Quasispecies dynamics and molecular evolution of human norovirus capsid P region during chronic infection
Beatrice Carlsson,¹ A. Michael Lindberg,² Jesús Rodríguez-Díaz,¹,³ Kjell-Olof Hedlund,⁴ Bengt Persson⁵ and Lennart Svensson¹

¹Division of Molecular Virology, Department of Clinical and Experimental Medicine, Medical Faculty, 58185 Linköping University, Linköping, Sweden
²School of Pure and Applied Natural Sciences, University of Kalmar, Kalmar, Sweden
³Department of Microbiology, School of Medicine, University of Valencia, Valencia, Spain
⁴Swedish Institute for Infectious Disease Control, Solna, Sweden
⁵Bioinformatics, IFM, Campus Valla, Linköping University, Linköping, Sweden

In this novel study, we have for the first time identified evolutionarily conserved capsid residues in an individual chronically infected with norovirus (GGII.3). From 2000 to 2003, a total of 147 P1-1 and P2 capsid sequences were sequenced and investigated for evolutionarily conserved and functionally important residues by the evolutionary trace (ET) algorithm. The ET algorithm revealed more absolutely conserved residues (ACR) in the P1-1 domain (47/53, 88%) as compared with the P2 domain (86/133, 64%). The capsid P1-1 and P2 domains evolved in time-dependent manner, with a distinct break point observed between autumn/winter of year 2000 (isolates P1, P3 and P5) and spring to autumn of year 2001 (isolates P11, P13 and P15), which presumably coincided with a change of clinical symptoms. Furthermore, the ET analysis revealed a similar receptor-binding pattern as reported for Norwalk and VA387 strains, with the CS-4 and CS-5 patch (Norwalk strain) including residues 329 and 377 and residues 306 and 310, respectively, all being ACR in all partitions. Most interesting was that residues 343, 344, 345, 374, 390 and 391 of the proposed receptor A and B trisaccharide binding site (VA387 strain) within the P2 domain remained ACR in all partitions, presumably because there was no selective advantage to alter the histo blood group antigens (HBGA) receptor binding specificity. In conclusion, this study provides novel insights to the evolutionary process of norovirus during chronic infection.

INTRODUCTION
Norovirus is a major cause of acute epidemic non-bacterial gastroenteritis worldwide, affecting individuals of all ages (Glass et al., 2000; Hedlund et al., 2000). Human noroviruses (NoV) are organized into three open reading frames (ORFs), of which ORF2 (VP1) encodes the capsid protein (58–60 kDa; Green et al., 2001; Hardy, 2005). Each capsid protein is folded into two major domains, the S (shell) and the P (protruding) domains that are connected via an eight aa hinge region. The P domain is further subdivided into P1-1, P1-2 and P2, the last located on the most exterior part of the capsid and therefore predicted to bear antigenic determinants affecting the immunological response and host specificity, including the receptor-binding pocket (Cao et al., 2007; Prasad et al., 1999).

The NoV capsid protein has a high sequence diversity, probably due to poor proofreading and post-replicative activities associated with RNA replicases, but sequence diversity is probably also driven by immune selection mechanisms. These properties facilitate the development of a complex mixture of related viruses within an infected individual, referred to as quasispecies (Domingo et al., 1998; Domingo & Holland, 1997; Eigen, 1993).

Generally, a quasispecies population consists of a quantitatively dominant genome, surrounded by a cloudlike multitude of sequences differing from the majority sequence to various extents (Domingo et al., 1998). Within this heterogeneous cloud of genetic variants, mutants with increased capacity to adapt to various environments will be generated (Domingo et al., 1998; Forns et al., 1999).

In recent years the sequence-structure comparison evolutionary trace (ET) method, developed by Lichtarge and
coworkers, have been widely used to investigate conserved motifs in proteins, but also in virus (Chakravarty et al., 2005; Innis et al., 2000; Lichtarge et al., 1996a; Lichtarge & Sowa, 2002; Sowa et al., 2001). In this study, we have gained access to unique patient materials collected from a patient chronically infected with NoV. Currently very limited information is available concerning evolution of NoV during chronic infection. In this novel study we have for the first time identified evolutionary conserved aa residues and quasispecies behaviour in an individual chronically infected for 3 years with NoV.

**METHODS**

**Patient samples.** The sample donor is an immunosuppressed patient who became infected with NoV in June 2000, and developed a chronic infection with four to eight episodes of diarrhoea per day lasting for several years. Early in the illness the patient also suffered from nausea and vomiting, which disappeared in later course of the infection (Nilsson et al., 2003). In this study faecal samples were collected from August 2000 to March 2003 (termed P1 to P17). These samples have previously been investigated for NoV and found to be positive for electron microscopy and PCR (Nilsson et al., 2003).

**RNA purification.** Extraction and purification of viral RNA from stool samples were performed using QIAamp Viral RNA Mini Spin protocol (Qiagen) according to the manufacturer's instructions.

**RT-PCR of NoV capsid P-region (P1-1 and P2 domain).** The reverse transcriptase reaction was performed by mixing 5 μl RNA, 3 μl 5 × first strand buffer (Invitrogen), 0.75 μl GeneAmp dNTPs (Applied Biosystems), 0.5 μl RNase OUT (Invitrogen), 0.5 μl 10 pmol Noro P5' primer (5'-ATGCCTGTGGCCACCTACTGTGGATCA-3'), 0.5 μl SuperScript II RNase H' (Invitrogen) and RNase-free water to a total volume of 15 μl. The reaction was performed at 42 °C for 60 min, followed by inactivation of the enzyme at 95 °C for 5 min.

To amplify the NoV capsid P1-1 and P2 domain a PCR containing 5 μl 10 × Native Plus Pfu polymerase buffer (Stratagene), 1 μl 10 mM GeneAmp dNTPs (Applied Biosystems), 1 μl 10 pmol Noro P3' primer, 1 μl 10 pmol Noro P5' primer, 5 μl cDNA, 2.5 U native Pfu DNA polymerase (Stratagene) and RNase-free water to a final volume of 50 μl was performed. The Noro P3' and Noro P5' primers are designed to amplify aa 215–314 (covering the P1-1 and the P2 domain, by stretching from the 3'-end of the S-domain into the 5'-end of the P1-2 domain). PCR was performed at 94 °C for 5 min followed by 30 cycles at 94 °C for 30 s, 55 °C for 30 s and 72 °C for 1 min. Finally, the last part of the PCR cycle was a 10 min elongation followed by 30 cycles at 94 °C for 30 s, 50 °C for 30 s, 72 °C for 1 min and a 10 min long elongation at 72 °C.

**Cloning of NoV capsid P-region (P1-1 and P2 domain).** Purified PCR products (using Stratagene PCR purification kit) were cloned into the pPCR-Script Amp SK (+) cloning vector (Stratagene), and transformation into XL10-Gold Kan Ultracompotent cells (Stratagene) was performed according to the manufacturer’s instructions. Transformants were examined for the presence of recombinant plasmids containing the desired inserts by using blue and white screening and PCR. Transformed cells (white colonies) were added to a PCR mixture containing 2.5 μl 10 × PCR buffer-MgCl2 (Invitrogen), 1 μl 50 mM MgCl2 (Invitrogen), 0.5 μl 10 pmol M13 forward M13 primer (5'-GTTTTCCCAGTCAGCAG-3'), 0.5 μl 10 pmol M13 reverse primer (5'-CAGGAAAACAGCTATGAC-3'), 2.5 mM dNTPs (Invitrogen), 1 U Taq polymerase (Invitrogen) and RNase-free water to a final volume of 25 μl. PCR was performed at 94 °C for 5 min followed by 35 cycles of 94 °C for 30 s, 50 °C for 30 s, 72 °C for 1 min and a 10 min long elongation at 72 °C.

**Nucleotide sequencing.** On average, 16 independent colonies from each cloned sample containing the desired insert were collected for sequencing, generating 147 partial (nt 643–1239; P1-1 and P2 domain) capsid gene sequences (GenBank accession nos EU449135–EU449279, EU955452 and EU955453). Nucleotide sequencing of cloned inserts was performed according to the manufacturer’s instructions using DYEnamic Dye Terminator kit (GE Healthcare) with standard M13 sequencing primers, and a MegaBACE 500 automated sequencer.

For accuracy, inserts were sequenced twice in both forward and reverse direction. Complete sequences were obtained by assembling overlapping contigs with DNASTAR (DNASTAR). A consensus sequence was also created for each time point isolate using DNASTAR. When making consensus sequences, truncated capsid sequences were removed to avoid bias.

**Sequence analysis**

**ET analysis.** Amino acids were aligned using CLUSTALW1.8 (Thompson et al., 1994) with default settings on the European Bioinformatics Institute server. The aligned protein sequences were submitted to the Cambridge University ET server (Lichtarge & Sowa, 2002) for construction of a phylogenetic tree based on a distance matrix from the PHYLIP package (v3.6beta) computing the consensus sequences. The tree was divided by an evolutionary time cut-off line (ETC) into 10 partitions, 1 to X, in order of increasing divergence, where the sequences could be grouped into different evolutionary classes based on their sequence similarities (Chakravarty et al., 2005; Lichtarge et al., 1996a; b; Lichtarge & Sowa, 2002). A class was formed by sequences originating from the same node and belonging to the same partition. Within the classes in each partition the sequences were separately aligned and then consensus sequences were constructed for the partition. Then, the consensus sequences from each partition were compared in order to detect trace residues. Trace residues have been found in three types (Chakravarty et al., 2005; Lichtarge et al., 1996a; b; Lichtarge & Sowa, 2002): absolutely conserved residues (ACR), that stayed completely conserved in all classes; class-specific residues, that were conserved within a class but varied between classes; and neutral residues, which were not conserved in any class.

**PredictProtein: secondary structure prediction.** The amino acid sequence from reference (GenBank accession no. AY2247441) and P17-14 (accession no. EU449274) NoV capsid sequences were submitted to the PredictProtein database (Rost et al., 2004). For accuracy of results the proposed secondary structure predictions obtained from the PredictProtein database were also compared to the PSIPRED secondary structure prediction server (Jones, 1999) at University College London, resulting in similar predictions supporting the proposed structure (data not shown).

**Construction of phylogenetic trees.** The 134 sequences were translated to aa sequences, aligned by CLUSTAL W using default settings, manually inspected and edited and the nucleotide sequences were then aligned according to the aligned amino acid sequences using DAMBE (Thompson et al., 1994; Xia & Xie, 2001). Phylogenetic reconstruction was performed using several criteria, including neighbour-joining (NJ) and maximum-likelihood (ML) using the GTR model of sequence evolution (Felsenstein, 1981; Rodriguez et al., 1990; Saitou & Nei, 1987). The reliability of the trees was estimated using non-parametric bootstrap with 1000 pseudoreplicates (Felsenstein, 1985). The ML tree was constructed using the PhyML software where nucleotide frequencies, proportion of invariable sites and rate of substitutions at variable sites were assumed to follow a gamma distribution (Guindon & Gascuel, 2003).
Isolate consensus sequences were aligned using CLUSTAL W-multialign with default settings on the Mobyle portal (Pasteur Institute). The tree was then generated by the NJPlot program (available on the Pôle Bio-Informatique Lyonnais World Wide Web server) (Perriere & Gouy, 1996).

**Recombination analysis.** Nucleotide sequences were aligned using CLUSTAL W-multialign with default settings on the Mobyle portal (Pasteur Institute). The aligned sequences were then analysed by the RDP2 software to search for possible recombinants using different conventional recombination detection methods such as GENECONV, SIMPLOT, BOOTSCAN, MAXIMUM χ^2 and CHIMAERA (Martin et al., 2005).

**RESULTS**

### Human NoV show a time-dependent evolution and quasispecies behaviour during chronic infection

To investigate virus evolution and quasispecies behaviour of the NoV P1-1 to P2 capsid domain, nucleotide sequences from 147 clones were investigated covering August 2000 to March 2003. Sequence analysis revealed that every isolate (except isolate number 3, isolated in October year 2000, and number 13, isolated in July year 2001) contained sequences with significant deletions and/or insertions (in total 12 sequences, can be viewed as Supplementary Fig. S1, available in JGV Online). Interestingly, 5 out of these 12 clones were found in a single isolate (isolate 11, collected in May 2001, clones P11-2, P11-12, P11-20, P11-26 and P11-211). None of the 12 truncated capsid sequences (nor the full-length clone P13-2, GenBank accession no. EU443234) could be translated into functional proteins due to the presence of multiple stop codons and were therefore removed from further analysis. All the remaining 134 full-length capsid sequences (P1-1 to P2) were translated into continuous protein sequences and were investigated for quasispecies distribution and presence of evolutionary conserved residues.

A phylogenetic tree consisting of the 134 sequences revealed that sequences isolated at the same time point (and thus belonging to the same isolate) showed a trend of clustering together in monophyletic groups (clouds) supporting a distribution of sequences according to the quasispecies theory (Supplementary Fig. S2, available in JGV Online) (Domingo et al., 1998; Eigen, 1993). To analyse the evolutionary trend, consensus sequences for each isolate were constructed and used for generating a phylogenetic tree (Fig. 1). The tree shows that the capsid proteins evolved in a time-dependent manner; P1 (isolated in August 2000) is located at the bottom of the tree, followed in chronological order by isolate P3 (isolated in October 2000) to P17 (isolated in March 2003) and P15 (isolated in October 2001). A break point between early isolates obtained between the autumn/winter 2000 (isolates P1, P3, P5, P7, P9, P11, P13, P15 and P17) were constructed using the DNASTAR software (DNASTAR) and used to generate a phylogenetic tree. The tree was constructed by the NJPlot program, and branch lengths for the time point consensus sequences showing longest nucleotide distances (P17, 0.024; P9, 0.010) are presented. The arrow shows the time-dependent evolution and a bar indicates the isolates referred to as late, early and ‘break point’.

Fig. 1. Consensus sequences reveal quasispecies pattern and a time-dependent evolution. Consensus sequences for the capsid proteins of each time point (designated P1, P3, P5, P7, P9, P11, P13, P15 and P17) were constructed using the DNASTAR software (DNASTAR) and used to generate a phylogenetic tree. The tree was constructed by the NJPlot program, and branch lengths for the time point consensus sequences showing longest nucleotide distances (P17, 0.024; P9, 0.010) are presented. The arrow shows the time-dependent evolution and a bar indicates the isolates referred to as late, early and ‘break point’.

ET analysis

To further investigate the issue of invariable aa residues that remained conserved during the evolutionary process, the sequence-structure ET algorithm was used (Chakravarty et al., 2005; Lichtarge et al., 1996a, b; Lichtarge & Sowa, 2002). When comparing P1-1 to the P2 domain, most of the ACR (from partition I to partition X) were found in the P1-1 domain 47/53 (88 %) compared with 86/133 (64 %) in the P2 domain (Figs 2, 3).

**ET partition-dependent classes.** The phylogenetic tree was divided into 10 partitions (I to X), sorting the sequences into different groups (Fig. 2, Supplementary Table S1, available in
Fig. 2. Evolutionary tree based on a distance matrix from the PHYLIP package (v3.6beta) generated from the European Bioinformatics Institute server, of 134 amino acid sequences of the P1-1 and P2 NoV capsid domains. The tree is divided into 10 partitions (I to X) shown as thin vertical lines, where the ETC is increasing from partition I to X. Node 0 is denoted the parent root, which is then divided into child nodes 1 and 2. In the same way, node 1 is the parent node of child nodes 3 and 4 and so forth.
JGV Online). The first partition, I (Fig. 2), contained all the 134 sequences in one class (designated group A, Supplementary Table S1), which are further subdivided in the following partitions in order of increasing divergence.

The March 2003 isolate is distinct from earlier time point isolates. Node 0 creates two branches in partition II (branches 1 and 2) (Fig. 2, Supplementary Table S1). It is interesting to note that branch 2 (class B2, Supplementary Table S1) contains all the P17 sequences obtained in March 2003, while branch 1 (class B1, Supplementary Table S1) contains all the other isolates. This indicates that the last isolate obtained in March 2003 is highly divergent from isolates obtained from earlier time points, a feature also observed in the consensus analysis (Fig. 1). The P17 sequences in branch 2 are further divided in partition IV, into node 6 separating out the P17-14 sequence (GenBank accession no. EU449274) and node 5 containing the remaining P17 sequences. This highly divergent nature of the P17-14 clone was also observed during the phylogenetic tree analysis, where P17-14 clearly was divergent from all other sequences (Supplementary Fig. S2).

Different conservation pattern observed for isolates obtained during early and late stages of infection; an evolutionary break point observed between year 2000 and year 2001.

Branch 1 divides into branches 3 and 4 in partition IV, separating out the P5-16 (GenBank accession no. EU449179) isolate (branch 4) (Fig. 2). Branch 3 is further divided in partition VI, indicating an evolutionary break point observed between year 2000 and year 2001. Branch 1 divides into branches 3 and 4 in partition IV, separating out the P5-16 (GenBank accession no. EU449179) isolate (branch 4) (Fig. 2). Branch 3 is further divided in partition VI, indicating an evolutionary
break point, generating branches 7 and 8 that are
separating the remaining isolates into two major classes;
D1 and D2 (Fig. 2, Supplementary Table S1). The upper
cluster (class D1, Supplementary Table S1) originating
from branch 7, contains the majority of the sequences
isolated at a later stage of the infection (May, July and
October 2001). The lower class (branch 8, class D2,
Supplementary Table S1) on the other hand, consists of
sequences obtained during the first months of infection:
August, October and December of 2000. Sequences from
isolate P7 (January 2001) and P9 (March 2001) are mixed
within both branches 7 and 8. A relationship between the
January 2000 and the March 2001 isolates was also found
in the phylogenetic tree for consensus sequences (Fig. 1). In
conclusion, the ET analysis revealed a distinct break point
in the conservation pattern for isolates collected between
the autumn/winter of 2000 (isolate P1, P3, P5) and the
spring to autumn 2001 (isolate P11, P13 and P15). Isolate
P7 (obtained in January 2001) and isolate P9 (collected in
March 2001) lie within this break point and thus contain
sequences that resemble both early and late isolates.

Amino acid positions of the putative carbohydrate-
binding pocket are conserved during chronic infection.
Two conserved class-specific surface patches termed CS-4
(residues 329, 373, 375 and 377) and CS-5 (residues 306
and 310) located at the dimeric interface of the Norwalk
(GI) crystal structure and suggested to be involved in
carbohydrate binding, have been reported (Chakravarty
et al., 2005). To investigate if these positions are also
evolutionarily conserved during chronic infection, the ET
algorithm was used. Amino acid residues corresponding to
positions 329, 373, 375 and 377 in the GII.3 strain were in
chronological order 346 (Gly, ACR from partition I to X),
392 (class specific in partition X), 394 (Arg, ACR from
partition IX to X) and 396 (Thr, ACR from partition I to X),
(Fig. 3, Supplementary Table S2, available in JGV Online).
Concerning the CS-5 surface patch, residues 322 (Gly) and
326 (Asp) (corresponding to positions 306 and 310, respec-
tively) were ACR in all partitions (Supplementary
Table S2). Altogether, this shows that most positions
were identical in all positions between the two genogroups,
except for residue 343. A second plausible binding pocket
(apart from the one observed in ligand binding) includes
residues 390–392, 395 and 443. Out of these, residues 390
(Val, corresponding to 402) and 391 (Asp, corresponding to
403) were found ACR in all partitions (Fig. 3, Supplementary
Table S2). Altogether, this shows that most positions
involved in potential carbohydrate binding found in GII.4
virus remain conserved during evolution in an individual
with constant histo blood group antigens (HBGA).

Earlier studies have suggested that RGD (located at the
beginning of the P2 domain) and NGR motifs (located at the
end of the P1-1 domain) might be of structural importance
for NoV binding to HBGAs (Tan et al., 2003). In this study,
the asparagine and glycine residues of the NGR motif (aa
263–265, GII.3 numbering) were ACR from partition I to X,
while the arginine residue was non-conserved. Concerning
the RGD motif (aa 287–289, GII.3 numbering), the arginine
and glycine residues were ACR from partition I to X, and
the aspartic acid residue in position 289 was substituted to a
ACR valine residue in all partitions.

During chronic infection most of the amino acid
substitutions are conserved or semi-conserved
and occur in the P2 domain
From the ET data and the phylogenetic analysis, the P17-14
close (GenBank accession no. EU449274) was most distant
from the reference sequence (GenBank accession no.
AY247441) (Fig. 2, Supplementary Fig. S2). Multiple
sequence alignment revealed 33 nt substitutions (data not
shown), scattered throughout the P1-1 and P2 domain
sequence. These changes occurred at an equal frequency at
all three codon positions. The 33 nt changes resulted in a
total of 18 aa alterations. Out of the 18 substitutions, 12
were conserved or semi conserved, and most surprisingly
all except one occurred in the P2 region (Fig. 4). The
substitution situated outside the P2 region was located at
residue 257 of the P1-1 domain. To examine if these 18 aa
mutations could possibly affect the protein structure,
secondary structure predictions were performed using the
PredictProtein server (Rost et al., 2004). This revealed only
minor changes in the secondary structure (Fig. 4), by
removal of a small b-sheet in position 370–371 and
affecting the length of a b-sheet in position 377–382 in the
P2 domain. These alterations were probably caused by the
nearby point mutations; position 368: Thr→Ser (polar,
hydrophilic), position 369: Gly→Asp (non-polar,
hydrophobic→charged, hydrophilic) and position 381: Phe→Ser (non-polar, hydrophobic→polar, hydrophilic).

No recombination detected within the P1-1 and
P2 capsid domains
Since RNA recombination is believed to be one of the
major driving forces in viral evolution (Worobey &
Holmes, 1999), including NoV (Bull et al., 2005, 2007;
Phan et al., 2007), we searched the 134 full-length (P1-1 and P2 capsid domains) NoV capsid nucleotide sequences for recombination events. The sequences were first aligned using CLUSTAL W-multialign with default settings on the Mobyle portal (Pasteur Institute), and then analysed by the RDP2 software to search for possible recombinants using different conventional recombination detection methods such as GENECONV, SIMPLOT, BOOTSCAN, MAXIMUM χ² and CHIMAERA (Martin et al., 2005). Some possible recombination events were located (e.g. involving the P17-14 isolate); however, none of the detected events were found to be significant by the recombination detection program used.

**DISCUSSION**

Currently there is no information about evolution of functionally conserved residues in the human NoV capsid during chronic infection. The aim of this study was to investigate quasispecies dynamics and identify evolutionarily conserved and thereby functionally important aa of the NoV capsid during chronic infection. By using the ET method, a direct connection between the conservation patterns of amino acid in aligned sequences and their functional importance can be investigated (Chakravarty et al., 2005; Lichtarge et al., 1996a; Lichtarge & Sowa, 2002). The method was most recently used to search for conserved capsid motifs between 56 different NoV strains isolated from different individuals (Chakravarty et al., 2005). In the P2 domain they found two class-specific surface patches, indicating a putative carbohydrate-binding site, one consisting of residues 329, 373, 375 and 377 and the other consisting of residues 306 and 310 (Norwalk numbering) (Chakravarty et al., 2005). Recently, Bu and coworkers confirmed by mutational analysis the importance of residues 329 and 377 in binding of Norwalk P-particles to the A trisaccharide (Bu et al., 2008). Similarly, we found residues 329 (Glu), 373 (Pro) and 377 (Ser) ACR in all partitions, while residue 375 was class-specific from partition I to V, and ACR (Leu) from partition VI to X. Further residues located near the CS-4 surface patch are residues in position 267, 322, 327, 331, 333, 334, 374 (Chakravarty et al., 2005). All these positions, except for 333 and 341, were found ACR from partition I to X, strongly suggesting that these positions play an important role in receptor binding.

When comparing the potential trisaccharide A and B interacting positions (343–345 and 374) (Cao et al., 2007), we found residues 344 (Val), 345 (Phe) and 374 (Pro) to be ACR in all partitions, and residue 343 to be class-specific from partition II to partition X. A second plausible binding pocket (except from the one located in the domain interface) includes residues 390–392, 395 and 396 (Cao et al., 2007). The importance of residues 393–395 for HBGA binding was recently confirmed by Lindesmith et al. (2008). Out of these, residues 390, 391 and 395 were found ACR (Asn, Lys and Phe, respectively) in all partitions, and residue 392 to be class specific in partition X.
Concerning the additional HBGA-binding site, including NGR and RGD motifs suggested to be involved in NoV binding (Tan et al., 2003), we found only certain parts of the NGR or RGD site conserved. Only the asparagine and glycine residues of the NGR site and the arginine and glycine of the RGD motifs remained conserved. Altogether, our data show similar receptor binding pattern to that previously found with other NoV strains (Cao et al., 2007; Chakravarty et al., 2005). However, it is important to remark that the virus investigated in this study belongs to genogroup II.3, and may therefore have different or additional binding pattern as compared with the VA387 (GII.4) and Norwalk virus (GI.1) strains (Tan & Jiang, 2005). Indeed, Bu and coworkers recently showed that, although Norwalk and VA387 strain (both bind A and H antigens in related regions within the P2 domain) have similar binding patterns, the interaction with the receptor includes different amino acids (Bu et al., 2008).

Among the absolutely conserved residues (from partition I to partition X) 47/53 residues (88%) were located in the P1-1 domain compared with 86/133 (64%) in the P2 domain. Based on secondary structure predictions of the P17-14 clone (the sequence diverging most from the reference sequence), it seems that the overall secondary structure has been preserved during evolution, probably caused by the high frequency of conserved or semi-conserved aa (12/18). In terms of biological relevance, however, it should be noted that the structural alterations of the β-sheets occurred in close proximity to proposed receptor-interacting residues in position 373–375 and 377 (Cao et al., 2007; Chakravarty et al., 2005), a fact that hypothetically may alter the conformation of a binding pocket.

Compared with Chakravarty et al. (2005) we found a higher number of ACR in the P2 domain and some differences within the surface patch residues (discussed above). The most reasonable explanation for our different observations is that we investigated evolutionarily conserved motifs in the capsid of a chronically infected individual with constant ABO, Lewis and secretor status, properties that affect NoV susceptibility (Le Pendu et al., 2006; Lindesmith et al., 2003; Thorven et al., 2005). In contrast, Chakravarty and coworkers investigated conserved motifs among different strains from different individuals, presumably also different ethnic groups, each with unique immunity and HBGA properties.

Since the viral evolution in this study was restricted to a single individual, it can be speculated that the receptor-binding specificity remained intact, and thus kept the P2 receptor-binding domain evolutionarily conserved. This might also explain the relatively high content of ACR (64%) residues found in the P2 domain, thus reflecting limited sequence diversity in this area. Lindesmith and coworkers have recently shown that the HBGA binding site of the NoV capsid is under heavy immune selection, and thereby probably allows GIL4 virus to persist by using positive selection and altering its HBGA specificity over time (Lindesmith et al., 2008). However, while no selective pressure to alter receptor specificity likely existed within the immunologically suppressed individual, immune evasion might have occurred. The patient in this study had normal serum concentrations of IgA, IgG and IgM (Nilsson et al., 2003), suggesting an intact humoral immunity, which could have been a driving factor of capsid evolution. This hypothesis is strengthened by the observed time-dependent evolution, suggesting the presence of a selective driving force giving rise to new and probably more adapted viruses with advantage over earlier virus variants.

When further analysing the phylogenetic tree generated from the ET, two interesting observations were made. The first observation was that the ET and consensus analysis indicate that the molecular makeup of the viral population was significantly altered between autumn/winter 2000 (isolates P1, P3 and P5) and spring to autumn 2001 (isolates P11, P13 and P15), thus implying an evolutionary break point between January 2001 (isolate P7) and March 2001 (isolate P9), which coincided with a change of clinical symptoms. Most interestingly, during the early part of infection the patient suffered from severe vomiting, nausea and diarrhoea, while in later stages of infection the vomiting and nausea subsided (Nilsson et al., 2003). The overall conclusion is that the sequences have evolved in a time-dependent manner where each isolate has descended from its predecessor in time, a feature observed elsewhere (Siebenga et al., 2007). This feature was also confirmed by the phylogenetic tree consisting of the 134 sequences (Supplementary Fig. S2), which revealed that sequences isolated at the same time point (and thus belonging to the same isolate) showed a trend of clustering together in swarms typical for a quasispecies distribution (Domingo et al., 1998; Eigen, 1993). Another feature supporting this finding is that 5 out of the 12 truncated capsid sequences were found during February 2001 (isolate P11), suggesting a prominent evolution activity during this period. This finding is in concordance with other studies, suggesting that the P2 domain of GII.4 virus is evolving in a time-dependent manner showing epochary evolution with periods of slow and high evolutionary rates over time (Lindesmith et al., 2008; Siebenga et al., 2007).

The second feature observed during the ET analysis was that NoV capsids isolated at the end of the collection period are distant from earlier isolates. The P17 isolate had the longest nucleotide distance compared with the other consensus sequences, which further supports epochal evolution. This finding is consistent with the idea of chronically infected patients as a reservoir for virus evolution, and thus a source for the emergence of new virus variants (Gallimore et al., 2004; Nilsson et al., 2003; Siebenga et al., 2007).

The evolutionary pattern and PCR analysis of the viral polymerase region suggests that the chronically infected individual is infected with only one viral strain (belonging to GII.3). However, heterogeneity in the initial infecting virus cannot be ruled out.
Since RNA recombination is believed to be one of the major driving forces in viral evolution, (Worobey & Holmes, 1999) including NoV (Bull et al., 2005, 2007; Phan et al., 2007), we searched the 134 full-length NoV capsid sequences (597 nt from the P1-1and P2 capsid domains) for recombination events. Some possible recombination events were located (e.g. the P17-14 isolate), but none of the detected events were found to be significant by the recombination detection program. While Lindesmith and coworkers have identified a recombination breakpoint at residue 265 located near the P1-1/P2 domain interface of NoV GII.4 strains (Lindesmith et al., 2008), this recombination break point was not found in this study. The absence of detected recombinants might be due to the fact that we did not have access to ORF1/ORF2 overlap sequences, previously found associated with NoV recombination break points (Bull et al., 2005, 2007; Phan et al., 2007). Thus, we cannot rule out that recombination occurred at the ORF1/ORF2 junction.

ACKNOWLEDGEMENTS

This work was supported by the Swedish Research Council grant 10392.

REFERENCES


