Evolutionary Trace Residues in Noroviruses: Importance in Receptor Binding, Antigenicity, Virion Assembly, and Strain Diversity

Sugoto Chakravarty,[1] Anne M. Hutson,[2] Mary K. Estes,[2] and B. V. Venkataram Prasad[1,3]*

Verna and Marrs Mclean Department of Biochemistry and Molecular Biology[1] and Keck Center for Computational Biology,[3] and Department of Molecular Virology and Microbiology,[2] Baylor College of Medicine, Houston, Texas

Noroviruses cause major epidemic gastroenteritis in humans. A large number of strains of these single-stranded RNA viruses have been reported. Due to the absence of infectious clones of noroviruses and the high sequence variability in their capsids, it has not been possible to identify functionally important residues in these capsids. Consequently, norovirus strain diversity is not understood on the basis of capsid functions, and the development of therapeutic compounds has been hampered. To determine functionally important residues in noroviruses, we have analyzed a number of norovirus capsid sequences in the context of the Norwalk virus capsid crystal structure by using the evolutionary trace method. This analysis has identified capsid protein residues that uniquely characterize different norovirus strains and provide new insights into capsid assembly and disassembly pathways and the strain diversity of these viruses. Such residues form specific three-dimensional clusters that may be of functional importance in noroviruses. One of these clusters includes residues known to participate in the proteolytic cleavage of these viruses at high pH. Other clusters are formed in capsid regions known to be important in the binding of antibodies to noroviruses, thereby indicating residues that may be important in the antigenicity of these viruses. The highly variable region of the capsid shows a distinct cluster whose residues may participate in norovirus-receptor interactions.

Norwalk-like caliciviruses (noroviruses) cause over 90% of epidemic nonbacterial gastroenteritis in humans. The single-stranded RNA genome of noroviruses is organized into three open reading frames (ORFs). ORF1 encodes a 200-kDa polyprotein that is processed into at least six nonstructural proteins, ORF2 encodes the 60-kDa capsid protein VP1, and ORF3 encodes the basic minor structural protein VP2 (10). Comparisons of the ORF1 and ORF2 nucleotide sequences show a wide genetic diversity in noroviruses (2, 10). A large number (>100) of norovirus strains have been sequenced all over the world. Distinguishing these strains has been a major effort in the epidemiology of noroviruses (8) that has resulted in their classification into two major genogroups, GI and GII, with each genogroup consisting of seven genetic clusters, and three arbitrarily classified minor genogroups, GIII, GIV, and GV (10, 21, 31, 43). Such classification has relied on comparisons of the genomic and capsid sequences of noroviruses. A more complete understanding of the sequence relatedness among the different norovirus strains will be possible if functionally important regions in these sequences are identified.

Little information is available about the molecular details of norovirus functions. Nucleoside triphosphatase and protease activities have been identified in two of the ORF1 proteins (5, 33, 39), while other enzymatic functions of the ORF1 proteins have mostly been deduced from their sequence similarities with the nonstructural proteins in picornaviruses. The experimental elucidation of VP1 functions and serotyping of noroviruses have not been possible because of the lack of propagation systems for these viruses. Cross-challenge studies

with human volunteers and infected patients, along with immunoelectron microscopy and solid-phase immunoelectron microscopy, were the only sources earlier to identify antigenic relationships between some of the native norovirus strains (20, 24). Despite these difficulties, empty recombinant capsids of noroviruses that show morphological and antigenic similarity with the native virions (18) have been used as surrogates of the virions for structural and functional studies including enzyme immunoassay-based antigenic studies on many different norovirus strains.

The capsids of noroviruses and all other caliciviruses are composed of a single predominant protein, VP1. Consequently, all major requirements for the assembly, receptor binding, host specificity, and antigenicity of noroviruses reside in VP1. Limited data on receptor binding and the antigenicity of noroviruses and only one norovirus crystal structure, that of the icosahedral T=3 recombinant Norwalk virus (rNV) capsid, are available. The rNV structure (Protein Data Bank [PDB] code 1IHM) shows that the VP1 subunit has a shell (S) and a protruding (P) domain consisting of a middle P1 and a distal P2 subdomain that is an insertion in the virus sequence (34). The S domain is primarily important in forming the icosahedral capsid shell (4). The P1 subdomain has been implicated in the antigenicity of noroviruses (11, 13), while the P2 subdomain has been suggested to bind to cellular receptors of these viruses (16). Available cryo-electron microscopy structural studies on various caliciviruses indicate that these viruses share a similar modular domain organization (7, 35).

The norovirus capsid protein sequences show high variability. The S domain sequences are 30% identical, while the sequence identity is only about 11 and 8% in the P1 and the P2 subdomains, respectively (this study). Such variability among norovirus sequences makes it difficult to use conventional sequence comparison methods to detect conservation patterns

* Corresponding author. Mailing address: Department of Biochemistry and Molecular Biology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030-3498. Phone: (713) 798-5686. Fax: (713) 798-1625. E-mail: vprasad@bcm.tmc.edu.

that may indicate possible functional sites on norovirus capsids. The evolutionary trace (ET) method (25) that exploits phylogenetic tree-based sequence comparisons along with crystal structure information has been successfully applied to detect functional sites in a wide variety of proteins (26, 28, 40). We have applied the ET method for the first time to a viral system to detect capsid protein residues that uniquely characterize different norovirus strains and that may be important in the assembly and function of these human pathogens.

MATERIALS AND METHODS

The complete amino acid sequences of the capsid protein of 56 different noroviruses (Fig. 1, sequences S1 to S56) including all prototype and antigenic strains (10), among others, were aligned by using ClustalW1.8 (42) with default parameters on the European Bioinformatics Institute server. The aligned sequences and the rNV coordinates were submitted to the Cambridge University ET server (17). By using a Phylip distance matrix computed from all the sequences, a phylogenetic tree was constructed (Fig. 1). The sequences on different branches of the tree were grouped into different evolutionary classes according to their degree of similarity. To generate these classes, an evolutionary time cutoff line first split the phylogenetic tree into 10 evenly distributed partitions, P01 to P10, in order of increasing divergence (Fig. 1). In a given partition, the sequences that originated from a common node on the phylogenetic tree and shared the evolutionary time cutoff line that created the partition formed a class. This ensured that the most similar sequences belonged to the same class while the more distant ones belonged to different classes. Ten partitions sufficed to avoid random inclusion of sequences in any class because no distinct class-specific surfaces were created after partition P10. In a given partition, sequences within different classes were separately aligned, and the resultant aligned classes were compared to obtain their consensus residues, called the trace residues, for that partition. Three types of trace residues were identified: those that remained invariant across all classes were designated absolutely conserved residues (ACRs), those that remained strictly conserved within a class but differed between various classes were designated class-specific residues, while trace residues that showed no conservation within any class were designated neutral (17, 25). The ACRs and the class-specific residues were mapped onto the rNV structure. Class-specific residues forming structural clusters in the vicinity of the ACRs were analyzed for possible biological significance by using the rNV crystal structure and available biochemical information due to the propensity of such residues to form functional sites in a large number of proteins (27, 44). Because trace residues could not be defined for single sequence branches, such branches were not considered as independent classes. The extent of exposure of the class-specific residues was computed, using cutoff values of 0.3 of the occluded surface packing indices of the atoms (32). Such calculations were done individually for each of the icosahedral interfaces to account for the different degree of exposure of a given residue located at such interfaces. In order to compare the ET-based classification of noroviruses with their known classification based on conventional phylogenetic analyses (10), 10 additional sequences belonging to genogroups GI through GV were included in the alignment and in the tree (Fig. 1). However, these additional sequences were not included in most of the subsequent ET analyses. A few sequences were included twice to ensure positive controls in the alignment and in the tree generation process.

RESULTS AND DISCUSSION

ET partition-dependent classes and genogroups. The different partitions P01 to P10 divide the phylogenetic tree into classes that vary with the partitions (Fig. 1). Individual partitions contain different numbers of classes (Table 1), where each class consists of a cluster of similar sequences originating from a given node within that partition. Nodes 1 and 2 create two branches in partition P01 (Fig. 1). One of these branches contains the murine norovirus sequence, while the other branch consists of the remaining norovirus sequences, indicating that the murine sequence may indeed constitute an independent and distinct norovirus genogroup (21). In partition P02, nodes 3 and 4 that diverge from node 1 distinguish the GI

sequences A1 to A20, referred to as class AG1, from the GII sequences A21 to A69, referred to as class AG2 (Fig. 1 and Table 1). Class AG1 includes the bovine sequences while class AG2 includes the swine and the Alphatron sequences (Table 1), indicating thereby that the bovine sequences are similar to GI sequences, while the swine and the Alphatron-like sequences are similar to GII sequences in partition P02. In partition P03, however, the bovine sequences branch off independently into class AG3 at node 6, while the remaining GI and GII sequences of partition P02 remain in clusters AG1a (sequences A1 to A17, corresponding to node 5) and AG2 (sequences A21 to A69 including the swine and the Alphatron, corresponding to node 4), respectively (Table 1 and Fig. 1). Thus, the bovine sequences are quite similar to the GI sequences only up to partition P02. In contrast, the swine and the Alphatron sequences bear a relatively stronger resemblance to the GII sequences because of their grouping together even in partition P03. It is only in the two nodes 7 and 8 of partitions P04 and P05 that the Alphatron-like sequences become distinct from the other genogroup II sequences (Table 1, class AG2b; Fig. 1). The AG2b class containing the Alphatron sequences remains unchanged beyond partition P05. Such partition-dependent similarities indicate that the bovine sequences may cluster in genogroup GI while the Alphatron sequences may cluster in genogroup GII, in contrast to the independent genogroup GIII and GIV status proposed previously for these sequences, respectively (31, 43).

Partition P06 creates further GI classes and distinguishes the swine sequences from other GII sequence classes. This partition creates two GI clusters, AG1b (A1 to A12) and AG1c (A13 to A18) from the AG1a cluster, and four GII clusters, AG2c (A21 to A35), AG2d (A36 to A40), AG2e (A41 to A65) and AG2f (A66 and A67), from the AG2a cluster (Table 1 and Fig. 1). Sequences in cluster AG1b are similar to the GI reference sequences NV (A1), Chiba (A4), Musgrove (A8), Southampton (SOV; A9), and the bs5_DE (A11), while sequences in cluster AG1c are similar to the GI reference sequences Desert Shield (DSV; A17) and Winchester (A18). The GII sequences A21 to A35 of cluster AG2c are similar to the Bristol virus (A32) reference sequence, while sequences A36 to A40 of cluster AG2d are similar to the Amsterdam (A38) and the Leeds (A39) reference sequences. The GII sequences A41 to A65 of cluster AG2e are similar to the Melksham (A42), Hillingdon (A44), Hawaii (HV; A54), Seacroft (A55), and the Toronto (A59) reference virus strains, while the swine sequences (A66 and A67) branch off into an independent GII class, AG2f in this partition (Table 1 and Fig. 1).

Further distinctions create new classes of similar sequences in partition P07 and beyond. Five new GI classes, AG1d to AG1h, and two new GII classes, AG2h and AG2i, are created in partition P07 (Table 1) where each new class contains at least one reference strain. The NV (A1) sequence shows similarity with the Japanese Aichi (A2) and KY89 (A3) strains, and these sequences branch off together into the AG1d class. The AG1e class consists of the Chiba (A4), Koblenz (A5), Valetta (A6), Thistlehall (A7), and the Musgrove (A8) strains. The AG1f class consists of the SOV (A9) and the Whiterose (A10) strains; the AG1g class consists of the bs5 (A11) and the Sindlesham (A12) strains, while the AG1h class consists of the
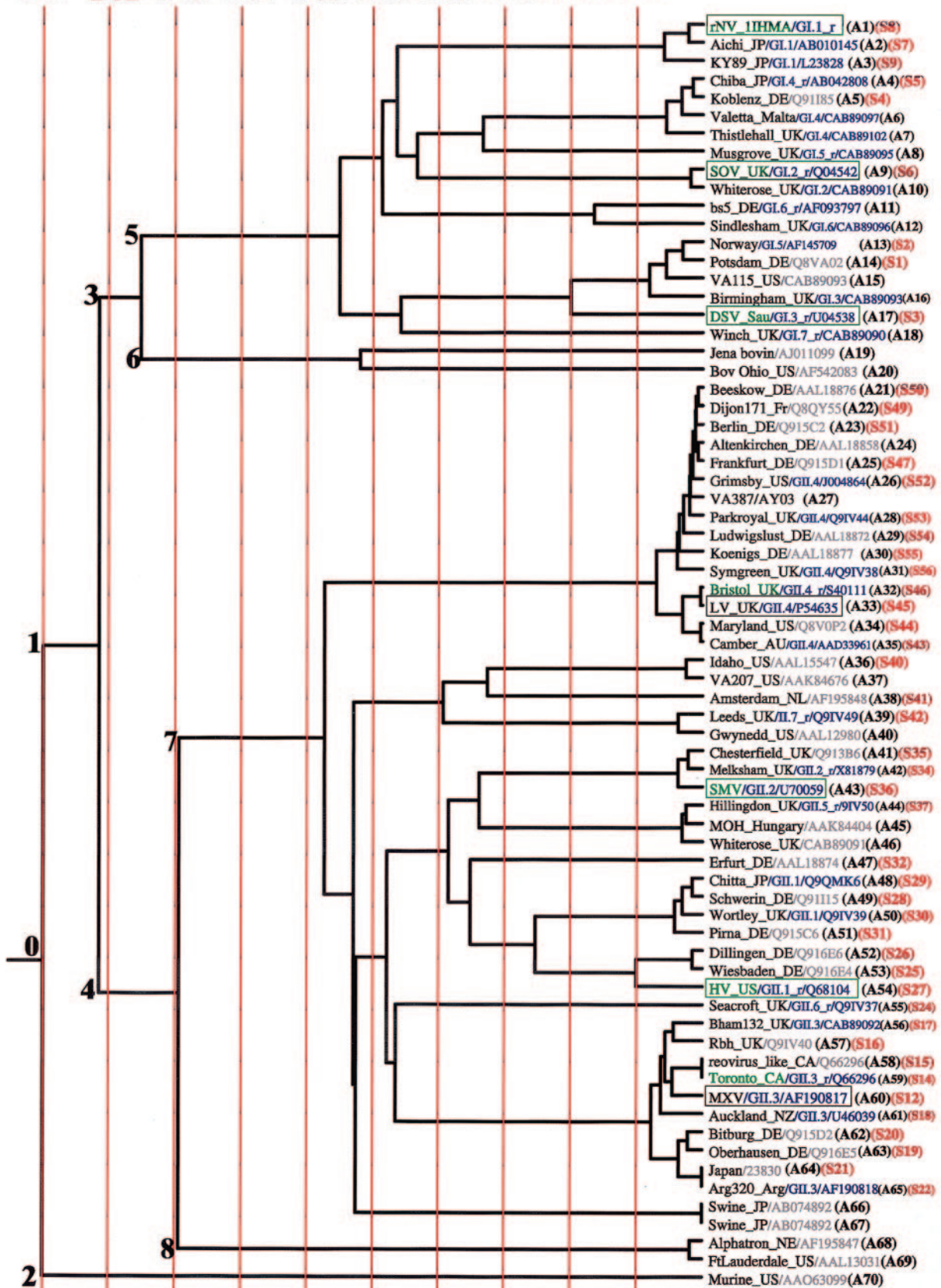
Norway (A13), Potsdam (A14), VA115 (A15), Birmingham (A16), and the DSV (A17) strains (Table 1). The AG2h class (A41 to A54) includes the Melksham (A42), Hillingdon (A44), and the HV (A54) reference strains, while the AG2i class (A56 to A65) includes the Toronto (A59) reference strains (Table 1). The swine strains remain in the AG2f class in all the remaining partitions (Table 1). One new GI class (AG1j) and five new GII classes (A2j, A2k, A2m, A2n, and A2o) are created in partition P08 from the classes AG1e, Ag2d, and AG2h of partition P07. The Idaho (A36) and VA207 (A37) strains are similar enough to occur in class AG2j, and Leeds (A39) and Gwynedd (A40) are also similar enough to occur in class AG2k of this partition. The AG2m class groups the similar Chesterfield (A41), Melksham (A42), and Snow-Mountain (SMV; A43) strains, while the Hillingdon (A44), MOH (A45), and Whiterose (A46) strains occur in the AG2n class. The Chitta (A48), Schwerin (A49), Wortley (A50), Pirna (A51), Dillingen (A52), Wiesbaden (A53), and HV (A54) strains are grouped into the AG2o class (Table 1 and Fig. 1). Partition P09 splits the AG2o class of partition P08 into two distinct classes, AG2q (A48 to A51) and class AG2r containing the A52 to A54 sequences. The other P08 classes remain unchanged in this partition. Similarly, partition P10 creates a new GII class, AG2s (Table 1 and Fig. 1). Thus, classes originate from distinct nodes in each partition. Such classes belonging to partitions P01 and P02 distinguish between the two major norovirus genogroups GI and GII, while classes in the remaining partitions P03 to P10 increasingly resolve the various sequence clusters within these genogroups (Table 1 and Fig. 1).

The ET classes (Table 1 and Fig. 1) correctly identify all known genetic clusters obtained from conventional phylogenetic analysis (10). Such clusters lie in different partitions because, unlike conventional phylogenetic analyses that determine the clusters by comparing all the sequences together by using arbitrary cutoff values of a sequence similarity index (2, 10), the ET classes are determined separately for each partition by comparing only those sequences that belong to the nodes originating in that partition. This results in well-defined, nonoverlapping ET classes unlike the conventional histogram-based cluster analyses that may show overlapping regions in the histogram (2). Besides, many new classes containing sequence similarities that may not be obvious in conventional phylogenetic analyses are created in the various ET partitions. Such similarities may be further understood from the conservation patterns of 56 of the GI and GII sequences (Fig. 1, S1 to S56) that form well-defined ET classes (Tables 1 and 2; Fig.

1). These results remain unchanged even when all the 70 sequences A1 to A70 (Fig. 1) are included in the analysis.

Partitioning of phylogenetic tree reveals hidden conservation patterns. The ET analysis reveals characteristic conservation patterns in the regions that appear to be variable in conventional sequence comparisons. This is a consequence of partitioning the phylogenetic tree (Fig. 1) and comparing the resultant sequence classes, in contrast to conventional sequence comparisons where all given sequences are compared together. In partition P01, all 56 sequences cluster into one class (Table 2). In partition P02, these sequences separate into the two classes GI and GII that correspond to the two major norovirus genogroups (Fig. 1). Class GI consists of sequences S1 to S10, while class GII consists of the sequences S12 to S56 (Table 2). No further classes are created in partitions P03 and P04. However, all subsequent partitions P05 to P10 create additional classes, each of which contains a number of sequence clusters (Table 2). Comparisons of classes in partitions P01 to P10 show interesting differences in their variable regions (Fig. 2A). Large variable regions are seen in the single class of partition P01, similar to results of a conventional comparison of all the 56 sequences taken together. However, when the sequences contained in classes GI and GII of partition P02 are separately aligned and the aligned classes are compared, class-specific and neutral trace residues emerge in the variable regions of partition P01. For example, residue 44 is variable if all 56 sequences are aligned together in the P01 partition, but it becomes a class-specific residue (X) in a comparison of the two aligned classes GI and GII in the P02 partition (Fig. 2A). This residue is a conserved Ala in class GI containing all the genogroup I norovirus sequences, while it is a conserved Pro in all the genogroup II norovirus sequences of class GII (Table 2). In contrast, residue 39S is neutral because it remains variable in both partitions P01 and P02, while residue 42A is an ACR because it is conserved in both of these partitions (Fig. 2A). Similar comparisons of classes in the subsequent P03 to P10 partitions reveal several class-specific residues in regions that appear to be highly variable in the preceding partitions (Fig. 2A). Thus, systematic comparisons of the different classes reveal class-specific conservation patterns that would otherwise be hidden if all the sequences were compared together.

Class-specific trace residues distinguish norovirus strains and may explain their antigenic diversity. The class-specific trace residues in different partitions uniquely distinguish different noroviruses including the antigenic strains. These include the NV (S8), SMV (S36), and the HV (S27) strains that

---

FIG. 1. The ET of the norovirus strains (A1 to A70) belonging to the two genogroups for the different partitions. A few randomly chosen duplicate sequences (A58/A59, A64/A65, and A66/A67) as positive controls in the sequence alignments and the tree construction appear correctly on the tree. The partition lines P01 to P10 shown in red indicate the extent of divergence (resolution) of the norovirus sequences. A subset, whose sequence numbers (S1 to S56) are shown in red, has been used for ET trace residue analysis. Sequences of known clusters in genogroups I and II are shown in blue as GI.n or GII.n, respectively, along with the appended "_r" if the sequence is a reference strain for the cluster (10). The EMBL and NCBI database accession numbers are shown following the strain names and cluster numbers if applicable. The strain names reflect the names of the places or regions where the strains were first isolated. The name of the country is shown as a two-letter code (except where the name is obvious) along with the strain name. These codes are as follows: AU, Australia; CA, Canada; DE, Germany; Fr, France; JP, Japan; NL, Netherlands; NZ, New Zealand; Sau, Saudi Arabia; UK, United Kingdom; US, United States. The prototype strains refer to the first strain that was isolated within a given cluster (10). The antigenic strains are shown within boxes, and the prototype strains are shown in green. Antigenic norovirus strains that are also prototypic are shown in green and are enclosed within boxes. The numbers (0 to 8) refer to the different nodes of the tree. Node 0 is the root node that is the parent node of child nodes 1 and 2. Similarly, node 1 is the parent node of child nodes 3 and 4, node 3 is the parent node of child nodes 5 and 6, and node 4 is the parent node of the child nodes 7 and 8.

TABLE 1. The evolutionary classes for partitions P01 to P10[a]

Partition	Evolutionary class		Sequences in class	Known norovirus reference sequence(s); contained within class[c]
	Total	ID[b]		
P01	1		A1–A70	A1, A4, A8, A9, A11, A17, A18, A32, A39, A42, A44, A54, A59
P02	3	AG1	A1–A20	A1, A4, A8, A9, A11, A17, A18
		AG2	A21–A69	A32, A39, A42, A44, A54, A59
P03	3	AG1a	A1–A18	A1, A4, A8, A9, A11, A17, A18
		AG3	A19, A20	
		AG2	A21–A69	A32, A39, A42, A44, A54, A59
P04, P05	4	AG1a	A1–A18	A1, A4, A8, A9, A11, A17
		AG3	A19, A20	
		AG2a	A21–A67	A32, A38, A39, A42, A44, A54, A55, A59
		AG2b	A68, A69	
P06	7	AG1b	A1–A12	A1, A4, A8, A9, A11
		AG1c	A13–A18	A17, A18
		AG2c	A21–A35	A32
		AG2d	A36–A40	A39
		AG2e	A41–A65	A42, A44, A54, A59
		AG2f	A66, A67	
		AG2b	A68, A69	
P07	11	AG1d	A1–A3	A1
		AG1e	A4–A8	A4, A8
		AG1f	A9, A10	A9
		AG1g	A11, A12	A11
		AG1h	A13–A17	A17
		AG2c	A21–A35	A32
		AG2d	A36–A40	A39
		AG2h	A41–A54	A42, A44, A54
		AG2i	A56–A65	A59
		AG2f	A66, A67	
		AG2b	A68, A69	
P08	14	AG1d	A1–A3	A1
		AG1j	A4–A7	A4
		AG1f	A9, A10	A9
		AG1g	A11, A12	A11
		AG1h	A13–A17	A17
		AG2c	A21–A35	A32
		AG2j	A36, A37	
		AG2k	A39, A40	A39
		AG2m	A41–A43	A42
		AG2n	A44–A46	A44
		AG2o	A48–A54	A54
		AG2i	A56–A65	
		AG2f	A66, A67	
		AG2b	A68, A69	
P09	15	AG1d	A1–A3	A1
		AG1j	A4–A7	A4
		AG1f	A9, A10	A9
		AG1g	A11, A12	A11
		AG1j	A13–A16	A17
		AG2c	A21–A35	A32
		AG2j	A36, A37	
		AG2k	A39, A40	A39
		AG2m	A41–A43	A42
		AG2n	A44–A46	A44
		AG2q	A48–A51	
		AG2r	A52–A54	A54
		AG2i	A56–A65	A59
		AG2f	A66, A67	
		AG2b	A68, A69	
P10	14	AG1d	A1–A3	A1
		AG1j	A4–A7	A4
		AG1f	A9, A10	A9
		AG1j	A13–A16	
		AG2c	A21–A35	A32
		AG2j	A36, A37	
		AG2k	A39, A40	A39
		AG2m	A41–A43	A42
		AG2n	A44–A46	A44
		AG2q	A48–A51	
		AG2s	A52–A53	
		AG2i	A56–A65	A59
		AG2f	A66, A67	
		AG2b	A68, A69	

[a] Details of the sequences belonging to these classes are shown in Fig. 1A.
[b] ID, identification, see Materials and Methods.
[c] From reference 10.

TABLE 2. Class-specific residues that distinguish different known and suggested antigenic norovirus strains belonging to different evolutionary classes for partitions P01 to P07 of the phylogenetic tree constructed from a subset (S1 to S56) of sequences of Fig. 1[a]

Partitions (from Fig. 1)	Evolutionary class		Sequence clusters in class	Antigenic norovirus strains in class	Class-specific residues and insertions that occur in the different evolutionary classes corresponding to the first few partitions
	Total	ID			
P01	1		S1–S56	DSV, SOV, NV, MXV, HV, SMV, LSV	
P02, P03, P04	2	GI	S1–S10	DSV, SOV, NV	44A, 70I, 104V, 106N, 163E, 201V, 203A, 204G, **226Q, 329H, 375W, 377S, 397S, 414F, 434P, 460H, 463D, 471G, 485P∗, 486N∗, 500V, 514K, 519A, 527G∗**
		GII	S12–S56	MXV, HV, SMV, LSV	44P, 70V, 104A, 106G, 163P, 201T, 203S, 204C, **226S, 329G, 375P, 377G, 397G, 414P, 434D, 460R, 463N, 471F, 500Y, 514A, 519G**
P05	3	GIa	S1–S10	DSV, SOV, NV	GI + [83L, 181R, **227K, 306G, 492P**]
		GIIa	S12–S42	MXV, HV, SMV	GII + [83L, 181R, **227K, 306G, 492P**]
		GIIb	S43–S56	LSV	GII + [83A, 181K, **227R, 306W, 492D**]
P06	7	GIb	S1–S3	DSV	GIa + [46A, **229K, 232S, 236L, 241L**]
		GIc	S4–S6	SOV	GIb: [199S, **229R, 232T, 236I, 310M**]
		GId	S7–S9	NV	GIb: [84S, 197G∗, **229R, 232T, 241L,**]
		GIIc	S12–S23	MXV	GIIa + [46T, 195G∗, 199V, **229K, 232T**]
		GIId	S25–S37	HV, SMV	GIIa: [65N, 241L, **509Q**]
		GIIe	S39–S42	HV, SMV	GIIa: [46V, **250V, 509Q**]
		GIIf	S43–S56	LSV	GIIa: [46A, **250L, 275V, 306W, 310D**]
P07	8	GIb	S1–S3	DSV	P06+ 43A, 65Q, 82D, 125I, 133A, 142A, 149I, 156E, 160V, 203C, 223N, **239N, 240T, 246V, 253M, 259H, 288C, 301L, 347∗, 356L, 389P, 401E, 422M, 446V, 452T, 504V**
		GIe	S4–S5	DSV	P06+ 43L, 65Q, 82D, 125I, 133Q, 142A, 149I, 156D, 160V, 203C, 223T, **239K, 240Y, 246I, 253M, 259Q, 288C, 301L, 347P, 356L, 389F, 401E, 422M, 446I, 452I, 504A**
		GId	S7–S9	NV	P06+ 43V, 65Q, 82D, 125V, 133G, 142A, 149I, 156D, 160V, 203C, 223T, **239S, 240S, 246A, 253M, 259N, 288A, 301L, 347T, 356L, 389L, 401E, 422M, 446A, 452T, 504V**
		GIIc	S12–S23	MXV	P06+ 43A, 65G, 82N, 125F, 133P, 142I, 149I, 156E, 160L, 203R, 223T, **239S, 240E, 246F, 253L, 259E, 288C, 301L, 347T, 356L, 389Q, 401H, 422R, 446Y, 452A, 504E**
		GIIg	S25–S32	HV	P06+ 43A, 65N, 82N, 125F, 133P, 142I, 149I, 156E, 160L, 203R, 223T, **239G, 240E, 246F, 253L, 259E, 288C, 301V, 347T, 356L, 389Q, 401L, 422R, 446Y, 452S, 504D**
		GIIh	S33–S37	SMV	P06+ 43A, 65N, 82N, 125F, 133P, 142I, 149I, 156E, 160L, 203R, 223T, **239G, 240E, 246F, 253M, 259E, 288C, 301I, 347T, 356L, 389Q, 401L, 422R, 446Y, 452S, 504D**
		GIIi	S39–S41	SMV	P06+ 43A, 65A, 82D, 125F, 133P, 142L, 149V, 156E, 160L, 203R, 223S, **239S, 240E, 246F, 253M, 259E, 288C, 301L, 347T, 356I, 389Q, 401N, 422R, 446Y, 452A, 504E**
		GIIf	S43–S56	LV	P06+ 43A, 65G, 82S, 125F, 133P, 142V, 149I, 156E, 160I, 203R, 223T, **239E, 240E, 246F, 253L, 259S, 288C, 301L, 347T, 356L, 389Q, 401R, 422R, 446Y, 452A, 504D**

[a] Only seven partitions P01 to P07 are shown because these are sufficient to clearly demonstrate all the relevant ET features discussed in the text. The remaining partitions can be viewed as supplementary data (http:/ncmi.bcm.tmc.edu/~chin). Although class-specific residues are shown beginning at residue 1 in Fig. 2, only residues in locations 44 and higher are shown here because of the lack of knowledge of some of the N-terminal residues in the NV crystal structure. All known antigenic strains are resolved in partition P07. All class identification (ID) codes contain three characters: the first two characters are GI for all genogroup 1 classes, and they are GII for all genogroup 2 classes. These class IDs identify individual classes occurring at a given ETC position. All the sequences of these classes are shown in Table 1. The hinge region residues and the P domain residues are boldfaced in contrast to the S domain class-specific residues. Residue numbers correspond to that of NV. ∗, insertions; P0n+, including all class-specific residues of the previous class denoted by the number n.
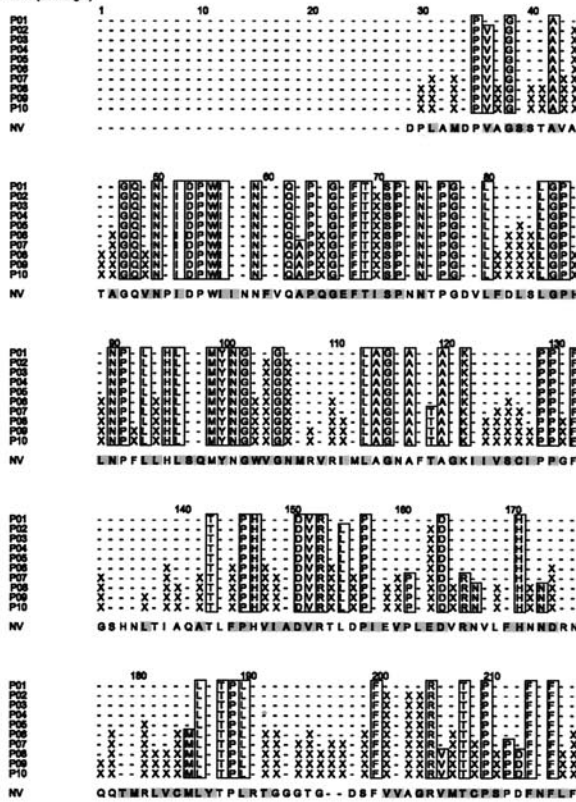
are known to be antigenically distinct from each other (10) and SOV (S6), DSV (S3), Lordsdale virus (LV; S45), and Mexico virus (MXV; S12) that have either been shown or suggested to be distinct antigenic strains (9). Partition P01 contains all these antigenic strains together (Table 2; Fig. 1). In partitions P02, P03, and P04, the class GI (genogroup I sequences S1 through S10), which includes the antigenic strains DSV, SOV, and NV, contains the following class-specific residues (NV numbering): 44A, 70L, 104V, 106N, 163E, 201V, 203A, and 204G in the S domain; residue 226Q in the hinge region; and residues 329H, 375W, 377S, 397S, 414F, 434P, 460H, 463D, 471G, 500V, 514K, and 519A in the P domain (Table 2). In contrast, the

corresponding residues in class GII (genogroup II sequences S12 through S56), which includes the antigenic strains MXV, HV, SMV, and LSV, are 44P, 70V, 104A, 106G, 163P, 201T, 203S, 204C, 226S, 329G, 375P, 377G, 397G, 414P, 434D, 460R, 463N, 471F, 500Y, 514A, and 519G (Table 2). Moreover, class-specific insertions at residues 485P, 486N, and 527G occur in class GI only (Table 2).
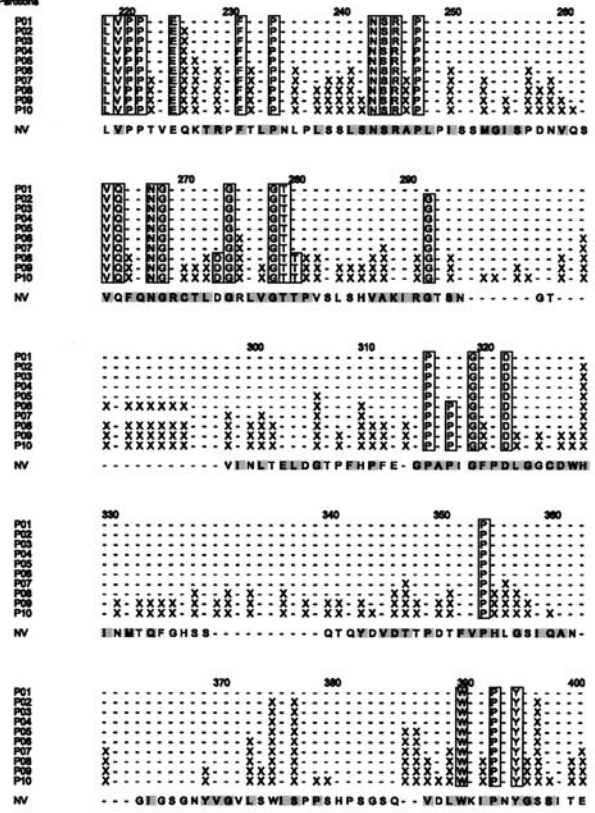
The GI class-specific residues of partitions P02 to P04, along with their structural neighbors as seen in the NV crystal structure, form 15 surfaces that may be important in distinguishing the different GI and GII strains. A majority (87%) of such surfaces contains at least one exposed residue, and six of the
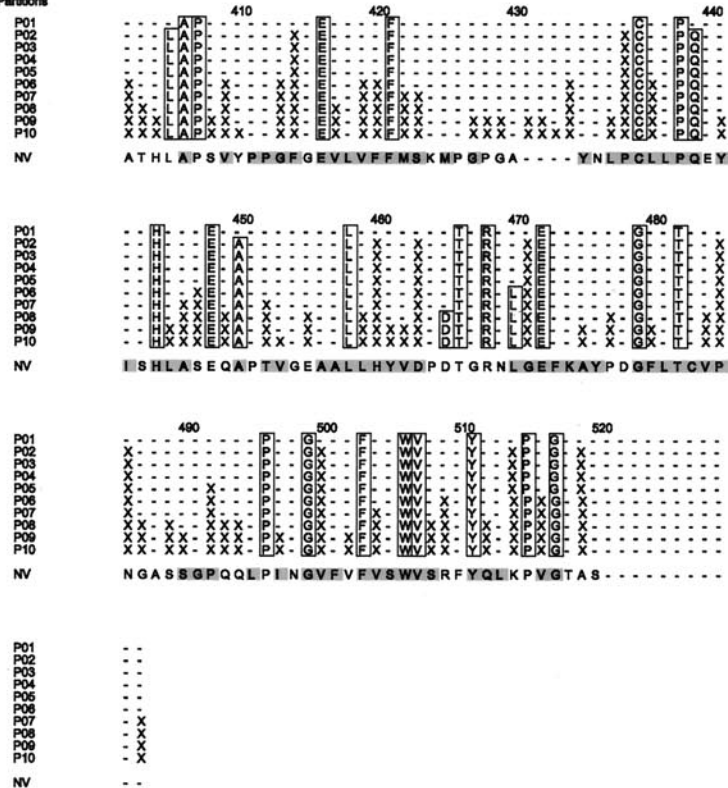
A

B

```
NV                  ------MMMASKDATSSVDGASGAGQLVPEVNASDPLAMDPVAGSSTAVA 44
Southampton         ------MMMASKDAPQSADGASGAGQLVPEVNTADPLPMEPVAGPTTAVA
DSV                 ------MMMASKDAPTNMDGTSGAGQLVPEANTAEPISMEPVAGAATAAA
Grimsby             ------MKMASNDANPSDGSTAN---LVPEVNNEV-MALEPVVGAAIAAP
Lordsdale           ------MKMASNDANPSDGSAAN---LVPEVNNEV-MALEPVVGAAIAAP
Japanese_oyster     LLPKVCMKMASNDATPSNDGAAG---LVPESNNEA-MALEPVVGASLAAP
                                                                    **

NV                  TAGQVNPIDPWIINNFVQAPQGEFTISPNNTPGDVLFDLSLGPHLNPFLL 94
Southampton         TAGQVNMIDPWIVNNFVQSPQGEFTISPNNTPGDILFDLQLGPHLNPFLS
DSV                 TAGQVNMIDPWIMSNYVQAPQGEFTISPNNTPGDILFDLQLGPHLNPFLS
Grimsby             VAGQQNVIDPWIRNNFVQAPGGEFTVSPRNAPGEILWSAPLGPDLNPYLS
Lordsdale           VAGQQNVIDPWIRNNFVQAPGGEFTVSPRNAPGEILWSAPLGPDLNPYLS
Japanese_oyster     VTGQTNIIDPWIRTNFVQAPNGEFTVSPRNSPGEILVNLELGPELNPYLA
                    **           *       *       *

NV                  HLSQMYNGWVGNMRVRIMLAGNAFTAGKIIVSCIPPGFGSHNLTIAQATL 144
Southampton         HLSQMYNGWVGNMRVRILLAGNAFSAGKIIVCCVPPGFTSSSLTIAQATL
DSV                 HLAQMYNGWVGNMRVKVLLAGNAFTAGKIIISCIPPGFAAQNISIAQATM
Grimsby             HLARMYNGYAGGFEVQVILAGNAFTAGKIIFAAVPPNFPAEGLSPSQVTM
Lordsdale           HLSRMYNGYAGGFEVQVILAGNAFTAGKVIFAAVPPNFPTEGLSPSQVTM
Japanese_oyster     HLAR----------------------------------------------
```
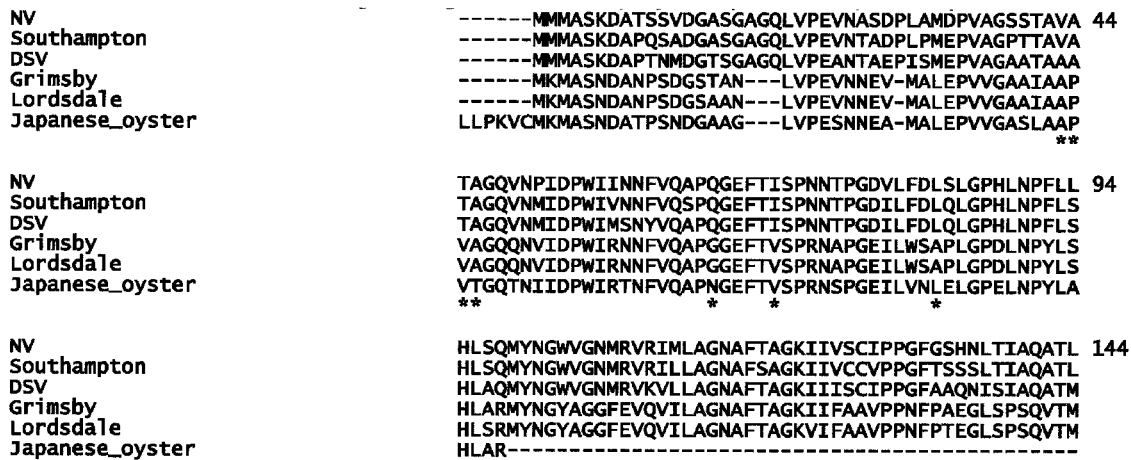
FIG. 2. (A) Locations of S and P domain trace residues for partitions P01 to P10. The trace residues belonging to a given partition occur in the horizontal row corresponding to the partition. Residues 1 to 218 and residues 219 to 401 are separated from each other by the partition numbers and are shown in the upper part of the figure, while residues 402 to 530 are shown in the lower part of the figure. The ACRs are shown as one-letter amino acid codes within boxes outlined in black, and the evolutionary class-specific residues are shown by the symbol X. The minus (−) signs denote nonconserved neutral residues. Residues are numbered according to the NV residues that are shown at the bottom of each group. (B) ClustalW alignment of partial norovirus sequence from Japanese oyster (NCBI accession no. BAC98461) with five other norovirus strains. Residues are numbered according to NV sequence. The asterisk denotes the class-specific locations that have been compared in the text.

surfaces are highly exposed with each surface containing more than three exposed residues (Table 3). Most of these exposed surfaces are located in the S and the P1 domains. The S domain exposed surfaces 1, 4, and 5 consist of residues interacting across the quasi-threefold and the icosahedral threefold and the fivefold axes. The remaining exposed surfaces (6, 11, and the common C-terminal end occurring in surfaces 13 to 15) consist of residues lying in the hinge region and the P1 domain (Table 3). Both of the P2 domain surfaces 7 and 8 are quite buried, although one of these surfaces (8) is relatively more exposed than the other. Interestingly, the P2 domain surface 7 is the only class-specific surface that includes the dimeric axis (Table 3).

The S and the P domain class-specific surfaces of the GII antigens are not known because of the lack of a GII crystal structure. However, it is reasonable to assume that the GII

TABLE 3. Surface-forming neighbors of class-specific residues that distinguish the GI noroviruses from the GII strains in partitions P02 to P04

Class-specific residues (NV numbering)	Surface	Neighbors forming the surfaces[a]
1044A[#]	1	1042A, 1043V*, 1045T*[#], 1046A
1070I	2	1069T, 1071S, 1202V[$%], 1074N, 1186L, 1080L, 1113L[$], 1152V, 1190L[%], 1201V[$]
1104V, 1106N	3	1036V[#], 1037A*[#], 1103W, 1105G, 1106N, 1107M, 1108R, 1162L, 1163E, 1164D, 1166R, 1169L*, 1170F, 1213D, 1214F, 1215N, 1217L
1163E	4	1033M*[#$], 1034D*, 1036V, 1106N, 1162L, 1164D, 1165V*[#], 1180M
1201V[$%], 1203A*, 1204G	5	1067E[$], 1068F[$], 1070I, 1071S, 1112M*[$], 1113L[$], 1114A*[$], 1200F[$%], 1202V[$], 1203A*[$], 1205R*[$%], 1512Q
1226Q	6	1225E*[#], 1227K*, 1229R, 1449Q*, 1228T
1329H[$], 1375W[$], 1377S	7	1326C, 1327D, 1328W, 1330I[$], 1342Q[$], 1343Y, 1344D, 1374S, 1376I, 1378P, 1379P, 1426P[$], 1427G[$], 1428P[$], 2338S, 2339S, 2340Q
1397S*	8	1353V, 1395Y, 1396G, 1398S, 1399I*
1414F	9	1074N, 1412P, 1413G, 1415G*, 1484V, 1512Q
1434P	10	1245R, 1265F, 1389L, 1390W, 1420F, 1421F, 1422M, 1432N*, 1433L, 1435C
1460H, 1463D, 1500V	11	1235N*, 1236L, 1459L, 1461Y, 1462V, 1464P*[$%], 1465D[$%], 1466T[$%], 1467G[$%], 1468R*[%], 1469N*, 1471G, 1472E, 1499G, 1501F, 1502V[$], 1504V
1471G	13	1461Y, 1470L, 1472E, 1516V, *1518T, 1519A*
1485P, 1486N	14	1417V, 1483C, 1484V, 1487G, 1488A*, 1490S, 1516V, 1517G, *1518T, 1519A*
1514K*[$%]	15	1484V, 1513L, 1515P*, 1516V

[a] These 4.0 Å sphere neighbors form class-specific surfaces (in boldface) in the NV crystal structure. The C terminus residues are italicized. To identify the different subunits, residues are denoted as xxxxN where xxxx is a four-digit number and N is the single-letter amino acid code of the residue. The first digit of xxxx refers to the subunit number (1, A; 2, B; 3, C) and the remaining three digits are the residue numbers; e.g., residue 1044 refers to 44 of A subunit; residue 2338 refers to residue 338 of B subunit. Superscript asterisk denotes exposed residues. The symbol # denotes residues interacting across the quasi-three fold axis, while the symbols §, $, and % denote residues interacting across the icosahedral twofold, fivefold, and threefold axes respectively.

class-specific residues that align with the GI surface-forming residues in a multiple sequence alignment interact to form surfaces in GII antigens as well. Characteristic differences among the class-specific capsid protein residues result in differences in the surfaces formed by these residues (Tables 2 and 3). Such differences uniquely distinguish all genogroup I noroviruses from those of genogroup II including the antigenic strains.

Similar comparisons of class-specific residues in the subsequent partitions P05 to P10 further distinguish the different norovirus antigenic strains belonging to a given genogroup and identify the class-specific trace residues that lead to such distinctions. The partition P05 contains the three classes GIa, GIIa, and GIIb (Table 2). The class GIa consists of the genogroup 1 sequences S1 to S10, while the remaining two classes GIIa and GIIb consist of the genogroup II sequences S12 to S42 and S43 to S56, respectively. The distribution of the antigenic strains in class GIa remains unchanged with respect to class GI of partition P02; the class GIIa retains the antigenic strains MXV, HV, and SMV from class GII of partition P02, while the antigenic strain LSV branches out independently into the genogroup II class GIIb (Table 2). The antigenic strains are further resolved in partition P06 where the strains DSV, SOV, and NV occur in genogroup I classes GIb, GIc, and GId, respectively. Among the genogroup II antigenic strains, MXV and LSV separate out into the distinct classes GIIc and GIIf, respectively, while HV and SMV occur in class GIId of this partition (Table 2). These two antigenic strains separate only in partition P07, where they occur in classes GIIg and GIIh, respectively (Table 2). Although all currently known antigenic strains are distinguished in partition P07, it is possible that additional antigenic strains occur in these partitions as well as in partitions P08 to P10. What about the class-specific residues that distinguish the strains belonging to a given genogroup? In partition P05, residues in locations 83, 181, 227, 306, and 492 uniquely distinguish the sequences in class GIIa of genogroup II from the sequences in class GIIb of the same genogroup. These residues are 83L, 181R, 227K, 306G, and 492P in class GIIa (and class GIa of genogroup I) but 83A, 181K, 227R, 306W, and 492D in class GIIb (Table 2). Similar variations in the class-specific residues distinguish more sequence clusters and the antigenic strains belonging to the classes GIb, GIc, GId, GIe, GIIc, GIId, GIIe, GIIf, GIIg, GIIh and GIIi of partitions P06 and P07 (Table 2). Variations in the class-specific residues for subsequent partitions P08 to P10 further distinguish between the sequences in these partitions (see supplemental material found at http:/ncmi.bcm.tmc.edu/~chin).

Detection of such subtle sequence variations by the ET method is a significant improvement over conventional cluster analysis methods in which sequence clusters are obtained by comparing the separation distance of all the sequences taken together. In conventional cluster analysis, distance histograms are plotted, using a distance filter in the form of an arbitrary cutoff value of the separation distance, and the highest peaks in these plots indicate the sequence clusters (2, 10). The fundamental drawback in this approach is that by comparing the separation distances of all the sequences together, no use is made of the sequence similarity information that is already embedded in the phylogenetic tree through its nodes and the connectivities among them. Consequently, the single-distance

filter cannot detect the tree-structured similarities at various nodes of the tree. Instead, the filter detects only the large peaks in the histogram that indicate only the gross similarity patterns between the sequences. Smaller peaks, corresponding to features of closely related sequence clusters, are often not visible in such histograms. In other words, cluster analysis that uses a single scalar value of the distance filter can discriminate only large differences between sequences such as that between distinct genogroups but cannot trace the subtle variations that exist between closely related norovirus antigenic strains. In contrast, because the ET method retains the connectivity information present within the tree while comparing the partition-based classes, it can detect very small differences between the class-specific residues of the antigenic sequences. Such fine variations in the class-specific residues in various partitions uniquely characterize the different norovirus antigenic and other strains, thereby explaining the diversity of existing and emerging norovirus strains.

Trace residues may uniquely characterize norovirus strains. Given a new norovirus strain whose capsid protein sequence may or may not be known completely, it is possible to identify the genogroup and the sequence cluster of the strain uniquely by systematically examining the class-specific trace residue locations. For example, if a new strain contains the class-specific residues of class GI, it would belong to genogroup I, while the strain would belong to genogroup II if it contained the class-specific residues of class GII (Table 2). Furthermore, if the strain belonged to genogroup II and it contained the class-specific residues 83L, 181R, 227K, 306G, and 492P, the strain would be similar to the genogroup II sequences S12 to S42 of class GIIa in partition P05 (Table 2). Otherwise, if the same genogroup II strain contained residues 83A, 181K, 227R, 306W, and 492D, it would be similar to the genogroup II sequences S43 to S56 of class GIIb in partition P05 (Table 2). Further categorization of the norovirus strain into a suitable cluster would be possible by comparing its residues at the class-specific locations corresponding to the subsequent P06 to P10 partitions of the phylogenetic tree. Such a procedure has been applied to the partial sequence of a norovirus isolate called Japanese oyster (29). On aligning this sequence with other norovirus sequences, locations 44 and 70 (NV numbering) are residues P and V, respectively (Fig. 2B). This indicates that the sequence belongs to class GII of partitions P02, P03, and P04 (Table 2). Furthermore, because position 83 is an L (Fig. 2B), this sequence belongs to class GIIa of partition P05 (Table 2). Further comparisons in partitions P06 and P07 show that locations 43, 46, 65, and 82 are A, T, G, and N, respectively, that categorize the Japanese oyster sequence in class GIIc along with the similar sequences S12 to S23 in (Table 2), which is consistent with published results (29). Similar comparisons can be continued for subsequent partitions to place the oyster sequence in smaller classes if required.

What about the murine (A70), bovine (A19 and A20), and the Alphatron-like (A68 and A69) sequences? All these sequences share some ACRs with GI and GII norovirus sequences when all the sequences are compared together in partition P01 (see supplemental material at http:/ncmi.bcm .tmc.edu/~chin). However, no class-specific residues characteristic of murine sequences can be defined because only one such sequence (A70) is known. More murine sequences are

needed to characterize them uniquely into appropriate ET classes on the basis of their class-specific residues. In contrast, the bovine sequences (Table 2, sequences A19 and A20) share nearly 50% ACRs and class-specific residues with other GI noroviruses in partitions P01 and P02 (see supplemental material at http:/ncmi.bcm.tmc.edu/~chin), after which the bovine sequences become independent, indicating thereby that these sequences may belong to a GI class instead of an independent genogroup as has been proposed (31). Similarly, the Alphatron-like sequences (Table 1, sequences A68 and A69) share class-specific residues with the GII sequences up to partition P03 (see supplemental material at http:/ncmi.bcm.tmc.edu/~chin), beyond which they branch off independently. This indicates that the Alphatron-like sequences are similar to GII sequences instead of being an independent genogroup (43). However, more such bovine and Alphatron-like sequences along with their structures are needed to evaluate whether such class-specific correlations with other GI and GII sequences are significant or random.

Clearly, the ET nodes and classes may formally define norovirus genogroups. By definition, any class must contain at least two sequences in order to define their class-specific residues. Let a node be designated "parent" if it has further "child" nodes diverging from it. Thus, 1 and 2 are child nodes of the parent (root) node 0, while 5 and 6 are child nodes of parent node 3 (Fig. 1). Distinct genogroups may be defined as those classes that originate in nodes having only the root as the parent node. Thus, ignoring the single murine sequence at node 2, as this sequence does not form a class, genogroups I and II are clearly the classes that belong to nodes 3 and 4 because these are the only nodes that have the root as their parent node (Fig. 1).

Such ET-based definitions and assignments of genogroups and classes have advantages. The main advantage of such assignments of genogroup and class is that knowledge of trace residues alone is sufficient to uniquely characterize complete or partial norovirus sequences. In addition, unlike conventional methods, such assignments may easily be automated to make the method cost-effective. As the trace residue locations are distributed throughout the genome, partial sequences from a large number of genomic regions may be used to carry out such analysis. However, the number of trace residues that should be known depends on the desired resolution with which the class of the isolate needs to be determined in the phylogenetic tree. The fewer the number of trace locations sequenced, the broader will be the classification due to the fewer number of class-specific differences considered (Table 2). In order to understand whether such differences have functional implications, the possible roles of the different types of norovirus class-specific trace residues need to be examined.

S domain: significance of the ACRs and class-specific residues in norovirus assembly. The rNV structure shows that a majority (65%) of the ACRs are hydrophobic, buried, and located in the S domain. Most of these residues (58%) interact with neighboring subunits across the icosahedral interfaces (Fig. 3A and B). Therefore, hydrophobic interactions between the S-domain ACRs are important in maintaining the icosahedral structure in noroviruses. The number of interactions (NI) between the ACRs across the icosahedral interfaces of rNV show a descending order $(NI)_{A-B5} > (NI)_{B5-C} >$

$(NI)_{A-A5} > (NI)_{C-B2} > (NI)_{ABC,}$ where the subscripts indicate the corresponding interfaces shown in Fig. 3A. The dimeric $(A-B_5)$ interface of the rNV structure contains the largest numbers of ACRs that participate in intersubunit interactions, while the quasi-trimeric (ABC) interface contains the minimum number of such interacting ACRs. The pentameric $(A-A_5)$ interface contains fewer interactions than the dimeric $(A-B_5)$ and the hexameric (B_5-C) interfaces but more than the number of interactions at the hexameric $(C-B_2)$ and the quasi-trimeric ABC interfaces. Such variations in NI across the different icosahedral interfaces match the variations in the calculated values of the buried surface areas (BA) of the rNV interfaces $(BA)_{A-B5} > (BA)_{B5-C} > (BA)_{A-A5} > (BA)_{C-B2} > (BA)_{ABC}$, which is in agreement with similar calculations shown on the VIPER website (36).

The interfaces and their BAs are important determinants of the energetics of the assembly of icosahedral viruses (6, 14, 19, 23, 37, 38). In modeling the assembly pathway of icosahedral capsids based on the energetics of the assembly process of the different subunits, the assumption is usually made that the association energies between the subunits during assembly are directly correlated with their BAs computed from X-ray coordinates of the assembled capsid (37), provided these BAs are not significantly altered due to interactions of the virus subassemblies with their environment. Following this assumption, the relative values of the NIs between the ACRs at these rNV interfaces (or the BAs of the corresponding interfaces) suggest a model for the assembly and disassembly of norovirus capsids. The inequality of these NIs indicates that the T=3 norovirus capsid should follow different pathways during its assembly and disassembly. The most stable capsid interface is likely to assemble first, followed by the assembly of the less stable interfaces. Because the stability of the capsid interfaces is related to their BAs (19, 37), it follows that the capsid interfaces with the largest BA should assemble the earliest, while those with progressively lower BA values should assemble later. Therefore, during assembly, monomers should associate first to form the AB_5 and CC_2 dimers (Fig. 3A). The dimers would associate among themselves to form $A-B_5$–$C-C_2$ (dimer of dimers) intermediates that form the pentamer around the fivefold axis. Association of the pentamers creates the quasi-threefold environment in a natural way. Disassembly of the virus is likely to follow the reverse pathway if no significant changes in the BAs of the subassemblies occur due to virus-environment interactions. Thus, the S domain ACRs, by virtue of their locations, appear to be important in the assembly and disassembly processes in noroviruses.

What about the S domain class-specific residues? A majority of such residues is also found to cluster near the NV icosahedral interfaces at the early P02 partition in the evolutionary tree (Table 4). The class-specific residues 70I, 201V, 203A, and 204G cluster together to form surface patch 1, while residues 104V, 106N, and 163E cluster to form the second surface patch 2 in genogroup I viruses (Table 4). The corresponding genogroup II residues are 70V, 201T, 203S, and 204C for surface 1 and 104A, 106G, and 163P for surface 2. Because of the absence of a GII crystal structure, it is not possible to precisely determine whether these residues form surfaces in these noroviruses. However, covariance matrices of the pairwise volume correlations of the substitutions at these residue locations
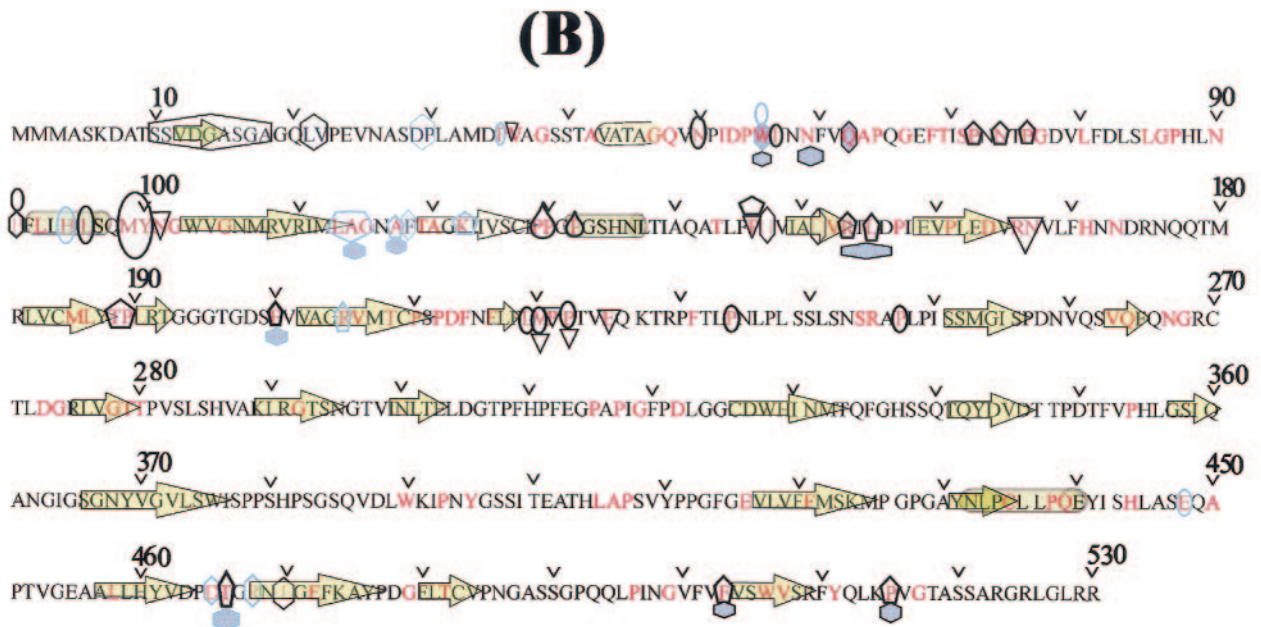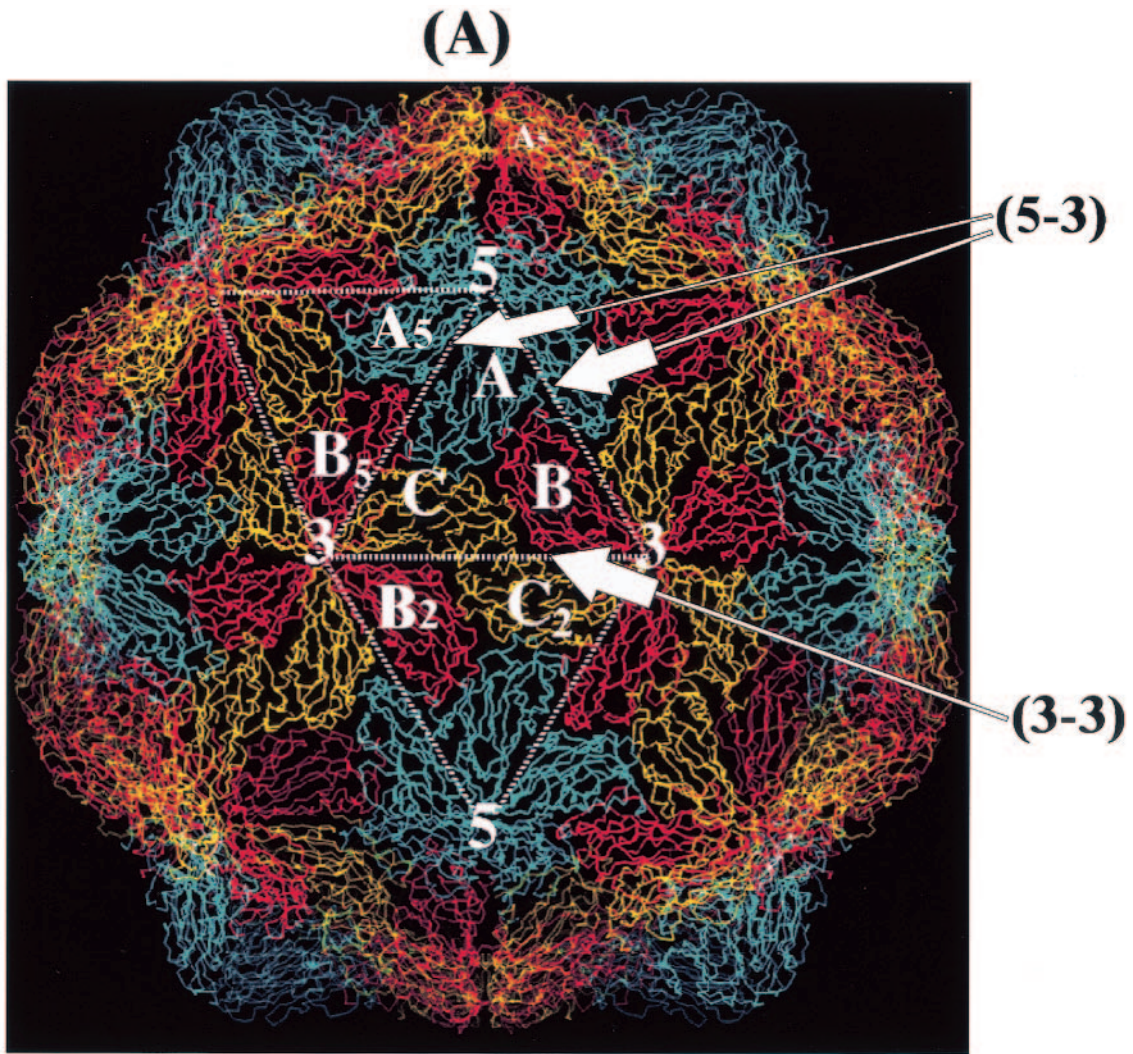
(A)

(B)

(5-3)

(3-3)

A₅ A₅ A

B₅ C B₃

3 B₂ C₂

5

MMMASKDATSSVDGASGAGQLVPEVNASDPLAMDIVAGSSTAVATAGQVQPIDPWQNNFVQAPQGEFTISENYIEGDVLFDLSLGPHLN

QFLLIISQMYNGWVGNMRVRIMLAGNAFLAGKIVSCHQQGSHNLTIAQATLFVVMALVQDPIEVPLEDVRNVLFHNNDRNQQTM

RLVCMLVFLRTGGGTGDSLVVAGRVMVCPSPDFNELHQVQDVFQKTRPFTLQNLPLSSLSNSRAQLPISSMGISPDNVQSVQPQNGRC

TLDGRLVGIPVSLSHVAKLRGTSNGTVINLTELDGTPFHPFEGPAPIGFPDLGGCDWHINMFQFGHSSQTQYDVDTTPDTFVPHLGSIQ

ANGIGSGNYVGVLSWISPPSHPSGSQVDLWKIPNYGSSITEATHLAPSVYPPGFGEVLVFEMSKMPGPGAYNLPLLPQEYISHLASEQA

PTVGEAALLHYVDPIQLNIGEFKAYPDGELTCVPNGASSGPQQLPINGVFWIVSWVSRFYQLKPVGTASSARGRLGLRR

TABLE 4. Class-specific residues present in the S and P domain surface patches for the P02 to P06 partitions[a]

Partition and domain	Surface patch	Class-specific residues	Nearest ACR(s)
P02	1	70I, 201V*, 203A, 204G	200F, 205R
S domain	2	104V, 106N, 163E	105G
	3	44A**	
P domain	**CS-1**	**226Q***	**225E**
	CS-2	**414F*, 463D, 485P, 486N, 514K**, 519A**	**515P**
	CS-3	**434P, 460H*, 500V**	**465D, 466T**
	CS-4	**329H, 375W*, 377S**	
P05	1	70I, 201V*, 203A, 204G	200F, 205R
S domain	2	104V, 106N, 163E, 174D, 178Q*	105G
	3	44A*	
	4	83L	
P domain	**CS-1**	**226Q*, 229R*, 227K***	**225E**
	CS-2	**414F*, 463D, 485P, 486N, 514K**, 519A, 520S****	**515P**
	CS-3	**434P, 460H*, 500V, 492P**	**465D, 466T**
	CS-4	**329H, 375W*, 377S**	
P06	1	70I, 192T*, 196T**, 199S*, 201V*, 203A, 204G	200F, 205R
S domain	2	103W, 104V, 106N, 163E, 174D, 178Q**, 183V, 94L	105G
	3	44A**, 46A, 218F	42A, 105G
	4	83L, 65Q, 82D*, 84S, 89L, 110R, 126S, 145F, 148V, 181R	129P
P domain	**CS-1**	**226Q*, 229R,* 227K***	**225E**
	CS-2	**414F*, 463D, 485P, 486N, 519A, 514K**, 520S***	**515P**
	CS-3	**434P, 460H*, 500V, 236L, 241L, 257P, 409I, 419V, 420F**	**465D, 466T**
	CS-4	**329H, 375W*, 377S, 373L***	
	CS-5	**306G, 310H**	

[a] These surface patches are subsets of all the surfaces shown in Table 2. The ACRs that occur nearest these surface patches are also shown. The P domain surface patches are shown in boldface. Accessible surfaces mostly consist of moderately and highly exposed residues that are shown with one and two asterisks, respectively.

show relatively high values for the diagonal terms in comparison with the off-diagonal terms at a 95% statistical confidence limit (data not shown). Assuming that correlated residue substitutions often imply functional interactions between such substituted residue locations (1, 3), this indicates a significant probability of interactions among such GII S domain class-specific residues across the icosahedral interfaces. The class-specific residue 44A, in the N-terminal helix of the genogroup I rNV structure near the quasi-sixfold (CB_2 and CB_5) interfaces (Fig. 3A and B), is a 44P in the genogroup II viruses. Such differences in the class-specific residues and their surfaces (Tables 4 and 5) in various partitions may be important in distinguishing subtle aspects of the assembly and disassembly processes in noroviruses of the two genogroups and in the clusters that are generated in the respective partitions (Table 4; Fig. 4A).

Hinge region: residues implicated in proteolytic cleavage form a class-specific surface patch. The class-specific residues of the hinge region connecting the S domain and the P1 subdomain indicate different functional specificities compared to the S domain. The exposed class-specific residues 226Q, 227K,

and 229R in the hinge region cluster to form a surface patch CS-1 around the exposed ACR 225E (Table 4 and Fig. 4B and C). Because this region undergoes proteolytic cleavage at high pH (12) and this patch starts forming in the P02 partition where the two genogroups diverge (Tables 2 and 4), residues in this patch may be critical for such cleavages in both norovirus genogroups. Although the exact significance of this cleavage site in the norovirus life cycle is not known, the evolutionary importance of the hinge region shown by the ET analysis indicates biological significance for this region.

P1 subdomain: observed antigenic regions contain class-specific surfaces. The class-specific residues in the P1 subdomain form two patches in surfaces 9 through 15 that contain the highly exposed C-terminal residues (Tables 3 and 4). These two patches CS-2 and CS-3 that begin developing in the P02 partition (Table 4 and Fig. 4B and C) may be antigenically significant in both norovirus genogroups. The patch CS-2, centered about the C-terminal ACR 515P, develops into a large exposed patch in partition P06. This patch, consisting of the exposed residues 414F, 514K, and 520S along with the buried residues 463D, 485P, 486N, and 519A (Table 4 and Fig. 4B and

FIG. 3. (A) Interactions of the S domains of the NV subunits across the different icosahedral symmetry axes. The locations of these axes are shown by the corresponding numbers. Subunits A, B, and C are related by a quasi-threefold (shown by a triangle) and C and C_2 are related by an icosahedral twofold axis. Subunits A and B_5 are related by quasi-twofold axes, and C is related to B_5 and B_2 by quasi-sixfold symmetry. Subunits A and A_5 are related by a strict fivefold axis. The broken lines joining the icosahedral symmetry axes are the icosahedral interfaces indicated by arrows. The line (5-3) joining the fivefold and the threefold axes includes the A-B_5 (dimeric), A-A_5 (pentameric), and the C-B_5 (hexameric) interface regions while the line (3-3) joining the two threefold axes, includes the C-B_2 (hexameric) interface region. Similarly, the quasi-trimeric A-B, A-C, and the B-C interface regions lie on the boundaries between the A, B, and the C subunits. (B) NV conserved residues with reference to the secondary structure and icosahedral interfaces. Filled green arrow, β-strand; filled green cylinder, helix. The ACRs are shown in red along with their locations with respect to the icosahedral symmetry axes that are shown, using different symbols: oval, twofold axis; inverted triangle, threefold axis; hexagon, quasi-sixfold axis; pentagon, fivefold axis; and teardrop, quasi-threefold axis. A black border around the symmetry axes symbols indicates a buried residue, while no border around the symbol indicates an exposed residue. Multiple symbols on ACRs indicate the presence of multiple symmetry axes about which these residue(s) interact with other subunits.

TABLE 5. Class-specific residues of the CS-4 surfaces[a]

Residue no.	Residue in GI strain (ID code)			Residue in GII strain (ID code)				
	NV* (A1)	Aichi (A2)	KY-89 (A3)	VA387* (A27)	Grimsby* (A26)	MXV* (A60)	MOH* (A45)	VA207* (A37)
326	C			K	K	K	E	L
327	D			I	I	V	V	L
328	W			Q	Q	F	F	Y
330	I			M	M	V	V	V
338	S			T	T	T	N	Del
339	S			R	R	R	R	Del
340	Q			A	A	A	A	Del
342	Q			K	K	E	D	Del
343	Y			A	A	A	A	Del
344	D			T	T	K	V	Del
374	S			T	T	T	T	T
376	I	V	V	V	V	V	I	I
378	P			V	V	I	L	I
379	P			I	I	N	G	Y
426	P							
427	G			Del	Del	Del	Del	Del
428	P			Del	Del	Del	Del	Del

[a] Class-specific residues of the putative receptor-binding surface (patch CS-4 in Table 4 corresponding to surface 7 in Table 3) of NV along with the corresponding ET-aligned residues of GI and GII norovirus strains for partition P06. Strain names and their identification (ID) codes are from Fig. 1. All residues are denoted by their one-letter amino acid codes and they are NV numbered. Blank entries indicate residues that are identical with NV. An asterisk indicates strains for which receptor-binding data are available (15). Del, deletion.

C), may be critical in defining the C-terminal epitopes reported in binding studies of noroviruses with monoclonal antibodies (13). The S domain patch 1 containing the exposed residues T192, T196, and S199 (Table 4) may be considered together with the CS-2 patch as part of a larger S and P domain surface patch that may be part of the epitopes (Fig. 4B). The moderately exposed CS-3 patch (Table 4), centered about the ACRs 465D and 466T that interact with the S domain across the fivefold axis, may define the additional nonoverlapping epitopes observed in NV (11).

Residue differences in the class-specific patches CS-2 and CS-3 between the two genogroups may explain the observed genogroup-specific reactivity differences of certain monoclonal antibodies (11). Residues 414F, 463D, 514K, and 519A that constitute patch CS-2 in all genogroup I noroviruses as shown in class GI of partition P02 (Tables 2 and 4), are 414P, 463N, 514A, and 519G in all genogroup II noroviruses as shown in class GII for the same partition (Table 2). In partitions P05 and P06, insertions at residues 485P and 486N of patch CS-2 in all genogroup II norovirus strains (Table 2) introduce additional genogroup-specific differences in this patch. Similarly, residues 434P, 460H, and 500V that comprise patch CS-3 in all genogroup I noroviruses (Table 2, class GI of partition P02) are 434D, 460R, and 500Y in all genogroup II noroviruses, as shown in class GII for the same partition (Tables 2). Such differences indicate that the patch CS-2 may be important in defining genogroup-specific epitopes that distinguish between the genogroup I antigenicity from that of genogroup II (11, 22).

Partitions P05 and P06 indicate cluster-specific modifications to the CS-2 and CS-3 patches. The CS-2 patch shows the addition of only one class-specific residue at location 520 in the P05 partition relative to the P02 partition (Table 4). In contrast, the patch CS-3 contains many differences in the P05 and

the P06 partitions relative to the P02 partition. These differences are in class-specific residues corresponding to the locations 492, 236, 241, 257, 409, 419, and 420 (Tables 2 and 4). Because of a larger number of cluster-specific variations in the CS-3 patch in comparison with the CS-2 patch, the CS-3 patch may be relatively more important in defining strain-specific epitopes that can distinguish the antigenicity of the different clusters in the two genogroups (22). Thus, the possible epitopes on the CS-2 and the CS-3 surface patches in the P1 subdomain may be part of genogroup-specific and strain-specific epitopes, respectively, in noroviruses. These surfaces are relatively more conserved in GI compared to the GII strains. Such conservation in these exposed NV surfaces may contribute to the observed homologous seroresponses of GI-infected patients to the recombinant NV antigen in contrast to that of GII infections (30). However, the hinge region patch CS-1 is the only class-specific patch that consists of residues that are nearly conserved between the two genogroups (Table 2). This patch may be involved in binding cross-reactive monoclonal antibodies that react with both the norovirus genogroups (22). The relatively sharp distinction in the seroresponses of GII-infected patients to the GII antigens HV and Toronto strains (30) may be understood from partition P07 in which HV (A54) lies in class AGIIh that is distinct from class AGIIi containing the Toronto (A59) antigen (Table 1). The class-specific residues 65N, 241L and 509Q in the GIIc and GIId classes of partition P06 mainly distinguish between these two strains (Table 2).

P2 subdomain: class-specific residues indicate a putative carbohydrate-binding site. Two class-specific surface patches are formed in the P2 domain. The patch CS-4 consists of the class-specific residues at locations (NV numbering) 329, 373, 375, and 377 (Table 4 and Fig. 4B). Residues corresponding to these locations in genogroup I noroviruses are 329H, 373L, 375W, and 377S. These residues are located near the only P2 domain cavity at the dimeric interface of rNV as seen in its crystal structure. Additional residues 267N, 322D, 327D, 374S, 331N, 333T, 334Q and 341T also lie in the vicinity of this cavity. Because all of these residue types are known to bind carbohydrate and sugar molecules and because blood group H antigens have been suggested to be the receptors for noroviruses from binding studies (16), it is possible that the CS-4 patch of this groove is involved in carbohydrate receptor binding of noroviruses. Recent studies have implicated a sequence motif that lies near one end of this cavity, in the binding of carbohydrate to noroviruses (41).

The two norovirus genogroups show significant differences in the class-specific residues of the CS-4 patch. Residues 329H, 375W, and 377S in genogroup I noroviruses of class GI are 329G, 375P, and 377G in the genogroup II noroviruses of class GII in the P02 partition (Table 2). Such residue variations in the CS-4 patch may explain the differences in carbohydrate binding to noroviruses of the two genogroups (16). Additional class-specific differences in this patch for the subsequent partitions may explain the host specificity in the receptor binding of the NV (GI) strain and the five GII (Grimsby, VA387, VA207, MOH, and MXV) strains (15). The earliest partition that distinguishes among these GII strains is P06 in which the Grimsby (A26) and VA387 (A27) strains belong to class AGIIc and VA207 (A37) belongs to class AGIId, while MOH (A45)
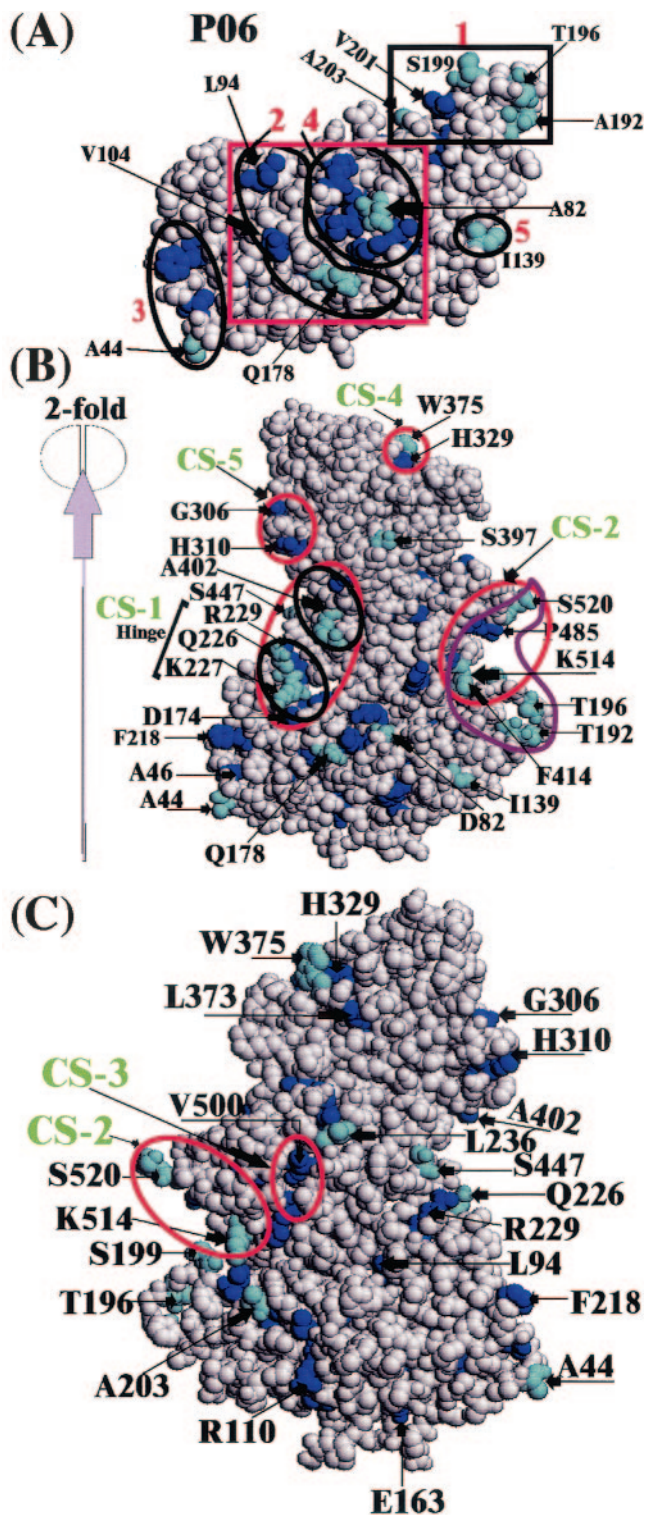
and MXV (A60) belong to class AGIIe (Table 1). The class-specific surface residues of the proposed receptor-binding CS-4 surface patch are identical in VA387 and Grimsby, whereas they are only similar but not identical in MXV and MOH (Table 5). This may explain why the VA387 and the Grimsby strains share very similar receptor binding patterns that are different from the patterns of the MXV and MOH strains (15). In addition, a greater similarity in the class-specific residues of the CS-4 surfaces (Table 5) may explain the relatively similar receptor binding patterns observed for the VA387 and the Grimsby strains in comparison with that of the MXV and MOH strains. Because these residues are highly conserved among the A21 to A35 strains of class AGIIc compared to that of the A41 to A65 strains of class AGIIe in partition P06 (Table 1), the strains in class AGIIc may share a greater similarity in receptor binding characteristics in comparison to the strains of class AGIIe. The VA207 strain of class AGIId in partition P06 (Table 1) shows markedly different residues in the CS-4 surface in comparison with other noroviruses (Table 5). These differences may explain the distinct receptor binding characteristics observed for the VA207 strain (15). Interestingly, the GII class AGIId shows (Table 1) very low conservation of the CS-4 surface, indicating highly nonhomologous receptor binding patterns that should be expected for the A36 to A40 strains of this class (Table 1 and Fig. 1). In contrast, NV, Aichi, and the KY-89 GI strains belonging to the AG1d class in partition P07 (Table 1) show high sequence identity among themselves for the CS-4 surface (Table 5), indicating a very similar receptor binding pattern expected for these strains.

A second class-specific surface patch CS-5 is also seen in the P2 subdomain. This patch, consisting of the buried residues 306G and 310H (Table 4; Fig. 4B and C), may define part of the same carbohydrate-binding site CS-4, or it may be involved in additional antigenic binding in noroviruses. The exact functional significance of this patch is not clear, similar to another surface patch that consists of the exposed residue 402A and the buried residue 447S near the CS-1 patch of the hinge region (Fig. 4B).

Conclusion. The ET analysis, hitherto carried out only in proteins, of the norovirus sequences has allowed identification of class-specific trace residues of the capsid protein that may be functional signatures defining the strain diversity in noroviruses. In contrast to the existing classification schemes based on conventional phylogenetic analysis, the ET approach has the advantage that the classification of the viruses in a given genus or family can be understood on the basis of putative functional sites deduced from evolution-related subtle variations in their sequences and independent biochemical evi-

FIG. 4. Space-filled representation of the proposed functionally important class-specific residues for the P06 partitions of the phylogenetic tree. The buried atoms of the class-specific residues are shown in blue and the exposed atoms of these residues are cyan. (A) The S domain surface patches 1 to 5 are outlined in black. The red box indicates that the S domain surface patches 2 and 4 may be considered as one single such patch. (B) The different class-specific surface patches CS-2, CS-4, and CS-5 of Table 4 are outlined in red. The surface CS-1 is outlined in black. Another possible class-specific surface consisting of the residues A402 and S447 lies near the CS-1 surface. The red outline enclosing the two class-specific patches outlined in black indicates that these patches may be considered as a single patch. The purple outline enclosing the S domain patch number 1 and the P domain CS-1 patch indicates that these patches may be considered as one single patch or a pair of possible neutralizing and nonneutralizing epitope sites. The direction of the icosahedral twofold axis relative to the capsid subunit is also shown. (C) A 180° rotated view of the figure in panel B. This figure indicates the surface patch CS-3 that is not visible in panel B. The surface patch CS-2 is also shown to indicate the relative location of the CS-3 patch.

dence. The generality of the approach allows similar analyses to be performed with nucleotide sequences. Our preliminary analysis of noroviruses ORF2 nucleotide sequences has identified class-specific nucleotides that vary between the strains of genogroups 1 and 2 (data not shown). Such nucleotides may possibly be used for designing primers for reverse transcription PCRs to identify genogroup-specific norovirus strains.

REFERENCES

1. **Altschuh, D., A. M. Lesk, A. C. Bloomer, and A. Klug.** 1987. Correlation of co-ordinated amino acid substitutions with function in viruses related to tobacco mosaic virus. J. Mol. Biol. **193:**693–707.
2. **Ando, T., J. S. Noel, and R. L. Fankhauser.** 2000. Genetic classification of Norwalk-like viruses. J. Infect. Dis. **181**(Suppl. 2):S336–S348.
3. **Atchley, W. R., K. R. Wollenberg, W. M. Fitch, W. Terhalle, and A. W. Dress.** 2000. Correlations among amino acid sites in bHLH protein domains: an information theoretic analysis. Mol. Biol. Evol. **17:**164–178.
4. **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.
5. **Blakeney, S. J., A. Cahill, and P. A. Reilly.** 2003. Processing of Norwalk virus nonstructural proteins by a 3C-like cysteine proteinase. Virology **308:**216–224.
6. **Caspar, D. L. D., and A. Klug.** 1962. Physical principles in the construction of regular viruses. Cold Spring Harbor Symp. Quant. Biol. **27:**1–24.
7. **Chen, R., J. D. Neill, J. S. Noel, A. M. Hutson, R. I. Glass, M. K. Estes, and B. V. Prasad.** 2004. Inter- and intragenus structural variations in caliciviruses and their functional implications. J. Virol. **78:**6469–6479.
8. **Glass, R. I., J. S. Noel, T. Ando, R. L. Fankhauser, G. Belliot, A. Mounts, U. D. Parashar, J. S. Bresse, and S. Monroe.** 2000. The epidemiology of enteric caliciviruses from humans: a reassessment using new diagnostics. J. Infect. Dis. **181:**S254-261.
9. **Green, J., J. Vinje, C. I. Gallimore, M. Koopmans, A. Hale, and D. W. Brown.** 2000. Capsid protein diversity among Norwalk-like viruses. Virus Genes **20:**227–236.
10. **Green, K. Y., R. M. Chanock, and A. Z. Kapiakan.** 2001. Human caliciviruses, p. 841–874. *In* Fields virology, 4th ed., Lippincott, Williams & Wilkins, Baltimore, Md.
11. **Hale, A. D., T. N. Tanaka, N. Kitamoto, M. Ciarlet, X. Jiang, N. Takeda, D. W. Brown, and M. K. Estes.** 2000. Identification of an epitope common to genogroup 1 "Norwalk-like viruses." J. Clin. Microbiol. **38:**1656–1660.
12. **Hardy, M., L. White, J. Ball, and M. Estes.** 1995. Specific proteolytic cleavage of recombinant Norwalk virus capsid protein. J. Virol. **69:**1693–1698.
13. **Hardy, M. E., T. N. Tanaka, N. Kitamoto, L. J. White, J. M. Ball, X. Jiang, and M. K. Estes.** 1996. Antigenic mapping of the recombinant Norwalk virus capsid protein using monoclonal antibodies. Virology **217:**252–261.
14. **Harrison, S. C., A. J. Olson, C. E. Schutt, and F. K. Winkler.** 1978. Tomato bushy stunt virus at 2.9Å resolution. Nature **276:**368–373.
15. **Huang, P., T. Farkas, S. Marionneau, W. Zhong, N. Ruvoën-Clouet, A. L. Morrow, M. Altaye, L. K. Pickering, D. S. Newburg, J. LePendu, 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.
16. **Hutson, A. M., R. L. Atmar, D. M. Marcus, and M. K. Estes.** 2003. Norwalk virus-like particle hemagglutination by binding to H histo-blood group antigens. J. Virol. **77:**405–415.
17. **Innis, C. A., J. Shi, and T. L. Blundell.** 2000. Evolutionary trace analysis of TGF-beta and related growth factors: implications for site-directed mutagenesis. Protein Eng. **13:**839–847.
18. **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.
19. **Johnson, J. E., and J. A. Speir.** 1997. Quasi-equivalent viruses: a paradigm for protein assemblies. J. Mol. Biol. **269:**665–675.
20. **Kapikian, A. Z., R. G. Wyatt, R. Dolin, T. S. Thornhill, A. R. Kalica, and R. M. Chanock.** 1972. Visualization by immune electron microscopy of a 27-nm particle associated with acute infectious nonbacterial gastroenteritis. J. Virol. **10:**1075–1081.
21. **Karst, S. M., C. E. Wobus, M. Lay, J. Davidson, and H. W. Virgin IV.** 2003.

22. **Kitamoto, N., T. Tanaka, K. Natori, N. Takeda, S. Nakata, X. Jiang, and M. K. Estes.** 2002. Cross-reactivity among several recombinant calicivirus virus-like particles (VLPs) with monoclonal antibodies obtained from mice immunized orally with one type of VLP. J. Clin. Microbiol. **40:**2459–2465.
23. **Lee, W.-M., and W. Wang.** 2003. Human rhinovirus type 16: mutant V1210A requires capsid-binding drug for assembly of pentamers to form virions during morphogenesis. J. Virol. **77:**6235–6244.
24. **Lewis, D.** 1991. Norwalk agent and other small round structured viruses in the United Kingdom. J. Infect. **23:**220–222.
25. **Lichtarge, O., H. R. Bourne, and F. E. Cohen.** 1996. An evolutionary trace method defines binding surfaces common to protein families. J. Mol. Biol. **257:**342–358.
26. **Lichtarge, O., and M. E. Sowa.** 2002. Evolutionary predictions of binding surfaces and interactions. Curr. Opin. Struct. Biol. **12:**21–27.
27. **Madabushi, S., H. Yao, M. Marsh, D. M. Kristensen, A. Philippi, M. E. Sowa, and O. Lichtarge.** 2002. Structural clusters of evolutionary trace residues are statistically significant and common in proteins. J. Mol. Biol. **316:**139–154.
28. **Mihalek, I., I. Res, H. Yao, and O. Lichtarge.** 2003. Combining inference from evolution and geometric probability in protein structure evaluation. J. Mol. Biol. **331:**263–279.
29. **Nishida, T., H. Kimura, M. Saitoh, M. Shinohara, M. Kato, S. Fukuda, T. Munemura, T. Mikami, A. Kawamoto, Y. Kato, K. Nishi, K. Kozawa, and O. Nishio.** 2003. Detection, quantitation, and phylogenetic analysis of noroviruses in Japanese oysters. Appl. Environ. Microbiol. **69:**5782–5786.
30. **Noel, J. S., T. Ando, J. P. Leite, K. Y. Green, K. E. Dingle, M. K. Estes, Y. Seto, S. S. Monroe, and R. I. Glass.** 1997. Correlation of patient immune responses with genetically characterized small round-structured viruses involved in outbreaks of nonbacterial acute gastroenteritis in the United States, 1990 to 1995. J. Med. Virol. **53:**372–383.
31. **Oliver, S. L., A. M. Dastjerdi, S. Wong, L. El-Attar, C. Gallimore, D. W. Brown, J. Green, and J. C. Bridger.** 2003. Molecular characterization of bovine enteric caliciviruses: a distinct third genogroup of noroviruses (Norwalk-like viruses) unlikely to be of risk to humans. J. Virol. **77:**2789–2798.
32. **Pattabiraman, N., and K. Ward.** 1995. Occluded molecular surface: analysis of protein packing. J. Mol. Recogn. **8:**334–344.
33. **Pfister, T., and E. Wimmer.** 2001. Polypeptide p41 of a Norwalk-like virus is a nucleic acid-independent nucleoside triphosphatase. J. Virol. **75:**1611–1619.
34. **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.
35. **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.
36. **Reddy, V., P. Natarajan, B. Okerberg, K. Li, K. Damodaran, R. Morton, C. Brooks III, and J. E. Johnson.** 2001. Virus particle explorer (VIPER), a website for virus capsid structures and their computational analyses. J. Virol. **75:**11943–11947.
37. **Reddy, V. S., H. A. Giesing, R. T. Morton, A. Kumar, C. B. Post, C. L. Brooks III, and J. E. Johnson.** 1998. Energetics of quasiequivalence: computational analysis of protein-protein interactions in icosahedral viruses. Biophys. J. **74:**546–558.
38. **Rossmann, M. G., C. Abad-Zapatero, M. R. Murthy, L. Liljas, T. A. Jones, and B. Strandberg.** 1983. Structural comparisons of some small spherical plant viruses. J. Mol. Biol. **165:**711–736.
39. **Someya, Y., N. Takeda, and T. Miyamura.** 2002. Identification of active-site amino acid residues in the Chiba virus 3C-like protease. J. Virol. **76:**5949–5958.
40. **Sowa, M. E., W. He, K. C. Slep, M. A. Kercher, O. Lichtarge, and T. G. Wensel.** 2001. Prediction and confirmation of a site critical for effector regulation of RGS domain activity. Nat. Struct. Biol. **8:**234–237.
41. **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.
42. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.
43. **Vinje, J., and M. P. Koopmans.** 2000. Simultaneous detection and genotyping of "Norwalk-like viruses" by oligonucleotide array in a reverse line blot hybridization format. J. Clin. Microbiol. **38:**2595–2601.
44. **Yao, H., D. M. Kristensen, I. Mihalek, M. E. Sowa, C. Shaw, M. Kimmel, L. Kavraki, and O. Lichtarge.** 2003. An accurate, sensitive, and scalable method to identify functional sites in protein structures. J. Mol. Biol. **326:**255–261.

STAT1-dependent innate immunity to a Norwalk-like virus. Science **299:**1575–1578.