in the structure and function of P proteins, a series of P mutants with deletions and substitutions targeting the R clusters and their vicinity has been constructed (Table 2).

When one to all three arginines were deleted in VA387 (a genogroup II virus), the resulting P mutants lost their binding to all A, B, and H antigens completely when human saliva samples were used in the receptor-binding assay (PC-3, PC-4, and PC-5; data not shown). Deletion of the last 12 amino acids (PC-12) showed the same result. On the other hand, deletion

FIG. 3. Trypsin cleaves at both ends of the full-length capsid protein and the P mutants. (A and B) SDS-PAGE stained with brilliant blue (G-250) shows various sizes of the full-length capsid protein (rVA387) and the P mutants with different tags (indicated in panel C) before (A) and after (B) trypsin digestion (125 µg/ml for 2 h). Note that all proteins in panel B (P polypeptide) showed the same size of ~34 kDa. IXE indicates that this P protein was purified by ion exchange. (C) The schematic graphs show the linear structures of the capsid protein and various P mutants. The underlined sequence is the hinge. GS is from the BamHI site of the expression vector pGEX-4T-1.

TABLE 3. Molecular masses of norovirus capsid proteins and P mutants before and after trypsin digestion

Genogroup and capsid protein or P mutant	Calculated molecular mass (Da)	Molecular mass determined by mass spectrometry (Da) ^a	Molecular mass determined by mass spectrometry after trypsin digestion (Da) ^a
VA387			
Capsid protein	58,886.6	Not determined	34,202.9
H/P	36,084.6	36,065.8	34,293.7; 1,289.1 (electrospray)
P	35,201.5	35,170.2 (electrospray)	34,272.7; 34,257.3 (electrospray)
P-CDCRGDCFC	36,203.3	36,198.5	34,260.4
H/P-CDCRGDCFC	37,043.5	37,088.6	34,263.3; 1,289.1 (electrospray)
GSFLVPPTVESR ^b	1,287.5		
Norwalk virus			
Capsid protein	56,670.4	Not determined	31,302.9
H/P	33,596.7	Not determined	31,309.2
CNGRC-H/P	34,130.6	Not determined	31,311.4

^a Molecular masses were determined by MALDI-mass spectrometry unless indicated otherwise.
^b The trypsin-released N-terminal sequence of the H/P mutants. GS is two extra amino acids from the vector pGEX-4T-1 (BamHI site). The underlined sequence is the hinge.

of the last (alanine) or last two (alanine-leucine) amino acids of the P domain did not result in significant reduction of receptor binding (PC-1 and PC-2; Table 2). To further determine the role of individual arginines in the receptor binding of the P protein, additional mutants with single alanine substitutions in the R cluster were studied. The results showed that neither of the last two arginines was dispensable for the receptor-binding function (P-RRAAL and P-RARAL; Table 2), whereas the replacement of the first arginine alone did not significantly affect the binding (P-ARRAL). Deletion of the

two arginines at the end of Norwalk virus (NV; a genogroup I virus) P domain also resulted in loss of receptor binding (PC-2 NV; Table 2). These data indicated that the R cluster is required for the receptor-binding function of P proteins, which is likely a common feature of noroviruses.
The R cluster does not participate directly in receptor binding. In characterizing the above P mutants, we have observed that the otherwise inactive mutants PC-3 and PC-4 were binding strongly to HBGAs before the GST tag was removed (GST-PC-3 and GST-PC-4; Table 2; Fig. 2). The same results

FIG. 4. Molecular masses of the P polypeptide of VA387 determined by mass spectrometry. (A) The schematic graphs show the linear structures of the H/P mutant, its possible tryptic fragments with calculated molecular masses, and the results of mass spectrometry (also see Table 2). (B and C) Shown are molecular mass determinations of the P polypeptides from VLP (B) or from H/P mutant (C) by MALDI-MS. (D) Shown is the molecular mass of the trypsin-released N-terminal peptide of the H/P mutant determined by electrospray ionization-MS.

were also obtained for P-N-2-C-5, a P mutant with deletion of the first two and the last five amino acids of the P domain (Table 2). However, the GST compensation of the impaired binding activities did not occur in two mutants with damaged receptor-binding interfaces (29, 30, 32) (P-T₃₃₈A and CNGRC-T₃₃₈A; Table 2; Fig. 2). Thus, we conclude that the R cluster is not directly involved in the receptor-binding interface; instead, the deletion or replacement of the arginines may affect the binding by changing the charge distribution within the P domain and/or affect the conformation of the receptor-binding interface. In fact, a P mutant with a replacement of the R cluster with three histidines (P-HHHAL-CDCRGDCFC) maintained the receptor-binding ability (Table 2), consistent with our hypothesis (see below).

The R cluster is required for P particle formation of the P protein. Our previous study demonstrated that an isolated P domain can form a P particle with high receptor-binding activity (32). To examine if the R cluster also affects P particle formation, mutant P proteins were characterized for their ability to form P particles by gel filtration. The results showed that none of three P mutants, with deletions of partial or the entire R cluster, formed P particles (Table 2, PC-3 to PC-5), even at a high protein concentration (5 mg/ml; data not shown). On the other hand, the P-HHHAL-CDCRGDCFC mutant formed P particles with a low efficiency (~5%; Table 2). These results indicated that the R cluster is required for P particle formation.

The R cluster is sensitive to trypsin digestion. It has been shown that an ~32-kDa soluble P protein (P polypeptide) exists in stools of Norwalk virus-infected patients (9, 11), and the N terminus of this protein has been mapped to a highly conserved trypsin digestion site near the beginning of the P domain (Fig. 1) (11). Sequence analysis revealed many other trypsin digestion sites in the P domain, including three to four sites within the R cluster (Fig. 1) (data not shown). Since the R cluster is important for receptor binding and P particle formation, we examined if these trypsin sites within the R cluster are functional in vitro. When the VA387 VLP and six P mutants with different N- or C-terminal-fused tags were treated by trypsin, all resulted in a single major polypeptide with a similar size of ~34 kDa (Fig. 3; Table 3), indicating that trypsin cleaved these proteins at the same sites on both ends of the P domain. Mass spectrometry analysis revealed that all these trypsin-trimmed proteins had the same molecular mass of 34.3 kDa (Fig. 4; Table 3). Sequence inspection indicated that the cleavages should occur at the conserved N-terminal trypsin site (R-T; Fig. 1) (11) and the first trypsin site (R-R) of the R cluster (Fig. 1 and 4), leaving a P polypeptide of 312 amino acids with a calculated molecular mass of 34.316 kDa (Fig. 4). The accuracy of the mass spectrometry analysis was also confirmed by detection (1,287.5 Da) of the released N-terminal peptide (GSFLVPPTVESR) with the predicted molecular mass of 1,291.1 Da (Fig. 4A, C, and D; Table 3).

A similar C terminus trypsin cleavage also occurred in a genogroup I virus, i.e., the prototype Norwalk virus (NV). When the NV VLPs and two NV P mutants (H/P and CNGRC-H/P) were digested with trypsin, a P polypeptide of ~32 kDa was observed (Fig. 5A; Table 3) by SDS-PAGE, although the NV VLP was less sensitive to trypsin digestion than that of VA387, and partial digestion of the NV capsid

FIG. 5. P polypeptide of Norwalk virus. (A) SDS-PAGE stained with brilliant blue (G-250) showed the trypsin-digested Norwalk virus (NV) VLPs, the H/P mutant, and VA387 VLPs. Sterically labeled protein was a major copurified bacterial protein. Proteins with (+) or without (-) trypsin digestion are indicated. (B) The schematic graph shows the linear structure of the NV H/P mutant and its predicted P polypeptide with calculated molecular masses. (C and D) Molecular masses of the P polypeptides from NV VLPs (C) and the NV H/P mutant (D) determined by MALDI-mass spectrometry. (E and F) Shown are chromatographs of gel filtrations of NV (E) and VA387 (F) VLPs. The gel filtration column was calibrated by a calibration kit (Amersham Biosciences) as follows: 2,000 kDa (void), 44.3 ml; 669 kDa, 50.1 ml; 440 kDa, 56.2 ml; 232 kDa, 65.7 ml; 146 kDa, 70.8 ml; 67 kDa, 76.7 ml; 47 kDa, 82.7 ml; 25 kDa, 92.9 ml; and 13.7 kDa, 97.3 ml. AU, absorbance unit.

protein was seen (Fig. 5A). Mass spectrometry analysis showed that these trypsin-released P polypeptides had an apparent molecular mass of 31.3 kDa (Fig. 5C; Table 3), indicating that trypsin also cleaved at the first trypsin site of the R cluster (R-G; Fig. 1B and 5B). These data suggest that the proteolytic cleavage of the R cluster may be a common feature of noroviruses. To confirm the forms of the VLPs used in this study, an aliquot of the VLPs was loaded on a gel filtration column of Superdex 200. A single peak at void volume (>800 kDa) for

FIG. 6. Trypsin digestion ablates receptor-binding capability and P particle formation of the P domain. (A and B) Shown are receptor-binding curves of the mutants P (A) and P-CNGRC (B) before (left panels) and after (middle panels) trypsin digestion (125 $\mu\text{g/ml}$ for 2 h). The letters O, A, B, and N represent saliva with blood types O, A, and B and saliva nonsecretor, respectively. (C) Chromatographs of gel filtrations of CNGRC-H/P mutant before (left) and after (middle) trypsin digestion show a shift of the major P protein peak. The minor peak in the void volume (~ 45 ml) represents some minor copurified bacteria proteins that do not affect our experiment (32). The gel filtration column was calibrated as described in the legend to Fig. 5. The right panels are the SDS-PAGE stained with brilliant blue (G-250) showing the result of the trypsin digestion. Au, absorbance unit.

both NV and VA387 VLPs was seen (Fig. 5E and F), indicating that the VLPs were not dissociated at the time they were digested by trypsin.

Trypsin-treated P proteins lost their abilities of receptor binding and P particle formation. Consistent with the studies of P mutants with deletions and substitutions of the R cluster (Table 2), the trypsin-treated P mutants of both VA387 and NV lost their abilities of receptor binding and particle formation (Fig. 6; Table 4; see mutants P-CNGRC and P-CGGC). It was noted that the P-CDCRGDCFC mutant is the only mutant that only partially lost its abilities of receptor binding and P particle formation (Fig. 7) (see below). A trypsin treatment of NV or VA387 VLPs also led to a partial loss of their receptor-binding activities (Table 4) due to an incomplete digestion of the VLPs, as shown by SDS-PAGE (Fig. 5A) and Western blot analysis (data not shown).

Structural requirement for specific trypsin cleavage. Previous studies showed that the NV VLP was not susceptible to trypsin digestion unless it was denatured under certain conditions (11). In the current study, we found that VA387 VLP is more sensitive to trypsin than NV VLP in $1\times$ PBS (pH 7.4), and in both cases at least a small portion of VLPs was resistant to trypsin digestion (Fig. 5) (data not shown). Furthermore,

during the treatment of the P particles from different P mutants (Table 4), we found that $\sim 50\%$ of the P-CDCRGDCFC (VA387) protein was resistant to trypsin cleavage, even with an extended incubation time of 24 h (Fig. 7A). The uncleaved P-CDCRGDCFC protein retained the ability to form P particles and bind to HBGA receptors, while the cleaved protein lost its binding to HBGA (Fig. 7). These data suggested that the trypsin sensitivity of VLPs and P particles is structure/conformation dependent, which may be affected by the environmental conditions. Such conditional cleavages may be an important regulatory factor in the viral replication by generating different forms of the capsid and/or the P protein and even the virions.

DISCUSSION

In this study, we demonstrated that a highly conserved R cluster at the C terminus of the norovirus capsid protein plays an important role in the structure and function of the viral capsid. More importantly, we also found that this R cluster is sensitive to proteolytic cleavage. A treatment of norovirus capsid proteins with trypsin resulted in removal of the S domain (9, 11, and this study) and the majority of the

TABLE 4. Trypsin treatment ablated the abilities of receptor binding and P particle formation of the P domain mutants

Genotype and capsid protein or P domain mutant ^a	P particle and/or dimer formation ^b	Receptor-binding affinity for saliva type ^c :		
		A	B	O
VA387				
VLP (VA387)	VLP	++++	++++	
VLP + trypsin	Not determined	++	++	
P	Dimer/P particle ^d	+++	+++	
P + trypsin	Not determined	—	—	
H/P	Dimer	++	++	
H/P + trypsin	Not determined	—	—	
P-CNGRC	P particle	++++	++++	
P-CNGRC + trypsin	Dimer	—	—	
P-CGGGC	P particle	++++	++++	
P-CGGGC + trypsin	Dimer	—	—	
H/P-CDCRGDCFC	P particle	++++	++++	
H/P-CDCRGDCFC + trypsin	Dimer	+	+	
P-CDCRGDCFC	P particle	++++	++++	
P-CDCRGDCFC + trypsin	Dimer/P particle	+++	+++	
Norwalk virus				
H/P	Dimer	++		++
H/P + trypsin	Not determined	—		—
CNGRC-H/P	P particle	+++		+++
CNGRC-H/P + trypsin	Not determined	—		—

^a Proteins were treated by trypsin at a concentration of 250 µg/ml for 2 h.^b Determined by gel filtration.^c A, B, and O represent A-, B-, and O-type saliva, respectively. The number of plus signs represents the relative binding affinity.^d P particle formation of this mutant is P protein concentration dependent (32).

R cluster, leaving an ~34.3-kDa (VA387) or an ~31.3-kDa (Norwalk virus) P polypeptide that lost the ability to form P particles and bind HBGA receptors. Proteolytic digestion of the prototype Norwalk viral capsid protein has been shown in vivo (9, 11), although direct evidence of such proteolytic cleavage of the R cluster remains lacking. Sequence analysis revealed many potential proteolytic cleavage sites across the entire capsid protein, including ~15 trypsin recognition sites within the P domain. However, proteolytic cleavage only occurs in the S domain and at the ends of the P domain (Fig. 1). These results raise a question about the role of the proteolytic cleavage in norovirus replication and/or pathogenesis.

Noroviruses replicate and cause diseases mainly in the intestine, which is rich in proteinases. One simple explanation of the role of proteolysis in this context is that it removes the extra unassembled capsid proteins during viral replication, which otherwise might interfere with the progeny viruses attaching to host cells to initiate a new infection. It is known that noroviruses contain a subgenomic RNA responsible for synthesis of the viral structural proteins. The 5' end of the subgenomic RNA shares the regulatory elements with the genomic RNA, possibly for high levels of transcription and translation (10). Thus, the proteolytic cleavage could be a necessary step to remove those over-produced viral capsid proteins. This process could occur within the infected enterocytes or in the lumen of the intestine. Proteinases may digest only unassembled capsid proteins (11) or empty VLPs without viral RNA.

The above hypothesis, however, does not explain why the

proteolysis of the capsid protein was incomplete, leaving the majority of the P domain (P polypeptide) intact. The high level of conservation of the two trypsin digestion sites (Fig. 1) that are responsible for producing the P polypeptide indicates this selective proteolytic process is a general feature of different noroviruses. Thus, the nondigested 31- to 34-kDa P polypeptides are likely to have a biologic function in the replication, host immune response, and/or pathogenesis of norovirus. One possibility is that this soluble P polypeptide serves as a decoy for the progeny viruses to escape from the host immunity (Fig. 8). Evidence supporting this hypothesis include the following points: (i) the P polypeptide does not bind host receptors, therefore it will not compete with the progeny viruses for receptors in a new infection; (ii) the P polypeptide is no longer sensitive to proteolytic digestion (9, 11, and this report), and therefore it is able to provide a persistent decoy effect; (iii) in spite of the loss in receptor-binding ability, the P polypeptide remains highly antigenic (M. Tan and X. Jiang, unpublished data); and (iv) the P polypeptide does not form P particles, therefore it would have a higher molar ratio than the P particle- or VLP-forming protein as a decoy against host immunity (see Fig. 8).

The P polypeptide has been shown to be the major product of the capsid protein, in addition to the authentic virus, after passing through the intestine (9, 11). However, we could not exclude the possibility of other forms of the P proteins in the earlier stages of viral replication and/or pathogenesis, particularly during viral RNA and protein synthesis and virion assembly inside the enterocytes before release into the intestinal lumen. In fact, partial digestion of norovirus VLPs has been observed in vitro in this study (Fig. 5). The numbers and types of proteases inside the enterocytes are expected to be fewer compared with those in the intestinal lumen. Thus, selective or partial proteolytic cleavage of capsid proteins could occur, which may result in different forms of the capsid protein with or without function.

A previous study of Norwalk virus VLP suggested that trypsin may cleave only dissociated, soluble capsid protein, leaving the well-formed VLP intact (11). This result is consistent with the hypotheses described above. However, we found in this study that trypsin can cleave the Norwalk and VA387 VLPs isolated from the sucrose gradient, but the digestion of Norwalk VLP was incomplete even after an extended incubation of 16 h (Fig. 5) (data not shown). Similar partial digestion was also observed for P particles; about half of the VA387 P particle of P-CDCRGDCFC (32) was never cut by trypsin (Fig. 7). These data indicated that there could be two or more forms of VLPs, or P particles in the case of the P-CDCRGDCFC mutant, differing in their susceptibility to trypsin. Since such putative sensitive and nonsensitive forms of the studied capsid and P proteins share the same primary sequences, their differences should be attributed to their tertiary or quaternary structure, which could be in turn environment dependent. At present it remains unknown what conditions favor the structure formation that is susceptible to trypsin digestion, but the mechanism of conditional cleavage of the capsid protein/VLP may help to explain the hypotheses proposed above.

In summary, this study further extended our understanding of the structure-function relationship of norovirus capsid pro-

FIG. 7. P particle of P-CDCRGDCFC was partially resistant to trypsin cleavage. (A) SDS-PAGE stained with brilliant blue (G-250) shows partial cleavage of the P-CDCRGDCFC mutant by trypsin (125 μ g/ml for 3 h), while the same trypsin cut other P mutants (PCAGAC, P-CNGRC, and P-CGGGC) completely. The P mutants incubated with (+) or without (-) trypsin are indicated. (B) Shown are receptor-binding curves of the P-CDCRGDCFC mutant before (left) and after (right) trypsin digestion. The letters O, A, B, and N represent saliva with blood types O, A, B, and B and saliva nonsecretor, respectively. (C) Chromatographs of gel filtrations of the P-CDCRGDCFC mutant before (left) and after (right) trypsin digestion show a partial shift of the major P protein peaks. The minor peak in the void volume at $\sim$ 45 ml represents some copurified bacteria proteins (32). The gel filtration column was calibrated as described in the legend to Fig. 5. (D) Shown are receptor-binding activities of the two peaks representing the P particle of P-CDCRGDCFC and its P polypeptide, respectively. The two peaks are from panel C. The peaks of $\sim$ 45 ml did not show detectable receptor-binding activity. The letters A and B represent saliva with blood types A and B, respectively. OD, optical density; AU, absorbance unit.

teins. The critical role of the R cluster in the receptor binding and the proteolytic inactivation of this function represent an interesting model of how noroviruses could adapt to the host and further take advantage of the host resources for their replication and survival in stages of attachment, penetration, releasing of the viral genome, spread of viruses from host to host (pathogenesis), and escaping from host immunity. Understanding of this model may provide insight not only into norovirus but also other viruses in general. However, many issues need to be resolved before we can fully elucidate this mechanism. For example, genogroup-specific motifs of the R clusters have been found (Fig. 1), raising the question of whether these

motifs are genogroup specific in functional terms. In other words, are these motifs interchangeable between genogroups? In addition, the conditions favoring proteolytic cleavage of the VLPs and the P particles, as well as the location and stages of the cleavage, are critical for further understanding of viral replication and pathogenesis. Finally, norovirus P particles could be potentially used as antivirals or vaccines against norovirus diseases because they are easily produced, with high receptor-binding capability (32), and are highly immunogenic (Tan and Jiang, unpublished). Studies to further characterize the mechanism of VLP and P particle formation may help to design candidates for these purposes.

FIG. 8. Possible roles of the P polypeptide in vivo. Norovirus capsids or capsid proteins are cleaved by trypsin into P polypeptides (P). These P polypeptides do not form P particles and do not bind to HBGA receptors, and therefore they do not compete for viral receptors with authentic viruses. On the other hand, these P polypeptides remain antigenic (data not shown) and may function as decoys to protect authentic viruses from host immunity. In addition, the P polypeptides may have other unknown functions.

ACKNOWLEDGMENTS

The research described in this article was supported by the National Institute of Allergy and Infectious Diseases (R01 AI37093 and R01 AI55649).

We thank Stephen Macha at the core facility of mass spectrometry at the University of Cincinnati for the molecular mass measurements of the capsid protein and sequence analyses. We also thank Brad Jennings, Weiming Zhong, and Shale Amsalu for technical assistance and Tibor Farkas and Chao Wei for helpful discussions.

REFERENCES

- Berke, T., and D. O. Matson. 2000. Reclassification of the Caliciviridae into distinct genera and exclusion of hepatitis E virus from the family on the basis of comparative phylogenetic analysis. *Arch. Virol.* **145**:1421–1436.
- Bertolotti-Ciarlet, A., L. J. White, R. Chen, B. V. Prasad, and M. K. Estes. 2002. Structural requirements for the assembly of Norwalk virus-like particles. *J. Virol.* **76**:4044–4055.
- Chakravarty, S., A. M. Hutson, M. K. Estes, and B. V. Prasad. 2005. Evolutionary trace residues in noroviruses: importance in receptor binding, antigenicity, virion assembly, and strain diversity. *J. Virol.* **79**:554–568.
- Fankhauser, R. L., S. S. Monroe, J. S. Noel, C. D. Humphrey, J. S. Bresee, U. D. Parashar, T. Ando, and R. I. Glass. 2002. Epidemiologic and molecular trends of "Norwalk-like viruses" associated with outbreaks of gastroenteritis in the United States. *J. Infect. Dis.* **186**:1–7.
- Glass, R. I., J. Bresee, B. Jiang, J. Gentsch, T. Ando, R. Fankhauser, J. Noel, U. Parashar, B. Rosen, and S. S. Monroe. 2001. Gastroenteritis viruses: an overview. *Novartis Found. Symp.* **238**:5–25.
- Graham, D. Y., X. Jiang, T. Tanaka, A. R. Opekun, H. P. Madore, and M. K. Estes. 1994. Norwalk virus infection of volunteers: new insights based on improved assays. *J. Infect. Dis.* **170**:34–43.
- Green, J., J. Vinje, C. I. Gallimore, M. Koopmans, A. Hale, D. W. Brown, J. C. Clegg, and J. Chamberlain. 2000. Capsid protein diversity among Norwalk-like viruses. *Virus Genes* **20**:227–236.
- Green, K. Y., T. Ando, M. S. Balayan, T. Berke, I. N. Clarke, M. K. Estes, D. O. Matson, S. Nakata, J. D. Neill, M. J. Studdert, and H. J. Thiel. 2000. Taxonomy of the caliciviruses. *J. Infect. Dis.* **181**(Suppl. 2):S322–S330.
- Greenberg, H. B., J. R. Valdesuso, A. R. Kalica, R. G. Wyatt, V. J. McAuliffe, A. Z. Kapikian, and R. M. Chanock. 1981. Proteins of Norwalk virus. *J. Virol.* **37**:994–999.
- Hardy, M. E., and M. K. Estes. 1996. Completion of the Norwalk virus genome sequence. *Virus Genes* **12**:287–290.
- Hardy, M. E., L. J. White, J. M. Ball, and M. K. Estes. 1995. Specific proteolytic cleavage of recombinant Norwalk virus capsid protein. *J. Virol.* **69**:1693–1698.
- Harrington, P. R., L. Lindesmith, B. Yount, C. L. Moe, and R. S. Baric. 2002. Binding of Norwalk virus-like particles to ABH histo-blood group antigens is blocked by antisera from infected human volunteers or experimentally vaccinated mice. *J. Virol.* **76**:12335–12343.
- Harrington, P. R., J. Vinje, C. L. Moe, and R. S. Baric. 2004. Norovirus capture with histo-blood group antigens reveals novel virus-ligand interactions. *J. Virol.* **78**:3035–3045.
- Huang, P., T. Farkas, S. Marionneau, W. Zhong, N. Ruvoen-Clouet, A. L. Morrow, M. Altaye, L. K. Pickering, D. S. Newburg, J. LePendou, and X. Jiang. 2003. Noroviruses bind to human ABO, Lewis, and secretor histo-blood group antigens: identification of 4 distinct strain-specific patterns. *J. Infect. Dis.* **188**:19–31.
- Huang, P., T. Farkas, W. Zhong, M. Tan, S. Thornton, A. L. Morrow, and X. Jiang. 2005. Norovirus and histo-blood group antigens: demonstration of a wide spectrum of strain specificities and classification of two major binding groups among multiple binding patterns. *J. Virol.* **79**:6714–6722.
- Hutson, A. M., F. Airaud, J. LePendou, M. K. Estes, and R. L. Atmar. 2005. Norwalk virus infection associates with secretor status genotyped from sera. *J. Med. Virol.* **77**:116–120.
- Hutson, A. M., R. L. Atmar, and M. K. Estes. 2004. Norovirus disease: changing epidemiology and host susceptibility factors. *Trends Microbiol.* **12**:279–287.
- Hutson, A. M., R. L. Atmar, D. Y. Graham, and M. K. Estes. 2002. Norwalk virus infection and disease is associated with ABO histo-blood group type. *J. Infect. Dis.* **185**:1335–1337.
- Hutson, A. M., R. L. Atmar, D. M. Marcus, and M. K. Estes. 2003. Norwalk virus-like particle hemagglutination by binding to histo-blood group antigens. *J. Virol.* **77**:405–415.
- Hutson, A. M., S. Chakravarty, R. L. Atmar, B. V. Prasad, and M. K. Estes. 2004. Loss of carbohydrate binding with point mutations of Norwalk virus virus-like particles. Second International Calicivirus Conference, Dijon, France.
- Jiang, X., M. Wang, D. Y. Graham, and M. K. Estes. 1992. Expression, self-assembly, and antigenicity of the Norwalk virus capsid protein. *J. Virol.* **66**:6527–6532.
- Lindesmith, L., C. Moe, S. Marionneau, N. Ruvoen, X. Jiang, L. Lindblad, P. Stewart, J. LePendou, and R. Baric. 2003. Human susceptibility and resistance to Norwalk virus infection. *Nat. Med.* **9**:548–553.
- Marionneau, S., A. Cailleau-Thomas, J. Rocher, B. Le Moullac-Vaidye, N. Ruvoen, M. Clement, and J. Le Pendu. 2001. ABH and Lewis histo-blood group antigens, a model for the meaning of oligosaccharide diversity in the face of a changing world. *Biochimie* **83**:565–573.
- Marionneau, S., N. Ruvoen, B. Le Moullac-Vaidye, M. Clement, A. Cailleau-Thomas, G. Ruiz-Palacios, P. Huang, X. Jiang, and J. Le Pendu. 2002. Norwalk virus binds to histo-blood group antigens present on gastrointestinal epithelial cells of secretor individuals. *Gastroenterology* **122**:1967–1977.
- Prasad, B. V., M. E. Hardy, T. Dokland, J. Bella, M. G. Rossmann, and M. K. Estes. 1999. X-ray crystallographic structure of the Norwalk virus capsid. *Science* **286**:287–290.
- Prasad, B. V., M. E. Hardy, X. Jiang, and M. K. Estes. 1996. Structure of Norwalk virus. *Arch. Virol. Suppl.* **12**:237–242.
- Prasad, B. V., R. Rothnagel, X. Jiang, and M. K. Estes. 1994. Three-dimensional structure of baculovirus-expressed Norwalk virus capsids. *J. Virol.* **68**:5117–5125.
- Ravn, V., and E. Dabelsteen. 2000. Tissue distribution of histo-blood group antigens. *APMIS* **108**:1–28.
- Tan, M., R. S. Hegde, and X. Jiang. 2004. The P domain of norovirus capsid protein forms dimer and binds to histo-blood group antigen receptors. *J. Virol.* **78**:6233–6242.
- Tan, M., P. Huang, J. Meller, W. Zhong, T. Farkas, and X. Jiang. 2003. Mutations within the P2 domain of norovirus capsid affect binding to human histo-blood group antigens: evidence for a binding pocket. *J. Virol.* **77**:12562–12571.
- Tan, M., and X. Jiang. 2005. Norovirus and its histo-blood group antigen receptors: an answer to a historical puzzle. *Trends Microbiol.* **13**:285–293.
- Tan, M., and X. Jiang. 2005. The P domain of norovirus capsid protein forms a subviral particle that binds to histo-blood group antigen receptors. *J. Virol.* **79**:14017–14030.
- Tan, M., W. Zhong, D. Song, S. Thornton, and X. Jiang. 2004. E. coli-expressed recombinant norovirus capsid proteins maintain authentic antigenicity and receptor binding capability. *J. Med. Virol.* **74**:641–649.