6.0 Discussion
Contents

6.1 Genetic relatedness of HAV strains 82
6.2 Recombination in HAV 82
6.3 Non-coding region 83
6.4 Coding region 84
6.5 Phylogenetic analysis 85
6.6 Analysis of GBS-IND strain 85
6.7 HAV genotypes and subgenotypes 92
6.8 Future studies 96
6.0 Discussion

Molecular biomarkers have been used to determine the genetic relatedness of organisms, both to identify and track modes and chains of transmission and to characterize the evolution of organisms in host populations. Advances from disciplines such as bioinformatics and molecular evolution are among the supplementary tools in the process of understanding of both the epidemiology of agents within host populations and the dynamics of genetic changes within the agent itself.

It is well known that a high degree of antigenic conservation exists among HAV strains from different geographic regions. This has ultimately resulted in identification of a single serotype and restriction on the use of seroimmunological approach for differentiation of HAV strains. An alternative approach that was evolved to distinguish globally available HAV isolates is based on the diversity of genomic sequences of HAV strains. (Robertson et al. 1992:1365-1377). The comparative nucleotide sequence analysis has proved to be a powerful tool for genotyping, identification of source of infection, epidemiological tracking and links of various strains and for examination of genetic divergence. This approach has been also found to be advantageous in epidemiologic investigations to overcome the problems related to the long incubation period of disease, variable ratios of clinical to subclinical infections in different regions, the sporadic as well as epidemic nature of disease transmission and frequently observed unclear links between persons involved in outbreaks. Identification of origin of the contamination and links between geographically separated outbreaks (Hutin et.al 1999:595-602; Sanchez et.al 2002:4148-4155) has accelerated the measures needed to be taken for prevention and control of disease.

Several studies have been carried out for genotyping of HAV. However, many of these have analyzed HAV strains from USA, Sweden, Japan and France-the countries with low endemicity of hepatitis A as compared to those from developing regions such as South America, India, North and Central Africa, known to have hyperendemic pattern of hepatitis A (Robertson et al. 1992:1365-1377; Costa Mattioli et al. 2003:3191-3201). Based on nucleotide sequence data it has been described that HAV genotypes have unique geographic distributions (Nainan et al.2006: 63-79). Genotype I has been reported to be most prevalent worldwide and that subgenotype IA is more common than IB. Due to common circulation of these subgenotypes identification of source of infection or chain of transmission only by genotyping was difficult. Subgenotypes IA and IB have been often found in North and South America, Europe, China and Japan. Besides these subgenotypes, majority of the remaining human HAV strains have been placed in genotype III that is further classified into subgenotype IIIA and IIIB. Subgenotype IIIB
was responsible for some cases of HAV infection in Denmark and Japan (Mitsui et al. 2006:1015-1024; Robertson et al. 1992:1365-1372). PA21, the first IIIA isolate recovered from feral owl monkey in 1980 was the only strain recovered from nonhuman primate. All other isolates of IIIA were recovered from humans. Genotype IIIA was reported from India, Sri Lanka, Nepal, Malaysia and USA and has re-emerged in Europe and is becoming more prevalent than formerly assumed (Jansen et al. 1990:2861-2871; Robertson et al. 1992:1365-1377; Khanna et al. 1992:118-124; Costa-Mattioli et al. 2003: 3191-3201; Stene-Johansen et al. 2005:2739-2745). Recently, outbreaks caused by HAV genotype IIIA strains have been reported in Estonia, Italy, the Netherlands, Norway and the UK (Stene-Johansen et al. 1999:3725-3729; O'Donovan et al. 2001: 469-473; Tallo et al. 2003:187-193; Spada et al. 2005:956-964; Tjon et al. 2005:360-366).

Even though, the characterization of full-length genome has been done for some representative strains, a large number of isolates has been characterized by using short genome fragments from C terminus of the VP3 region, N terminus of the VP1 region, VP1/2A junction region, entire VP1 and VP1-2B region (Jansen et al. 1990:2867-2871; Robertson et al. 1992:1365-1377; Costa Mattioli et al. 2001:233-240; Costa Mattioli et al. 2002:9516-9525; Nainan et al. 2005:957-963). However, full-length genome sequencing is essential, as possibilities of recombination events occurring in other genomic regions cannot be ruled out.

In the past two decades, the data on entire/nearly entire nucleotide sequences of a total of 33 human and 1 simian HAV strains have been accumulated in GenBank/EMBL/DDBJ database. Thirty-three human HAV strains include 16 strains of IA, 8 strains of IB, 5 strains of IIIA (including 3 Japanese strains which were made available in March 2007 in the GenBank), 2 strains of IIIB and one strain, each of IIA and IIB subgenotypes. Among simian genotypes (IV-VI), complete genome is known only for genotype V (strain AGM27). For the strains of each of the genotype IV and VI nucleotide sequences of only 2.5kb and 168bp are available respectively.

In the present study full length genomes of HAV strains recovered from three hepatitis A patients, two with self limiting disease and the one associated with GBS subtype AMAN were sequenced to characterize HAV genotype IIIA from India and to find out the viral genomic changes under different clinical condition.

Similar to other RNA viruses, HAV exists in vivo as distributions of closely related variants namely quasispecies (Sanchez et al. 2003: 34-42). The molecular basis of this genetic variability may be the high error rate of the viral RNA-dependent RNA polymerase and the absence of proofreading mechanisms. Quasispecies dynamics is characterized by the continuous generation of variant viral genomes, competition among
them and selection of the fittest mutant distributions in any given environment. To avoid
the possible selection of a single clone for sequencing after cloning of the PCR
products, we sequenced directly PCR products, which would reflect majority of DNA
population.

6.1 Genetic relatedness of HAV strains

Comparison of full-length Indian IIIA sequences with GenBank IIIA sequences
showed PNI in the range of 94 – 97 and PAAI above 99. Analysis carried out separately
in P1, P2 and P3 domains showed PNI and PAAI in the range of 93-98 and 98-100
respectively. PNI values were in the range of 97-99 and 96-100 for 5'NCR and 3'NCR
respectively. Overall, the highly conserved nature of HAV genome from different
geographic locations such as Norway, Germany, Japan and India was observed. The
highest number of nucleotide identity/variations observed between any two strains is not
reflected at highest number of amino acid identity/variations for most of the genomic
regions indicating that all nucleotide substitutions were not necessarily converted into
amino acid substitutions.

Comparison of Indian strains with the strains other than genotype III in the
coding region and at full genome level showed PNI in the range of 80-83. However,
PAAI values were in the range of 96-97 in P1 domain, 92-93 and 91-92 at P2 and P3
domain respectively among human strains, thus, indicating the negative selection of
amino acid substitutions among HAV strains and highly conserved nature of capsid
proteins.

6.2 Recombination in HAV

Cocirculation of multiple genotypes or subgenotypes has been reported in some
regions of the world. Cocirculation of subgenotypes IA and IB is reported from South
Africa, Brazil, Israel and United States (Taylor et al. 1997:273-279; Villar et al.
2004:1779-1787; Nainan et al. 2005:957-963; CDC unpublished data), and
subgenotypes IA and IIIA are reported from the Central Asian Republics of the former
Soviet Union (CDC unpublished data). Subgenotypes IA, IB, IIIA and IIIB are
cocirculating in Japan (Endo et al. 2007:116-127). Limited studies carried out to
characterize HAV strains from India have shown cocirculation of genotype IA, IB along
with predominance of genotype IIIA (Arankalle et al. 2006::760-769; Hussain et al.

RNA viruses take advantage of all known mechanisms of genetic variation
including mutation and genetic combination to ensure their survival. In case of HAV,
genetic recombination has been observed among strains in cell culture (Lemon et al. 1991:2056-2065; Beard et al. 2001:1414-1426; Gauss-Muller and Kusov 2002:2183-2192) and was assumed not to occur in nature. However, in progressing years, dual infections of a single individual with HAV strains of different subgenotypes were reported (de Paula et al. 2003:223-228; Chitambar et al. 2007:85-93; Coppola et al. 2007:e73-e77). The first recombinant strain 9F94 isolated from an infected patient was a recombinant of genotype IIB and IB with crossover point in VP1 capsid protein (Costa-Mattioli et al. 2003: 51-59). Similarly FG isolate (X83302) is also a recombinant of genotypes IA and IB with a crossover point within the 3D region (Morace et al. unpublished data, Endo et al. 2007:8-17). Recombination requires double infection of a single cell. Therefore, prerequisites of in vivo recombination are cocirculation of different genotypes in a given geographic area and infection of a susceptible subject with at least two sources of infection in a short period of time. However, Chitambar et al. (2007:185-193) have shown presence of two genotypes IB and IIIA in sewage samples. Such a source is sufficient to infect susceptible individual and chances of recombination event to occur are more. Complete genome sequence analysis has been considered to be essential in determining the frequency of intra and intertypic recombination of HAV and the newly emerging genetic and antigenic variants (Costa-Mattioli et al. 2003: 51-59). In the study presented here all three strains of HAV showed absence of genetic recombination (Figure 7, 8, 9).

6.3 Non-coding region

Higher PNI observed in 5′NCR as compared to full length genome or any other part of the genome of different genotypes clearly indicated the importance of this region for HAV. 5′NCR constitutes internal ribosomal entry site and pyrimidine rich tract. These are important for cap independent translation of the viral message and therefore, for viral replication. Nucleotide substitutions in the 5′NCR might result in significant functional changes in the virus. 5′NCR of all IIIA and IIIB strains including those of the present study showed 18 common deletions along with one addition. Fifteen of these are in the first pyrimidine rich tract at 128-142 nucleotide positions. Shaffer et al. (1994:5568-5578) have shown that the PY1 domain (99-138) forms an ordered structural element downstream of the putative 5′ pseudoknots of HAV, which is not required for efficient replication in cultured cells. In contrast, an extended single stranded region immediately downstream of the PY1 domain, which includes nt 140-144 is required for efficient replication of the virus at physiologic temperatures (Shaffer et al. 1994:5568-5578). 5′ NCR of HAV interacts with cell specific proteins especially at the pyrimidine rich tract. The cellular specificity of HAV replication is probably due to the
presence of such cell specific nuclear factors (Jia et al.1996:2861-2868). Graff et al. (1997:1841-1849) also demonstrated the importance of mutations in 5'NCR that are responsible for host cell dependent cell culture adaptation. All these findings may suggest that the viral host range is determined in part by differences in the sequences of 5'NCR. Repeated efforts using different cell lines, which included BGM, VeroE6, and BSC-1 adaptation of IEM and ELISA positive strains of IIIA of the present study was not possible. Globally genotype IIIA is underreported on the aspects of tissue culture isolation and molecular characteristics. PA21 recovered from feral owl monkey, is the only IIIA strain reported to be isolated in tissue culture. Some wild type strains have not been adapted to cell culture despite intensive efforts. Nucleotide deletions, additions or substitutions in 5'NCR may play major role in tissue culture adaptation.

6.4 Coding region

6.4.1 Analysis of synonymous and nonsynonymous mutations

For evaluating evolutionary significance of variable protein coding sequences among diverged species in a quantitative fashion, one of the powerful tools is to compute nonsynonymous and synonymous substitution rates termed as Ka and Ks respectively (Kimura 1983:367; Gillespie 1991:336; Li 1997:463-464). Since Ka and Ks represent the numbers of substitutions per nonsynonymous and synonymous site, respectively, these parameters are used to partition the targeted sequences into three basic scenarios namely negative (purifying) selection, positive (adaptive) selection and neutral mutation.

The first study to estimate Ks and Ka values was reported by Sanchez et al for examination of the mutation rate in capsid proteins of HAV strains associated with outbreak caused due to consumption of coquina clams. (Sanchez et al. 2003:452-259). The Ks values were 7.5 times lower in the VP0 (VP4+ VP2) and VP3 capsid proteins and 5 times lower in the VP1-2A region for the outbreak stains than that of GenBank sequences while the Ka values were 13.4, 3 and 16 times lower respectively. The Ks/Ka ratios for the outbreak sequences of HAV strains were observed to be extremely high for the VP0 (VP4+VP2) and VP1-2A region and considerably high for the VP3 indicating divergence in HAV strains.

The complete ORFs of two Indian strains (CP-IND and PN-IND) and GenBank genotype IIIA strains (n=5) were analyzed in a similar manner by using synonymous and nonsynonymous mutations. Ks values for both GenBank and two Indian strains were apparently similar. However, the Ks/Ka ratios for Indian strains (except GBS-IND) were markedly lower than that of GenBank strains indicating higher number of nonsynonymous changes within the former (Table 7). This observation is important as it
suggests higher genomic variability within the strains from endemic region than that of low endemic region.

6.5 Phylogenetic analysis

Ching et al (2002:53-60) have illustrated phylogenetic trees, which were constructed by using all nucleotide positions, only nonsynonymous positions and only synonymous positions of P1 domain. Nonsynonymous and synonymous positions were evaluated independently to determine the differential evolution reflecting mutations resulting from the presence and absence of selective pressure. Phylogenetic trees using all nucleotide positions and only synonymous positions showed clustering of human genotype I, II and simian genotype V strains of HAV on one branch and that of human genotype III and simian genotype IV on another branch (Ching et al. 2002:53-60). It was also described in the same study that the two simian strains on separate branches might suggest separate evolutionary pathways that do not reflect host dependent coevolution. However, tree based on only nonsynonymous positions contains a branch that separates simian strains from human strains. We carried out similar analysis by using all nucleotide positions, complete ORF and more number of strains (n=28 in place of 11 described by Ching et al. 2002) to clarify the discrepancy. It is of interest to note that the simian strains in the synonymous tree appeared on the separate branches with a little better support (62%) as against sub branching of simian strains on human branch with 41% support indicated in earlier study. In addition to this, both the trees based on all nucleotide positions and only nonsynonymous positions showed separate branches for simian strains with 70% and 97% bootstrap support respectively (Figure 10 & 11). It is worth noting that the clustering of human strains according to genotype and subgenotype was observed with bootstrap support of 100% for nonsynonymous and all nucleotide positions and 62-100% for synonymous positions. These data suggest host specific evolution among human HAV strains and may reflect selective pressure resulting in mutations and changes for specific virus host interactions, such as that with the immunodominant neutralization epitope (Ping et al. 1988: 8281-8285; Nainan et al. 1992:984-987).

6.6 Analysis of GBS-IND strain

6.6.1 Clinical status and molecular changes

Hepatitis A usually occurs as a self-limiting infection of the liver. However, variable spectrum of disease ranging from subclinical to severe form also has been reported. Complications of hepatitis A that have been described include persistent
infection, fulminant hepatic failure, renal failure, late recovery, clinical relapse, cholestatic hepatitis and chronic hepatitis (Saunders et al. 1979: 569-584; McDonald et al. 1989: 223-228, Chio and Bakir 1992: 413-416; Glikson et al. 1992: 14-23; Schiff 1992: 518-520, Inoue et al. 1996: 322-324; Bendre et al. 1999: 1107-1112). In addition the neurological syndromes such as myelitis, peripheral neuropathy, myeloradiculopathy, mononeuritis and exacerbation of multiple sclerosis have been described in association with hepatitis A (Adams and Asbury 1980: 2097-3030; Owen et al. 1980: 2307-2309; Bosch et al. 1986: 685-687, Pelletier et al. 1985: 53-56). GBS is a rare disorder of the nervous system that causes rapidly progressing muscle weakness that, in severe cases can result in paralysis. The most common infections that have been reported to precede the GBS were cytomegalovirus, herpes, Epstein-Barr virus and viral hepatitis. It is not clear why some patients develop GBS and others don’t. The patient investigated in the present study was classified as a case of GBS subtype AMAN on the basis of physical, CNS and electrodiagnostic examinations. Analysis of host antibody response indicated strikingly high anti-HAV IgG level, while genomic analysis of the causative agent detected genotype IIIA (Chitambar et al. 2006: 1011-1014) circulation of which has been commonly observed in India (Hussain et al. 2005: 16-24; Chitambar et al. 2007: 85-93).

It is known that genomic variations in HAV strains contribute significantly in replication/adaptation of virus in in vitro and in vivo systems. Wild type HAV grows very slowly in cell culture with no apparent cytotoxicity. However, after passaging it grows faster. Some strains grow efficiently in some cell culture systems but not in others (Day et al. 1992: 6553-6540). Some wild type strains could not be adapted to tissue culture with continuous efforts. On the other hand, cytotoxic HAV variants were detected in cell culture (Beneduce et al. 1995: 299-309; Zhang et al. 1995: 686-697; Brack et al. 1998: 3370-3376). Usually adaptation of HAV to tissue culture leads to loss of virulence in humans and other primates due to genomic mutations (Karrcn et al. 1988: 338-345, Mao et al. 1989: 621-624) and suggesting that viral factors may be involved in determining the severity of the disease. Overall, the data on HAV mutations obtained from in vitro and in vivo studies on attenuated and cytopathic viruses (Karrcn et al. 1988: 338-345; Tedeschi et al. 1993: 16-22; Brack et al. 1998: 3370-3376) have suggested that the mutations in HAV themselves may also have a potential to alter clinical status of the disease (Yokosuka 2000: 91-97). Though the severity of hepatitis A could be attributed to the host factors (Stapleton et al. 1995: S9-S14, Sata et al. 1996: 812-817), the variant virus may need further examination for the possible potential virulence. Single report on familial cluster of fulminant hepatitis A infection suggested that some strains of HAV might be particularly pathogenic (Durst et al. 2001: 453-454). Thus, even though HAV has been studied extensively, correlation of the molecular
characteristics of the virus and the clinical status of disease has not been established so far.

We analyzed GBS-IND strain of the present study by using Ks/Ka ratio in the light of sequence data from other Indian strains and GenBank strains. Interestingly, capsid proteins VP1, VP2 and VP3 showed 2 to 5 times lower Ks/Ka ratio in GBS-IND strain than that of other Indian strains when both were compared against GenBank strains thereby indicating presence of higher number of nonsynonymous changes in the former (Table 8). It is known that VP1 and VP3 are major immunodominant proteins and hence high number of nonsynonymous changes observed in GBS-IND strain is of significance.

Previously VP1/2A junction region, 5'NCR, P2, P3 and full length genomes were analyzed in search of genomic changes that may lead to severity of the disease. Very recently molecular changes in VP1/2A junction region and 5'NCR have been shown to be useful in differentiation of acute and severe clinical presentation of HAV (Munne et al. 2007). However, Fujiwara et al. have reported failure of VP1/2A junction region in differentiation of clinical status of disease (Fujiwara et al. 2003:124-134). In the study presented here no difference within amino acid sequences of GBS-IND and other IIIA strains was observed. It has been reported that nucleotide variation in central portion of 5'NCR might influence the outcome of disease in case of FHF patients (Fujiwara et al. 2002:82-88). In the present study, GBS-IND strain also showed 2 unique nucleotide substitutions in 5'NCR, located in second pyrimidine rich tract and near the end of intra ribosomal entry site region. However, engineering the suitable constructs and examining their effects by in vivo studies may explain the role of such substitutions in clinical presentation of disease. Fujiwara et al. showed more mutations in the HAV strains obtained from FHF and severe acute hepatitis patients than those of acute hepatitis patients in the central part of 2B region (Fujiwara et al. 2007;560-566). The same group of authors reported few however, nonspecific amino acid substitutions in 2C protein of FHF patients as compared to acute/severe acute hepatitis A patients indicating the association between severity of hepatitis A and amino acid variations in 2C region of HAV (Fujiwara et al 2007; 871-877). It is suggested that the 2C protein may have a central role in replication of most of HAV strains in cell culture as well as in virulence in animal models (Raychoudhari et al 1998:7467-7475). Comparisons between chimeras of human and simian strains have suggested that sequences within the central region known to be highly conserved may encode important structures necessary for 2C function (Raychoudhari et al 1998:7467-7475). In the light of these data it is worth noting that GBS-IND strain of the present study showed a unique, heterologous amino acid substitution at position 251 in conserved region of 2C region. It
may be noted that mutations in conserved residues of the poliovirus 2C protein inhibited virus replication resulting in low yield of virus (Mirzayan and Wimmer 1992:547-565; Teterina et al 1992:1977-1986). Such a correlation could not be extrapolated to GBS-IND strain as it was found to persist for long time in the patient. However, it may suggest its association with severity of the disease. It has been described that amino acid substitutions in 3D region of SLF88 strain probably altered the polymerase efficiency resulting in increased replication and more aggressive disease (Ching et al. 2002:53-60). Genome of GBS-IND strain showed no mutations in 3D region inspite of its presence in the patient for several days.

6.6.2 Antigenic variability

The dominant antigenic determinants on the HAV capsid consist of neutralizing site, which is conformational in nature and distributed over regions of VP1 and VP3 on each capsomer (Luo et al.1988:503-514; Lemon et al. 1991:20-23). Antigenic variants of HAV have been identified on the basis of reactivity of monoclonal antibodies with neutralization epitopes present on VP1 and VP3 capsid proteins. So far described variants include cell culture derived mutants and strains from monkey, Uruguay and shellfish borne outbreak from Spain. Neutralization escape mutations for the human HAV strain HM-175 were identified at 70 and 74 amino acid position of the VP3 protein and 102, 171 and 221 of VP1 protein (Ping et al. 1988:8281-8285; Ping and Lemon 1992:2208-2216) and those for HAS15 strain were identified at 65, 70 and 71 of VP3 protein and 104, 105 and 232 of VP1 protein (Nainan et al. 1992:984-987). Neutralization escape mutants have been reported among simian strains in VP1 at 102, 105, 174, 178 and 221 amino acid position and those for VP3 protein at 65 and 70 amino acid position (Nainan et al.1991:1685-1689; Tsarev et al. 1991:1677-1683). Uru-3 strain has been reported to have deletions of 15 amino acids in the VP1 region. This region contained 3 amino acids (Ser 102, Asn 104 and Asn105), which were reported to be able to induce an escape response in neutralization experiments (Costa-Mattioli et al. 2002:9516-9525). Since, these residues align with recognized neutralization immunogenic sites in human rhinovirus (HRV) 14 and poliovirus (PV) type 3, it is suggested that deletion found in Uru-3 strain would alter antigenic structure of this virus and that this may be the first antigenic variant found in humans. Two other antigenic variants were detected from shellfish borne outbreak, one in a discontinuous epitope defined by monoclonal antibody K3-4C8 and a second in a linear VP1 epitope of the virus.

On the basis of nuclotide sequencing and deduced amino acid sequences, Sanchez et al. (2003:452-459) found amino acid residues variable at 1 position in VP4
(residue no. 5), four positions in VP2 (residue no. 40, 44, 89, 180), six positions in VP3 (residue no. 32, 45, 72, 92, 145, 239) and 9 positions in VP1 (residue no. 25, 28, 148, 156, 174, 208, 216, 241, 271) as variable amino acids in the GenBank and outbreak sequences. It was assumed that all these substitutions were located at the capsid surface and hence most of these were expected at HAV antigenic sites. However, only three of these substitutions from VP1 and VP3 proteins have been found directly involved in HAV epitopes while few more residues from the same proteins were located close to the antigenic sites. In case of GBS-IND strain of the present study deduced amino acid sequence of VP1, VP3 and VP4 capsid proteins showed heterologous and unique amino acid substitutions in B cell epitope region and/or closed to epitope region (Figure 15, 17, 18).

It is known that synonymous single nucleotide polymorphisms do not produce altered coding sequences, and therefore, they are not expected to change the function of the protein in which they occur. However recently, conformational changes in a protein caused due to synonymous nucleotide changes leading to codon replacements have been reported (Kimchi-Sarfaty et al. 2007: 525-528). When frequent codon is changed to infrequent codon, rate of protein translation might be affected. This in turn might affect co-translational protein folding leading to altered specificity of the protein. Since synonymous mutations are the most prevalent mutations in HAV, these are needed to be checked for changes in codons and also for their role in disease severity. Codon usage studies in comparison to other HAV strains may be of singnificance in case of GBS-IND strain to find out the association if any with neuropathy.

GBS-IND strain showed absence of HAV antigen in blocking ELISA that used polyclonal anti HAV antibodies for capturing and probing. This may be due to viral load that was below the detection limit of ELISA. It has been shown that high mutation rate among RNA viruses create a cloud of potentially beneficial mutations at the population level, which afford the viral quasispecies a greater probability to evolve and adapt to new environments and challenges during infection. Selection of variant occurs at population level rather than on individual variants (Vignuzzi et al. 2006: 344-348). In case of HAV also quasispecies distribution and selection of fittest variant has been shown to occur (Sanchez et al. 2003: 34-42). It is possible that polyclonal anti HAV antibodies detected in GBS patient’s serum were against population of HAV variants and the mutant that was selected is neutralization resistant for these antibodies. This may explain the reason behind the long duration of persistence of HAV RNA inspite of high anti HAV titres. In this context it is essential to examine the epitopes on capsid proteins of GBS-IND strain by use of monoclonal antibodies. However, this remains to
be explored due to low load of virus in the fecal specimen (Table 1) and failure to isolate the GBS-IND strain in cell culture inspite of extensive efforts.

6.6.3 Estimation of surface charge distribution on capsid proteins

In order to find out the effect of unique and heterologous amino acid substitutions on capsid proteins of GBS-IND strain, homology-modeling approach was used. It is known that the patterns, which are apparently hidden at sequence level, become evident when mapped onto structure. In the absence of experimentally derived 3D structure data on HAV, attempt was made to predict 3D structure of individual VP1, VP2 and VP3 capsid proteins of HAV strains of the present study by using knowledge-based homology modeling approach. Modeling was done by using Modeller 8 V2 (Sali et al. 1993:779-815). VP1 is the major surface accessible protein in the mature virion of picornaviruses (Hogle et al. 1985:1358-1365; Mateu et al. 1995:298-306; Ping et al. 1988:8281-8285; Ping et al. 1992:2208-2216). Among picornaviruses, poliovirus showed the highest resolution and resulted in models that were stereo-chemically more favorable in VP1, VP2 and VP3 capsid proteins, when models obtained from respective structures of Mengovirus, Human Rhinovirus and Foot and Mouth Disease virus were compared. Overall, structural conservation was observed between the model and respective template. Difference in surface charge was clearly evident only in VP1 protein (and not in VP2 and VP3 proteins) of GBS-IND strain, when compared with that of the respective proteins of other HAV strains (Figure 20). The VP1-29R/K residue observed in all global strains of HAV being polar, positively charged and basic was seen to have resulted in a localized positive charge distribution on the VP1 surface. However, the presence of E at position 29 in the GBS-IND strain resulted in the negative and acidic charge distribution (Figure 20). Such a charge difference might change the receptor specificity or binding affinity of HAV. However, this assumption needs to be supported by experimental evidence from in vivo and/or in vitro studies. In influenza virus, single amino acid change in the hemagglutinin, the receptor binding protein, has been reported to alter receptor-binding specificity (Rojers et al. 1983:76-78). Similar observation was made for mouse neurovirulence caused by poliovirus type 2 Lansing due to 3 amino acid changes at position 29, 30 and 31 of VP1 protein (Monica et al. 1986:515-525). Amino acid substitutions resulting possibly in negative charge in the capsid caused binding of poliovirus type 2 Lansing to a different cellular receptor in mouse central nervous system and subsequently in paralysis of mice (Monica et al. 1986:515-525).

As reported previously detection of plasma HAV RNA after 20 days of illness may predict a prolonged course of the disease (Sagnelli et al. 2003:1-6; No authors
listed, Russian article in ZH Mikrobiol Epidemiol Immunobiol, 2007:20-23). In GBS patient HAV-RNA persisted up to 126 days despite remarkably high anti-HAV titres (Chitambar et al. 2006:1011-1014). Direct detection of HAV antigen in blood has been difficult (Hollinger et al. 1975:1464-1466) because fibronectin can bind to HAV and mask antigenic determinants required for immunological detection (Seelig et al. 1984:335-347). HAV capsid polypeptides and viral RNA have been detected in IgM circulating immune complexes isolated from experimentally infected chimpanzees (Margolis and Nainan 1990:31-37). In addition, detection of HAV RNA by PCR has been correlated with the presence of infectious virus (Bhattacharya et al. 2003:181-187). Therefore, long duration of HAV-RNA in GBS patient was probably indicative of neutralization escape or low level of antibodies that could neutralize the virus. Neutralizing epitopes of HAV are known to be complex and conformational in nature and distributed over regions of VP1 and VP3. It may be noted that in the present study heterologous amino acid changes seen in the GBS-IND strain were located in defined B cell epitope region of VP1, VP3 and VP4 capsid proteins (Figure 15, 17, 18). It has also been reported that neutralization resistance in Foot and Mouth Disease could be generated due to amino acid substitutions that occurred outside the neutralization epitope present on the surface of FMD virus (Parry et al. 1990:569-572). However, such a mechanism of neutralization escape cannot be documented in the absence of definitive crystallographic studies on HAV.

Recently medium resolution images of HAV particle have been obtained by cryo-electron microscopy (R.H. Cheng, unpublished data). Even though the work is in progress, the new view of the HAV particle suggests significant differences in its structure compared with other picornaviruses (Martin and Lemon 2006:s164-s171). The availability of this 3D structure would throw more light on the specific antigenic sites and conformational epitopes of HAV.

The GBS patients with preceding CMV, EBV, and Campylobacter jejuni infections have been described to be seropositive for anti ganglioside antibodies and to suffer from severe axonal degeneration (Hughes et al. 1999:74-97; Hadden et al. 2001:758-765; Yuki 2001:29-37). Molecular mimicry between gangliosides and epitopes of infecting pathogen and/or epitopes expressed by the infected cells has been assumed to cause generation of anti ganglioside antibodies (Shimoya et al. 1997:106-110; Ang et al. 2000:1453-1458; Koga et al. 2006:547-555). However, the proof for such an assumption would need the examination of epitopes of GBS-IND strain and anti ganglioside antibodies in the GBS patient of the present study. Finally, the construction of chimeric full length cDNA clones carrying the sequences specific for the capsid protein of GBS-IND strain followed by cell culture experiments and monoclonal antibody
studies would be useful for detection of viral risk factors and the mechanism involved in the development of GBS.

6.7 HAV genotypes and subgenotypes

Genomic characterization of HAV has been mainly carried out by genotyping of strains from different geographic regions of the world. Most commonly used VP1/2A junction region classifies globally available strains in to seven genotypes on the basis of >15% nucleotide variation between isolates and subgenotypes with >7.5% to <15% nucleotide variation (Robertson et al. 1992:1365-1372). Full-length genomes of three Indian genotype IIIA strains along with GenBank HAV strains (n=25) of all genotypes (IA, IB, IIA, IIB, IIIA, IIIB, V) were used to determine the percent nucleotide identities in different genomic regions of HAV. Since full genome sequence for genotype IV and VI are not available as yet partial nucleotide sequences of these genotypes were included for analyses to represent the genotypes. During analyses of these nucleotide sequences, most of the genomic regions satisfied the criteria of PNI values for genotyping and subgentyping except 5’NCR, 3’NCR and VP4 regions which showed overlapping PNI ranges (Table 12).

Comparative studies on nucleotide sequences have reported 5’NCR as the most conserved region among I, II and III HAV genotypes (Brown et al. 1991:5828-5838; Cohen et al. 1987:50-59; Bradly et al. 1984:373-386). Similar observations were made in the present study with additional genotypes and strains suggesting the importance of 5’NCR for HAV and also its utility in diagnostics. VP4 region along with 5’NCR due to their conserved nature as indicated by overlapping PNI ranges between genotypes and subgenotypes can also be recommended in diagnostic PCR. Although PNI of 3’NCR among HAV strains was equivalent to that of 5’NCR, due to its small size (60-64nt in length) and location at extreme end it has limited utility in molecular diagnostics.

The strains of all genotypes and subgenotypes were analyzed to find out the conserved amino acid stretches through out the genome. One well-conserved area was observed in the 2C region from 2C-108 to 2C-261 among both human and simian isolates. RNA synthesis is the main function of 2C region and long highly conserved stretch is indicative of its importance. Among human strains well-conserved long stretches were observed in capsid proteins VP1, VP2 and VP3. This is consistent with existence of single serotype of HAV. It is worth noting that unique amino acid substitutions observed in GBS-IND strain at 2C-251 position, VP1-91 position and VP3-50 position lie within these conserved stretches. It is possible that such substitutions may alter the functions of proteins.
6.7.1 Evaluation of different genomic regions for genotyping of HAV

The most commonly used genomic region to define HAV genotypes is the 168bp VP1/2A junction region, which classified globally available strains in to seven genotypes. The two strains classified as genotype II (CF-53 strain) and genotype VII (SLF88 strain) using VP1/2A junction region were reclassified as subgenotype IIA and IIB after full length genome sequencing resulting into only three human (I-III) and three simian (IV-VI) genotypes (Ching et al. 2002:53-60; Lu et al. 2004:2943-2952). VP1/2A junction region, however, could not differentiate the three antigenic variants of HAV reported recently (Costa Mattioli et al. 2002:9516-9525; Sanchez et al. 2002:4148-4155). While classifying DL3 strain of China Guo-Dong et al suggested that it may not be so accurate to define a genotype only by comparison of VP1/2A junction region (Guo-Dong et al. 2003:499-504). The analysis of complete VP1 gene sequence identified the emergence of novel variant with a deletion of 15 amino acids in immunodominant region of VP1, which could not be detected using either VP1 amino terminus or VP1/2A junction region (Costa Mattioli et al. 2002:9516-9525). Other genomic regions used for genotyping include the C terminus of the VP3 region and the 390 bp region of VP1-2B region (Hutin et al. 1999:595-602; Nainan et al. 2005:957-963). Limitations of various genomic regions during genotyping of different global strains were observed. To overcome these problems a comparative analyses of 240 HAV isolates from a number of communities using the 350 nucleotide fragment of the VP1-2B and 900 nucleotide fragment from the entire VP1 region was carried out at CDC laboratory. This study documented concordance in all instances for genotype assignment (CDC unpublished data). Utility of VP1-2B region using 315bp fragment was also confirmed by using over 3,000 HAV isolates from the United States and 12 other countries.

As mentioned in previous section with exceptions of 5'NCR, VP4 and 3'NCR most of the other regions showed PNI values satisfying the criteria of distinguishing genotypes and subgenotypes. Further to confirm the reliability of various genomic regions for phylogeny, analysis was carried out by three different phylogenetic methods that included NJ, ML, MP methods. Complete VP1 gene was taken as a gold standard. Bootstrap support above 80% was considered as a good support.

5'NCR, VP4, 2B, 3A, 3B and 3'NCR were the regions, which did not show good bootstrap support and/or grouping similar to gold standard by all three phylogenetic methods (Table 14). Region 3C displayed clustering of genotypes similar to gold standard by all three methods and bootstrap support above 80%, however, HMH strain from Germany is the only strain which is off grouped from other genotype IIIA strains in ML method of PAUP software. MP method showed some variations in clustering of
strains of genotype IV and V for VP2, VP3 and VP1 regions, however it is not supported by good bootstrap support. Failure of the MP method may be due to higher number of strains used in the analyses than recommended (n=15) and inability of the method to use all sequence information (Table 16).

Table 16: Advantages and disadvantages of methods used for phylogenetic analysis

<table>
<thead>
<tr>
<th>Name of the method</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neighbor Joining (NJ) with Kimura 2 parameter</td>
<td>Distance based method</td>
<td>Sequence information is reduced, gives only one possible tree</td>
</tr>
<tr>
<td></td>
<td>Corrects for multiple hits</td>
<td>Strongly dependent on the model of evolution used.</td>
</tr>
<tr>
<td></td>
<td>Consider the difference between transition and transversion rates</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Assumes that the four-nucleotide frequencies are the same and that rates of substitution do not vary among sites.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fast and thus suited for large datasets and for bootstrap analysis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Usually finds a tree that is quite close to the optimal tree</td>
<td></td>
</tr>
<tr>
<td>Maximum Parsimony (MP)</td>
<td>Parsimony-Informative site are the sites that has at least two different characters, each represented at least twice</td>
<td>Only informative sites are analyzed. Other variable sites are not used for constructing an MP tree, although they are informative for distance and maximum-likelihood methods</td>
</tr>
<tr>
<td></td>
<td>It searches all possible tree topologies for the optimal (minimal) tree</td>
<td>There need not be a unique Maximum Parsimony tree</td>
</tr>
<tr>
<td></td>
<td>The possible optimal tree is built by adding the number of changes at each informative site for each tree.</td>
<td>Slow in comparison with distance methods</td>
</tr>
<tr>
<td></td>
<td>The tree that requires the least number of changes is chosen as the best tree</td>
<td>Does not correct for multiple mutations</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Give only the branching pattern and not the branch lengths</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fails to get best tree with large number of taxa (&gt;15)</td>
</tr>
<tr>
<td>Maximum Likelihood (ML)</td>
<td>The method is statistically well founded and is most widely accepted</td>
<td>It is not simple to apply</td>
</tr>
<tr>
<td></td>
<td>Can be applied to nucleotide or amino acid sequences</td>
<td>It is computation ally very intense (limits number of taxa and length of sequence)</td>
</tr>
<tr>
<td></td>
<td>It uses all the sequence information (unlike MP method)</td>
<td>The result is dependent on the model of evolution used.</td>
</tr>
<tr>
<td></td>
<td>Usually the most ‘consistent’ of the methods available [i.e. least affected by (bootstrap) sampling error]</td>
<td>Violations of the assumed model can lead to incorrect trees.</td>
</tr>
</tbody>
</table>
Irrespective of advantages and disadvantages of each of the three phylogenetic methods used for analysis (Table 16), VP3, 2A, 2C, and 3D were found to be the genomic regions, which clustered the strains according to their genotypes, and/or subgenotypes with high bootstrap support as defined by VP1 region that is included as gold standard in this study (Table 14).

6.7.2 Likelihood mapping analysis

The quartet topology weighting introduced by Strimmer et al. (1997:210-213) led to three values that add up to 1.0. Plotting these three posterior weights into a 3 dimensional coordinate system makes all points fall into a triangular surface, a so-called simplex. This lead enables a way to analyze phylogenetic information in datasets, likelihood mapping (Strimmer and von Haeseler 1997:6815-6819). Plotting all quartets in a likelihood-mapping diagram provides an overview of how many quartets cannot be resolved. The higher this percentage, the less suited a dataset is for phylogenetic analysis. In view of this, percentages of unresolved quartets and completely resolved quartets were estimated (Table 15). Genomic regions namely complete genome, 2C, 2A, VP3-2B, 3D, VP1, VP3, VP2, VP1-2B, 3C, 5\'NCR, VP1/2A junction region, VP1 amino terminus, 3A, 2B, VP3 C terminus, 3B. VP4 and 3'NCR showed percentage of unresolved quartets in increasing order. Out of these regions complete genome, 2C, 2A, 3D, VP1 and VP3 have classified the strains according to their genotypes and subgenotypes with high bootstrap support (Table 14). Of these complete genome sequence has shown lowest unresolved quartet value and hence most suitable for phylogenetic analysis. However, it is always not feasible to obtain the data for full genome. Alternatively, 2C appears to be suitable genomic region for genotyping or phylogenetic analysis, which has not been explored till today.

To summarize, the study presented here describes characterization of full-length genome sequences of HAV genotype IIA strains from Indian subcontinent for the first time. Till date, complete genome data on this genotype available in GenBank represent the strains only from low endemic regions. In endemic regions, inadequate hygienic and sanitary conditions enhance the dissemination of virus resulting in widespread infections. A high density of population also elevates perpetual cycle of exposure-infection-exposure of the virus. Therefore, the strains circulating in endemic regions may be very different from that of low endemic regions. The present study conducted for molecular characterization of full-length genome sequences of Indian strains of HAV helps to enrich the GenBank database by including the data from endemic region. Analysis of complete genomes available from different strains showed extensive
nucleotide variations, which were not reflected at amino acid level indicating well-conserved nature of HAV proteins. Higher number of nosynonymous changes observed in Indian strains as compared to GenBank strains suggests variability in the strains from endemic regions. Phylogenetic analyses carried out by using nosynonymous, synonymous and all nucleotide positions showed separate branching of human and simian strains and indicated host specific evolution of HAV strains. HAV strain associated with GBS showed higher number of amino acid substitutions as compared to each of the other genotype IIIA strains including Indian strains. Among these substitutions, seven were unique to GBS-IND strain as compared to sequence data of all HAV strains deposited in GenBank. Majority of these substitutions were heterologous and involved in or close to predicted epitope regions. Amino acid substitution at VP1-29 position in GBS-IND strain resulted in a localized change in the surface charge from positive and basic to negative and acidic as monitored by homology modeling approach and suggested a possibility of change in receptor binding specificity. Evaluation of different genomic regions of HAV using various phylogenetic methods confirmed the utility of 5′NCR and VP4 region in diagnostics due to their conserved nature and identified VP3, 2A, 2C and 3D as suitable genomic regions comparable to VP1 region used as a gold standard for genotyping. Likelihood mapping analysis indicated complete genome sequence as the most suitable choice for HAV genotyping that was followed by 2C region, which has not been explored till date.

6.8 Future studies

Molecular characteristics of HAV strains of genotype IIIA that have been identified in the present study emphasize the need for further characterization of HAV strains that cause sporadic infections and outbreaks of hepatitis A in India and other endemic regions. The data on GBS-IND strain indicates that there is a need to perform in vitro and in vivo studies using proteins expressed by cDNA chimeras constructed with unique substitutions and monoclonal antibodies for establishment of the relationship between clinical spectrum of disease and molecular characteristics of HAV strains. Codon usage studies will be of significance if there is a change in codon due to nucleotide changes which may be either synonymous or nonsynonymous which may affect epitope and protein structure and ultimately virus neutralization. Utility of 2C region of HAV genome needs to be confirmed for genotyping by using large size of samples, which would include all genotypes and subgenotypes from different geographical locations.