Characterization and genetic variability of *Hepatitis A virus* genotype IIIA

Kathrine Stene-Johansen, Tom Øystein Jonassen† and Kjell Skaug

Division of Infectious Disease Control, Norwegian Institute of Public Health, PO Box 4404, Nydalen, NO-0403 Oslo, Norway

Molecular epidemiological studies of hepatitis A outbreaks in Norway showed the emergence of *Hepatitis A virus* (HAV) genotype IIIA in association with parenteral transmission among haemophiliacs and intravenous drug users. The complete genomic sequence of one of these outbreak isolates, NOR-21, was determined. This is the first complete genomic sequence of HAV genotype IIIA. Phylogenetic analysis showed that genotype IIIA/NOR-21 was genetically distinct from the other human and simian genotypes. Phylogenetic analysis of the nucleotide sequences clearly distinguished the different HAV genotypes, regardless of the genomic region used for analysis, whereas the amino acid sequences showed a more vague distinction between human HAV genotypes I and II. In particular, the inferred phylogeny based on the capsid proteins showed that the human HAV strains were related more closely to each other than to the simian strains. The greatest variability and clearest distinction between genotypes were observed for the polymerase gene. The outbreak isolates of HAV genotype IIIA in this study showed greater nucleotide variability than is generally seen in outbreaks of genotype I. This high nucleotide variability, which may be characteristic of this HAV genotype, the mode of transmission in this outbreak or parallel introductions, is discussed.

**INTRODUCTION**

*Hepatitis A virus* (HAV) is the only member of the genus *Hepatovirus* within the family *Picornaviridae* (Melnick, 1992). The positive-sense, single-stranded RNA genome (7·5 kb) consists of (i) a highly conserved 5′ non-translated region (NTR); (ii) a single open reading frame (ORF) encoding a polyprotein of 2227 aa; and (iii) a 3′ NTR (Brown et al., 1991; Melnick, 1992). The single ORF is divided into three functional regions termed P1, P2 and P3. P1 encodes the capsid polypeptides VP1–VP4, whereas P2 and P3 encode the non-structural polypeptides (Totsuka & Moritsugu, 1999). HAV has been classified into three human (I–III) and three simian (IV–VI) genotypes (Robertson et al., 1992; Costa-Mattioli et al., 2003a). An HAV genotype is defined as a group of viruses with >85% nucleotide sequence identity. The HAV genotypes are further classified into subgenotypes with sequence variability of <7·5%. The only two strains representing genotype II, isolates CF-53 and SLF88, were recently classified as subgenotypes IIA and IIB, respectively (Ching et al., 2002; Costa-Mattioli et al., 2003b; Lu et al., 2004). Previously, SLF88 was classified as genotype VII. Almost all human HAVs belong to genotypes I and III, genotype I being the most prevalent, comprising at least 80% of circulating human strains (Robertson et al., 1992; Stene-Johansen et al., 1999; Pina et al., 2001; Costa-Mattioli et al., 2003a; Nainan et al., 2005). HAV genotype III is unique in that it includes strains that have been recovered from both humans and non-human primates. The first virus identified as genotype IIIA (PA21) was recovered from a feral owl monkey shortly after its capture and admission to a primate-holding facility in Panama in 1980 (Brown et al., 1989). PA21 was initially considered to be a simian virus as it was found in non-human primates, as well as being genetically distinct from the human HAVs known at that time. Closely related viruses have since been shown to be responsible for human infections in South-East and Central Asia (India, Nepal, Sri Lanka and Malaysia), the USA and Europe (Khanna et al., 1992; Robertson et al., 1992; Stene-Johansen et al., 1999; Costa-Mattioli et al., 2001; Pina et al., 2001). HAV genotype IIIA has been associated with outbreaks among intravenous drug users (IVDU) in Sweden (1979–1985), Norway (1997–1998), the UK (1998–1999) and Estonia (1998–99) (Robertson et al., 1992; Stene-Johansen et al., 1999; O’Donovan et al., 2001; Tallo et al., 2003). Only a few cases of hepatitis A have been associated with genotype IIB (Robertson et al., 1992). HAV genotype IIIA has re-emerged in Europe in the few last years and has particularly been
associated with outbreaks among IVDU. Surveillance of hepatitis A in Norway during outbreaks has suggested a greater genetic variability within genotype IIIA than within genotype IA. So far, limited information has been available on these emerging strains. Sequence information on the non-structural proteins is of particular interest, as the P1 region is subjected to strong negative selection of amino acid replacements (Costa-Mattioli et al., 2002; Sánchez et al., 2003b) and hence is conserved among HAV genotypes.

We have presented here characterization of outbreaks caused by genotype IIIA and the first complete genome sequence of HAV genotype IIIA (NOR-21). In this work, we have compared these HAV genotype IIIA isolates with the other human and simian genotypes to elucidate the characteristics and origin of this HAV genotype.

METHODS

Samples. Serum samples were obtained from patients with acute hepatitis A during an outbreak among IVDU in Norway during 1995–1999 (Stene-Johansen et al., 1999). In addition, serum samples were obtained from two patients in an outbreak of hepatitis A among four haemophiliacs, as well as samples of the two batches of coagulation factor VIII (Octavi; Octapharma) that were given to four haemophiliacs, as well as samples of the two batches of 1995–1999 (Stene-Johansen et al., 1999). In addition, serum samples were obtained from a Norwegian patient with hepatitis A in 1997. This patient was part of an outbreak of hepatitis A among IVDU from August to October 1998 (Stene-Johansen et al., 1999). Two co-circulating HAV genotypes were found during this outbreak. Genotype IA (NOR-17) was found in 73 cases from the beginning of the outbreak in 1995 until the summer of 1998 and was conserved throughout this period (Stene-Johansen et al., 1998, 1999). HAV genotype IIIA was detected in 17 cases during a period of 15 months from August 1997 (Fig. 1). These genotype IIIA viruses showed up to 4% nucleotide variability, but no amino acid variability, within the 350 nt region of the VP1–2PA junction of the HAV genome used for surveillance by molecular epidemiology. Virus identical to NOR-21 was detected in five cases associated with IVDU from August to November 1999, virus identical to NOR-21 was also detected in serum

RESULTS

Epidemiological background

The presented HAV genotype IIIA (NOR-21) was obtained from a Norwegian patient with hepatitis A in 1997. This patient was part of an outbreak of hepatitis A among IVDU in Norway during 1995–1999 (Stene-Johansen et al., 1999). Two co-circulating HAV genotypes were found during this outbreak. Genotype IA (NOR-17) was found in 73 cases from the beginning of the outbreak in 1995 until the summer of 1998 and was conserved throughout this period (Stene-Johansen et al., 1998, 1999). HAV genotype IIIA was detected in 17 cases during a period of 15 months from August 1997 (Fig. 1). These genotype IIIA viruses showed up to 4% nucleotide variability, but no amino acid variability, within the 350 nt region of the VP1–2PA junction of the HAV genome used for surveillance by molecular epidemiology. Virus identical to NOR-21 was detected in five cases associated with IVDU from August to October 1998 (Fig. 1). During an outbreak of hepatitis A among four haemophiliacs in Norway in July–August 1999, virus identical to NOR-21 was also detected in serum
from two of these patients, as well as in two batches of coagulation factor VIII given to these haemophiliacs.

Genetic characterization and comparison

The genomic sequence determined for the HAV genotype IIIA strain NOR-21 isolate included 7385 nt, encompassing nt 80–7478 relative to strain HM-175. The sequence was generated by primer walking on HAV RNA isolated from serum. The last few bases of the 5′ NTR present in genotype I sequences were not obtainable. By nucleotide comparison of HAV strains, the 5′ NTR was found to be highly conserved (data not shown). Compared with the HM-175 sequence, there was 91 % sequence similarity within the 5′ NTR region and 16 nt missing within the pyrimidine-rich tract. The predicted secondary structure of the 5′ NTR (Brown et al., 1991) was supported by compensatory substitutions within stems of the NOR-21 RNA except for positions 310, 372 and 479, where there were conflicting base pairings. The virus NOR-21, was, by comparison with earlier published sequences, shown to belong to HAV genotype IIIA. It was related most closely (97 %) to the earlier published sequences, shown to belong to HAV genotype IIIA. It was related most closely (97 %) to the early sequences recovered from an outbreak of hepatitis A among IVDU in Norway in 1997–1998 in comparison with IVDU outbreak strains from Eastbourne and reference strains HMH and PA21. A 342 bp region in the VP1–2PA junction was used for construction of the tree.

Fig. 1. Genetic relationship between genotype IIIA strains recovered from an outbreak of hepatitis A among IVDU in Norway in 1997–1998 in comparison with IVDU outbreak strains from Eastbourne and reference strains HMH and PA21. A 342 bp region in the VP1–2PA junction was used for construction of the tree.

Comparisons among HAV genotype IIIA strains

The HAV genotype IIIA strains (GA76, PA21, P27, HMH and NOR-21) showed 3–4 % nucleotide variation within the P1 region. Ten amino acid changes were unique to these HAV genotype IIIA strains compared with the other HAV genotypes in the P1 region (Table 2). Equivalent deletions were seen in the 5′ NTR sequences available for the NOR-21, PA21 and HMH strains. Five amino acids were different between the NOR-21 strain and the PA21 strain within the P1 and partial P2 region (975 aa) available for PA21. The HMH and NOR-21 strains showed 96·5 % nucleotide identity and four amino acid differences within the nearly complete coding region (6496 nt) available for the HMH strain. Three of these amino acids clustered in VP1 (positions 202–208) in HMH, whereas the other amino acid substitution (P2C-132) was only present in NOR-21.

DISCUSSION

Characterization of HAV genotype III is important for understanding the evolution and origin of HAV, as well as providing information on the dissemination of HAV and its pattern of transmission. NOR-21 is the first complete genome sequence of an HAV genotype IIIA and contributes to a more distinct characterization of this genotype. The magnitude of nucleotide and amino acid sequence diversity among different HAV genotypes varied among the individual genes (Table 1 and Fig. 3). The capsid proteins were highly conserved, particularly among the human HAV strains (Table 1 and Fig. 3). This is consistent with the existence of a single serotype of HAV (Sánchez et al., 2003a, b). In agreement with previously published data, we observed that similarity at the amino acid level was not reflected at the nucleotide level (Table 1), as can be seen by the lack of non-synonymous mutations (Costa-Mattioli et al., 2002; Sánchez et al., 2003b). The capsid proteins are the most conserved regions in the HAV genome, due to negative selection. The surface proteins are generally regarded as reflecting the genetic relatedness of the virus, but for HAV this relatedness might be biased. We have reported that the NOR-17 strain was completely conserved for several years during an outbreak among IVDU in Norway (Stene-Johansen et al., 1998, 1999). Molecular epidemiological studies of HAV are based on sequences within the capsid region (Costa-Mattioli et al., 2003a) and show that outbreak strains of genotype I are generally very conserved. In addition to the capsid region, the 3B and 3C sequence are shown in Fig. 2.

Table 2. Nucleotide and amino acid differences within the P1, P2 and P3 regions.

<table>
<thead>
<tr>
<th>Amino acid exchanges</th>
<th>P1 exchanges</th>
<th>P2 exchange</th>
<th>P3 exchanges</th>
</tr>
</thead>
<tbody>
<tr>
<td>342</td>
<td>10</td>
<td>24</td>
<td>21</td>
</tr>
</tbody>
</table>

The amino acid differences between representatives of each HAV genotype relative to the HM-175 strain for the complete coding region are illustrated in Fig. 3. The NOR-21 strain showed 64 unique amino acids within the coding region when compared with all available full-length sequences, as well as the P1 region for genotype IV. Half of these (33) were non-conservative substitutions dispersed within the P1, P2 and P3 region with five, 11 and 17 substitutions, respectively. A single amino acid insertion at P3C-215, as well as 24 amino acid substitutions, were common to NOR-21 and the simian strains.
proteins and the middle part of the 2C protein showed very few amino acid substitutions (Table 1 and Fig. 3), whereas numerous amino acid changes were seen within the polymerase 3D gene. The 3A protein of NOR-21 was the most divergent protein at both the nucleotide and amino acid levels.

**Fig. 2.** Phylogeny of HAV strains. (a) Phylogeny of P1 nucleotide sequences; (b) phylogeny of P1 amino acid sequences; (c) phylogeny of the polymerase 3D gene nucleotide sequences; (d) phylogeny of the polymerase 3D gene amino acid sequences. Numbers on nodes represent bootstrap values (950 replicas).

**Fig. 3.** Amino acid differences among representatives of the individual HAV genotypes and the HM-175 prototype strain. Vertical bars indicate positions of changed amino acids.
Phylogeny based on the nucleotide sequences clearly distinguished the different genotypes, regardless of the region used for analysis (Fig. 2). The distinction between HAV genotypes I and II was less clear when the amino acid sequences were analysed. When the polymerase gene was used for analysis, the clustering of subgenotypes was less obvious, and for the capsid region, even the separation into genotype I and II was vague. The inferred phylogeny based on the amino acid sequence of the capsid region also showed that the human HAV strains were related more closely to each other than to the simian strains. Similar clustering of the human HAV strains has also been obtained by phylogenetic analysis of non-synonymous sites (Ching et al., 2004).

The nucleotide comparisons in Table 1 showed a slightly higher variability in the VP1 region than in the polymerase 3D gene. However, phylogenetic analysis of the nucleotide sequences clearly distinguished the different genotypes, although the distinction was better for the polymerase gene than for the VP1 gene. Unique amino acid substitutions in the polymerase gene showed great variability among genotypes, whereas VP1 was highly conserved, particularly among the human genotypes. In addition, molecular epidemiological studies showed that VP1 is conserved. The conserved amino acid sequence in VP1 is subjected to negative selection and this will necessarily be reflected in the nucleotide sequence. Therefore, it will not reflect the genetic relatedness of HAV strains as much as regions that are not subjected to such structural constraints. Summarizing these observations, the polymerase gene might be a better candidate for the distinction of outbreak strains. So far, regions within VP1 have been useful for distinction of strains but, in light of our findings, the polymerase gene might be more suitable.

The origin of HAV genotype IIIA strains has been debated due to their diverse origin, but close sequence similarity (Brown et al., 1989; Robertson, 2001). PA21, isolated in Panama in 1980 from a feral owl monkey, was the first isolate of genotype III and the only isolate of simian origin (Brown et al., 1989). HAV genotype III has since been detected among humans in geographically diverse regions. The subgenotype IIIA seems to be endemic in South-East and Central Asia (Robertson et al., 1992). The best-characterized genotype III isolates are the GA76 isolate from an outbreak in Georgia in 1976 (Khanna et al., 1992), and the P27 strain (Costa-Mattioli et al., 2002), the NOR-21 strain and the HMH strain isolated recently in Europe. The NOR-21 strain shared 24 unique amino acid substitutions, as well as an insertion at the C terminus of the 3C protein, with the simian strains and showed considerably more amino acid changes than the two human isolates of genotype II relative to HM-175 (Fig. 3). Phylogenetic analysis and nucleotide and amino acid comparisons showed that genotype IIIA was more or less equally distant from the other human and simian genotypes. In contrast, analysis of the amino acid sequence of the capsid region suggested a closer relationship between the human strains, which may be explained by adaptation to the human host. The widespread and increasing distribution of this genotype among humans, and their close relatedness to PA21, suggest strongly that genotype III is of human origin.

We wondered whether the genetic variability was higher for genotype IIIA than genotype IA. During the outbreak of hepatitis A in the IVDU community in Norway in the period 1995–1999, we found two circulating clades of HAV. In the genotype IA outbreak, starting in 1995, the virus (NOR-17-like isolates) was highly conserved throughout a 4·5 year period (data not shown), whereas the genotype IIIA outbreak, starting in 1997, showed up to 4% nucleotide variability within a much shorter period. The general conservation of outbreak strains of genotype IA was not seen for HAV genotype IIIA. The difference in sequence

**Table 1.** Percentage identity of nucleotide and amino acid sequences between HAV genotype IIIA strain NOR-21 and strains representing the other HAV genotypes

<table>
<thead>
<tr>
<th>Protein</th>
<th>HM175 AA</th>
<th>GBM AA</th>
<th>CF53 AA</th>
<th>SLF88 AA</th>
<th>Cy145 AA</th>
<th>AGM27 AA</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>80</td>
<td>97</td>
<td>80</td>
<td>97</td>
<td>79</td>
<td>98</td>
</tr>
<tr>
<td>P2</td>
<td>80</td>
<td>97</td>
<td>80</td>
<td>97</td>
<td>79</td>
<td>98</td>
</tr>
<tr>
<td>P3</td>
<td>82</td>
<td>98</td>
<td>83</td>
<td>98</td>
<td>81</td>
<td>97</td>
</tr>
<tr>
<td>P1</td>
<td>78</td>
<td>96</td>
<td>77</td>
<td>95</td>
<td>77</td>
<td>96</td>
</tr>
<tr>
<td>P2</td>
<td>77</td>
<td>93</td>
<td>77</td>
<td>93</td>
<td>78</td>
<td>93</td>
</tr>
<tr>
<td>P3</td>
<td>79</td>
<td>92</td>
<td>78</td>
<td>91</td>
<td>80</td>
<td>92</td>
</tr>
<tr>
<td>P1</td>
<td>75</td>
<td>91</td>
<td>78</td>
<td>92</td>
<td>75</td>
<td>93</td>
</tr>
<tr>
<td>P2</td>
<td>78</td>
<td>94</td>
<td>77</td>
<td>94</td>
<td>80</td>
<td>93</td>
</tr>
<tr>
<td>P3</td>
<td>79</td>
<td>92</td>
<td>78</td>
<td>91</td>
<td>80</td>
<td>92</td>
</tr>
<tr>
<td>P1</td>
<td>78</td>
<td>94</td>
<td>77</td>
<td>94</td>
<td>79</td>
<td>94</td>
</tr>
<tr>
<td>P2</td>
<td>77</td>
<td>93</td>
<td>77</td>
<td>93</td>
<td>78</td>
<td>93</td>
</tr>
<tr>
<td>P3</td>
<td>79</td>
<td>92</td>
<td>78</td>
<td>91</td>
<td>80</td>
<td>92</td>
</tr>
<tr>
<td>P1–P3</td>
<td>79</td>
<td>94</td>
<td>78</td>
<td>94</td>
<td>79</td>
<td>94</td>
</tr>
</tbody>
</table>

**Table 2.** Unique amino acids for HAV genotype IIIA (strains GA76, PA21, P27, HMH and NOR-21) in the P1 region

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Capsid protein (amino acid position)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IIIA</td>
<td>P R A H K V Q T L K</td>
</tr>
<tr>
<td>V</td>
<td>T K P G E I R S S R</td>
</tr>
<tr>
<td>IV</td>
<td>S K P S E P R S G R</td>
</tr>
<tr>
<td>II</td>
<td>V I K P G E M R S A R</td>
</tr>
<tr>
<td>I</td>
<td>V I K P G E M R S A R</td>
</tr>
</tbody>
</table>

The origin of HAV genotype IIIA strains has been debated due to their diverse origin, but close sequence similarity (Brown et al., 1989; Robertson, 2001). PA21, isolated in Panama in 1980 from a feral owl monkey, was the first isolate of genotype III and the only isolate of simian origin (Brown et al., 1989). HAV genotype III has since been detected among humans in geographically diverse regions. The subgenotype IIIA seems to be endemic in South-East and Central Asia (Robertson et al., 1992). The best-characterized genotype III isolates are the GA76 isolate from an outbreak in Georgia in 1976 (Khanna et al., 1992), and the P27 strain (Costa-Mattioli et al., 2002), the NOR-21 strain and the HMH strain isolated recently in Europe. The NOR-21 strain shared 24 unique amino acid substitutions, as well as an insertion at the C terminus of the 3C protein, with the simian strains and showed considerably more amino acid changes than the two human isolates of genotype II relative to HM-175 (Fig. 3). Phylogenetic analysis and nucleotide and amino acid comparisons showed that genotype IIIA was more or less equally distant from the other human and simian genotypes. In contrast, analysis of the amino acid sequence of the capsid region suggested a closer relationship between the human strains, which may be explained by adaptation to the human host. The widespread and increasing distribution of this genotype among humans, and their close relatedness to PA21, suggest strongly that genotype III is of human origin.

We wondered whether the genetic variability was higher for genotype IIIA than genotype IA. During the outbreak of hepatitis A in the IVDU community in Norway in the period 1995–1999, we found two circulating clades of HAV. In the genotype IA outbreak, starting in 1995, the virus (NOR-17-like isolates) was highly conserved throughout a 4·5 year period (data not shown), whereas the genotype IIIA outbreak, starting in 1997, showed up to 4% nucleotide variability within a much shorter period. The general conservation of outbreak strains of genotype IA was not seen for HAV genotype IIIA. The difference in sequence
variability could be the result of different selective pressures, as well as the mode of transmission. Virus that is transmitted by the faecal–oral route must cross several barriers and is therefore subjected to strong selection. By parenteral transmission, there is a direct transfer from blood to blood, with limited barriers, so we assume that the selection pressure is lower than via the faecal–oral route. HAV genotype IIIA has been associated with parenteral transmission through blood products. This genotype also caused outbreaks in Sweden, the UK, Estonia and Norway in IVDU communities, where parenteral transmission of virus infections through needle-sharing practices is common. The sequence variability seen for HAV genotype IIIA might therefore be a result of parenteral transmission.

HAV genotype IIIA has been much less prevalent than genotype I and has not been detected in Norway before. We therefore assumed that the genotype IIIA strains had evolved from the same outbreak strain and not from parallel introductions of different HAV genotype IIIA strains to the same community. However, phylogenetic analysis, including PA21 and HMH as reference strains, clustered the outbreak strains into two groups, suggesting at least two introductions of genotype IIIA strains. Increased immunity in the IVDU community at the time of introduction of HAV genotype III may also suggest the occurrence of several small, limited outbreaks due to the lack of susceptible IVDU to HAV. HAV genotype IIIA seems to be endemic in parts of Asia and might therefore have been imported from these countries by transmission through travellers (IVDU or drug dealers) who import drugs illegally into Europe. If this is the case, it might explain why this genotype is so frequently associated with IVDU. The close relationship between the isolates from IVDU in the UK and Norway may support this theory by import from the same high-endemic regions on different occasions. Alternatively, the close relationship may indicate a more global spread of HAV among IVDU communities in Europe. The incidence of HAV is declining in many parts of the world (Jacobsen & Koopman, 2004). HAV genotype IIIA is emerging in Europe, especially among IVDU. Transmission of HAV in association with illegal import of drugs from endemic regions might cause an increase in the incidence of infection with HAV genotype IIIA strains. Whether the variability observed in the genotype IIIA outbreak in Norway during 1997–1998 was due to the nature of this genotype, its mode of transmission or frequent import from endemic regions needs to be further studied. Close surveillance by molecular epidemiology is important to understand better the pattern of transmission of this emerging genotype in Europe.

ACKNOWLEDGEMENTS

In our Division of Infectious Disease Control, we acknowledge the technical assistance of Hilde Elshaug and Einar S. Berg, and Hans Blystad for the provision of epidemiological data.

REFERENCES


Felsenstein, J. (1993). PHYLIP (phylogeny inference package), version 3.5c, Department of Genome Sciences, University of Washington, Seattle, USA.


