

***In silico* CD4+ T cell epitope mapping and HLA coverage analysis for proteins of human hepatitis E virus**

Bhawna Rathi, Rakesh Aggarwal, Sita Naik*

Biomedical Informatics Center, Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow, India

Submitted: 20 Nov. 2009; Revised: 3 Mar. 2010; Accepted: 21 Mar. 2010

Abstract

Hepatitis E virus (HEV) infection is a common cause of sporadic and epidemic viral hepatitis in developing countries, and is being increasingly recognized in many developed countries. The genome of HEV consists of three overlapping open reading frames that code for viral structural and non-structural proteins. Data on T helper cells epitopes in HEV proteins are limited. We therefore used NetMHCIIpan web server for *in silico* prediction of peptide sequence that may act as promiscuous helper T lymphocyte epitopes. This was followed by HLA coverage analysis using the population coverage tool of Immune Epitope Database to calculate the fraction of individuals in various populations who may be expected to respond to sets of such peptides based on the HLA genotypic frequencies. Three 9-mer peptide cores in the HEV capsid protein were predicted to be putative binders to several HLA-DR alleles. *In silico* data suggest that these three peptide cores may together be expected to act as T-cell epitopes in more than 80% of persons in 9 of the 12 populations studied. Each of these three peptide sequences was found to be fully conserved over 35 full length HEV genome sequences belonging to all four known HEV genotypes. It may be useful to validate the role of peptides containing these three core peptide sequences as T-cell epitopes in *in vitro* or *in vivo* experiments. If confirmed as T-cell epitopes in such studies, these peptides could be useful adjuvants to the HEV recombinant vaccines, currently under development either as additives or in a prime-boost approach.

Keywords: Hepatitis E virus, T-cell epitopes, major histocompatibility complex polymorphism, helper T cells.

INTRODUCTION

Infection with hepatitis E virus (HEV) is a major cause of acute viral hepatitis in several developing countries, in particular those in south and southeast Asia, northern Africa, the Middle East, etc (Aggarwal and Naik, 2009; Purcell *et al.*, 2008). In these regions, HEV infection causes large outbreaks of acute hepatitis (Naik *et al.*, 1992) as well as a large proportion of sporadic cases with this disease. The virus is transmitted predominantly by the fecal-oral route, usually through contaminated drinking water (Aggarwal and Naik, 2009; Purcell and Emerson, 2008). Person-to-person transmission appears to be uncommon (Aggarwal and Naik, 1994; Somani *et al.*, 2003). Characteristic epidemiological features of HEV infection include: a high disease attack rate among young adults, a relative

sparing of children, and particularly high attack rate and case fatality rate (15-20%) among pregnant women (Aggarwal and Naik, 2009; Purcell *et al.*, 2008). The infection is usually self-limiting; however, recently, chronic infection with HEV has been reported among organ transplant recipients who are receiving immunosuppressive drugs (Kamar *et al.*, 2008).

The genome of HEV consists of a 7.5-Kb long, single-stranded RNA of positive polarity, with three partially-overlapping open reading frames (ORF), and short non-coding regions at both 5' and 3' ends (Tam *et al.*, 1991). The ORF1, located at the 5' end of the genome, is nearly 5 Kb in length. It encodes a 1690-amino acid (aa) long polyprotein that contains motifs characteristic of viral nonstructural proteins such as RNA-dependent RNA polymerase, helicase, methyltransferase and a papain-like cysteine protease. The ORF2 is located at the 3' end of the viral genome and codes for a 660-aa long viral capsid protein. ORF3, the shortest of the three HEV ORFs, has an extensive overlap with ORF2 and a minor overlap with ORF1, and encodes a small immunogenic 122 to 123-aa phosphoprotein, which may play a role in virus-host interaction.

***Corresponding author:**

Sita Naik, MBBS, MD
Biomedical Informatics Center,
Sanjay Gandhi Postgraduate Institute of Medical Sciences,
Lucknow 226014, India
Email: sitanaik@gmail.com

Humoral immune responses against HEV have been studied in detail (Favorov *et al.*, 1992; Khudyakov *et al.*, 1993; Khudyakov *et al.*, 1999; Riddell *et al.*, 2000; Dawson *et al.*, 1992; Coursaget *et al.*, 1993; Shrestha *et al.*, 2007). These studies show a prominent antibody response directed against immunodominant antigenic epitopes contained in pORF2 and pORF3 (Favorov *et al.*, 1992; Khudyakov *et al.*, 1993; Khudyakov *et al.*, 1999; Riddell *et al.*, 2000; Dawson *et al.*, 1992; Coursaget *et al.*, 1993). These data have led to the use of HEV ORF2 and ORF3 proteins in diagnostic assays for HEV infection. IgM antibodies to these proteins appear during early phase of disease and last for 4-5 months (Kamar *et al.*, 2008). IgG antibodies follow thereafter; the duration of their persistence, though longer than that of IgM anti-HEV, however remains uncertain. Recombinant truncated HEV ORF2 protein has been tried as a vaccine and have shown good immunogenicity and protective efficacy against hepatitis E disease (Kaur *et al.*, 1992). B-cell epitopes have also been identified in HEV ORF1 protein (Shrestha *et al.*, 2007).

In contrast, data on cellular or T-cell immune responses during HEV infection are limited (Naik *et al.*, 2002; Pal *et al.*, 2005). These responses are believed to be particularly important for protection against viral infections, through their role in elimination of virus-infected cells and in providing help to B cells for antibody production. These responses can downregulate viral replication through a cytokine-mediated, non-cytolytic pathway, as has been shown to occur during infection with hepatotropic viruses, such as hepatitis B virus or hepatitis C virus (Ferrari *et al.*, 2003). We have previously shown existence of proliferative immune responses (Naik *et al.*, 2002; Srivastava *et al.*, 2007) among patients with acute hepatitis E. We have also mapped T-cell epitopes on HEV ORF2 protein (Aggarwal *et al.*, 2007). Other groups have also recently shown presence of cellular responses to HEV (Wu *et al.*, 2008; Shata *et al.*, 2007).

Activation of helper T lymphocytes requires recognition by these cells of viral peptides bound to the molecules of major histocompatibility complex (MHC) on antigen presenting cells (APC). The APCs take up viral proteins and process these into small peptides. These peptides bind to the resultant MHC molecules and the MHC-peptide complexes are transported to the APC surface. Host T-cells recognize and bind these complexes through the T-cell receptor molecules on their surface. The affinity of individual MHC molecules and peptides varies widely. MHC and peptide pairs with a high binding affinity to each other are associated with induction of a cellular immune response; in contrast, if the binding affinity is low, no immune response is generated.

Bioinformatic approaches allow for *in silico* prediction of affinity of a particular peptide sequence with various known HLA molecules. This may allow identification of specific peptide(s) that may be suitable for inclusion in subunit T-cell vaccines against a pathogen using the latter's genomic sequence data. We report here our data on *in silico* prediction of putative T-cell epitopes in HEV proteins that can bind with various known HLA-DR molecules on human CD4+ cells, and population coverage and epitope conservation analysis of these predicted T-cell epitopes.

MATERIALS AND METHODS

Retrieval of protein sequences

The protein sequences of pORF1, pORF2 and pORF3 proteins of a genotype 1 HEV strain of Burmese origin were downloaded from the Entrez database of the National Centre for Biotechnology Information, Bethesda, MD, USA (<http://www.ncbi.nlm.nih.gov>; sequence ID M73218). These peptides are 1690, 660 and 123 amino acids in length respectively.

CD4+ T-cell epitope prediction and HLA distribution analysis

Amino acid sequences of HEV proteins were uploaded to the NetMHCIIpan server (Nielsen *et al.*, 2008) (<http://www.cbs.dtu.dk/services/NetMHCIIpan/>) for prediction of linear 9-mer peptide core sequences in each of these proteins that could serve as potential T cell epitopes, and to identify the HLA-DR allele(s) that each such putative epitope could be expected to bind to. NetMHCIIpan server is a tool on the World Wide Web that can process amino acid sequence of a protein and identify short linear peptide sequences contained therein that can bind to one or more of the nearly 500 known human HLA-DR molecules and thus can serve as CD4+ T-cell epitope cores. The server uses an artificial neural network (ANN) approach. It has been trained on previously-available quantitative experimental laboratory data on binding of peptides from several proteins with various HLA-DR molecules. The computational method used by the server takes into account not only data on amino acid sequences of the epitope core and of the peptide-binding cleft of HLA-DR molecules, but also of amino acid residues that flank the peptide core and those in HLA-DR molecules that lie within interaction distance of the bound peptide. The server provides an output that shows 9-mer long potential HLA-DR restricted putative epitope cores, along with a set of HLA-DR alleles that each of these can be expected to bind to, and the binding affinity of each peptide and HLA-DR allele pair. After initial ANN training, the software can predict binding affinity

of novel peptide sequences to HLA-DR alleles included in the original dataset used to train the software, as well as to newly discovered HLA-DR alleles provided in the amino acid sequences are known.

Population coverage analysis

The data for predicted core peptides that could serve as T-cell epitopes and their ligand HLA-DR alleles were submitted to the 'Population Coverage Tool' of the Immuno Epitope Database (IEDB) web server (http://tools.immuneepitope.org/tools/population/iedb_input).

This tool calculates the fraction of individuals in a particular population that can be expected to respond to a particular putative epitope or epitope set, using as input the data on binding affinity of each peptide with various HLA-DR alleles (Bui *et al.*, 2006). This calculation also needs data on genotypic frequency of various HLA-DR alleles in a population; these data can either be provided as an external input, or used from pre-defined HLA-DR allele frequency data for several populations already available at the server website. We used this tool to calculate population coverage of HEV peptide core sets for populations in 11 geographical areas contained in the server database, as also for a northern Indian population using allelic frequency data based on a recent study (Agrawal *et al.*, 2008).

Epitope Conservancy Analysis

An effective vaccine should contain epitopes that are conserved across different genotypes of a pathogen. The putative peptide epitopes that were predicted to bind to a large number of HLA-DR molecules were therefore subjected to epitope conservancy analysis. This was done using the epitope conservancy analysis tool of the IEDB server http://tools.immuneepitope.org/tools/conservancy/iedb_input (Bui *et al.*, 2007). This tool computes degree of conservation of amino acid sequence within a predicted epitope, using a set of known protein sequences for various known isolates of a particular pathogen. For this analysis, pORF2 amino acid sequences from 35 isolates belonging to the four known HEV genotypes were used. A phylogenetic tree based on nucleotide sequences of the full length genomes was also drawn of these isolates using MEGA software (Tamura *et al.*, 2007).

RESULTS AND DISCUSSION

In silico prediction of CD4+ T-cell epitopes using NetMHCIIpan server identified 344 peptides (9-mer each) that could act as cores of putative T-cell epitopes in the ORF1 protein, 147 such peptides in the ORF2 protein and 26 such peptides in the ORF3 protein of genotype 1 HEV. Identification of the largest number of putative epitopes in the ORF1 proteins and the least

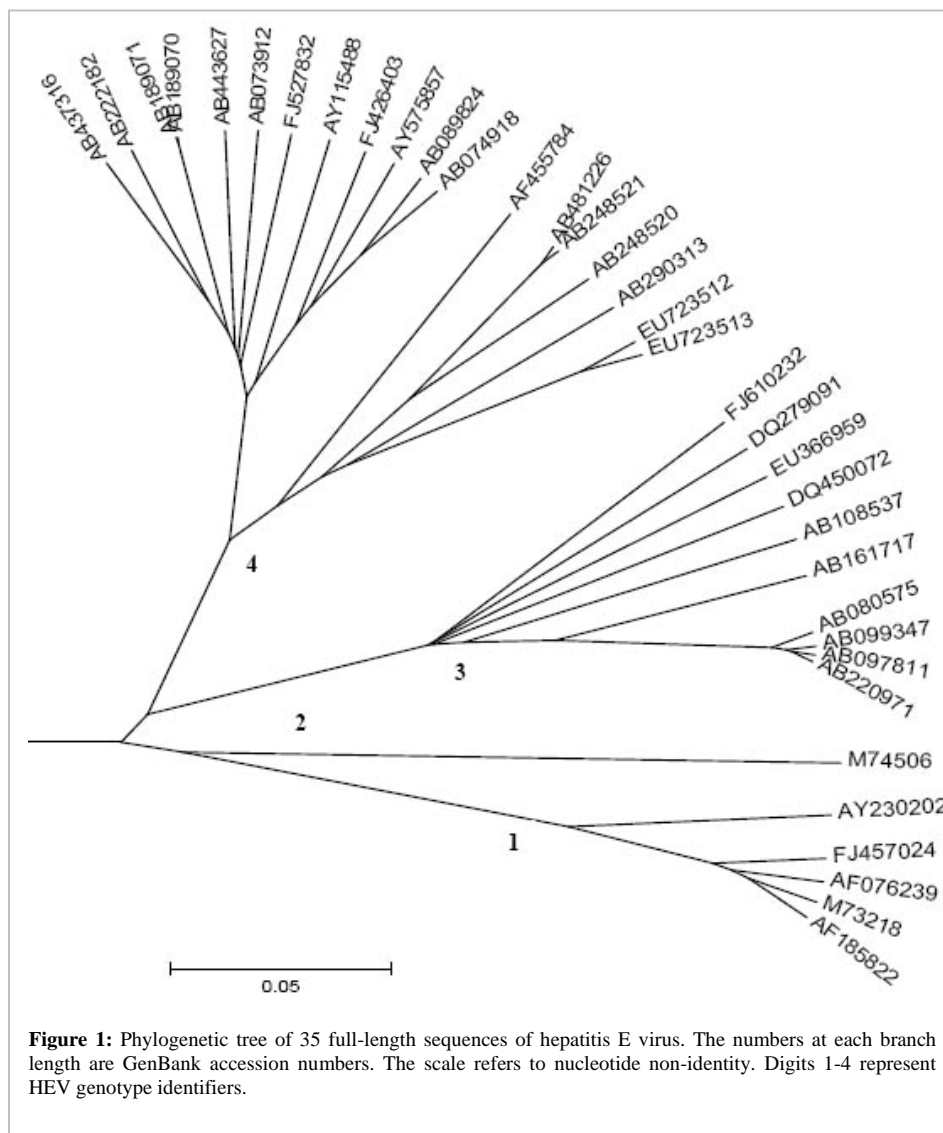
in ORF3 was as would be expected from the relative length of these proteins, i.e. 1690 aa for ORF1, 660 aa for ORF2 and 123 aa for ORF3. The details of putative CD4+ T-cell epitope cores in each of the HEV proteins and their respective binding HLA-DR alleles are available from authors. These tables include all peptide cores with binding affinities (IC50) of less than 500 nM, and include both weak (50-500 nM) and strong binders (<50 nM).

These data obtained on *in silico* prediction of T-cell epitopes of HEV along with their binding HLA-DR alleles were used as an input for calculation of population coverage rates. MHC loci are highly polymorphic and the HLA allelic frequencies vary considerably in different population groups. The population coverage rates obtained from this tool for each of the 11 geographical regions for which the allele frequency data were available at dbMHC as well as for the northern Indian population are shown in Table 1 [Supplementary data]. The coverage rates for all the putative epitopes contained within each protein were found to be lower in the South East Asian and European populations (Table 1).

ORF2 in HEV codes for the main viral structural protein, which acts as the viral capsid. Two different truncated recombinant ORF2 proteins have been used as vaccine candidates for HEV, because this protein has been shown to contain a B-cell neutralization epitope located between amino acids 458 and 607 (Meng *et al.*, 2001). In animal studies, recombinant HEV ORF2 proteins have been shown to elicit neutralizing antibodies against HEV and to induce protective immunity (Zhang *et al.*, 2002; Purdy *et al.*, 1993). In fact, one vaccine that includes a truncated HEV ORF2 protein (aa 112 to 607 of HEV ORF2 protein) has undergone a phase III clinical trial and was shown to be highly efficacious in protecting against clinical hepatitis E (Shrestha *et al.*, 2007).

In view of this, we focussed on ORF2 protein, and selected three predicted putative epitopes for pORF2 that showed binding to the largest number of HLA-DR alleles (Table 2 [Supplementary data]). The population coverage rates of these putative epitopes were calculated using the population coverage tool for each population (Table 3 [Supplementary data]). For most populations, the proportion of population covered by these three peptides taken together was only marginally less than that covered by all the putative HEV ORF2 epitopes (Tables 2 and 3). This suggests that a combination of these three peptides may be expected, based on *in silico* data, to activate T-cell responses in a fairly large proportion of people in most populations.

Epitope conservancy analysis was done to determine the conservation of these three putative epitopes in 35 HEV isolates belonging to all the four HEV genotypes.



A phylogenetic tree based on these isolates is shown in Fig.1. All the three putative epitopes, i.e. LRRQYNLST, YNLSTSPILT and FSVLRANDV were fully conserved among these 35 HEV isolates, indicating that these peptides are likely to induce T-cell responses equally effectively against all the four HEV genotypes.

Knowledge of T cell epitopes of a pathogen may be useful in several ways. First, this information may be useful in design of vaccines against a pathogen. Second, the peptides containing the T cell epitopes sequences can be used to study the induction and time course of cellular immune responses during and after an infection with the particular pathogen, and to study the role of such responses in disease pathogenesis. Finally, the knowledge may help develop specific diagnostics assays, based on cellular immune response, for exposure to an infectious agent.

Theoretically, it should be possible to develop a peptide-based vaccine, a relatively new method for prevention and control of infectious diseases, provided one or a few immunogenic peptides that act as T-cell epitopes can be identified in proteins contained in a pathogen. T-cell vaccines may also have the advantage of inducing T-cell memory, thus providing a longer duration of protection than vaccines that induce antibody responses (Robinson and Amara, 2005). In addition, these vaccines may be expected to be less likely to induce a response against self-antigens, thereby avoiding autoimmune responses (Ali and Francisco, 2008). The search for immunogenic peptides for use as vaccine candidates by the traditional *in vitro* and animal experiments is prolonged, cumbersome and expensive. However, *In silico* prediction of T-cell epitopes can markedly reduce the time and cost for development of such peptide-based vaccines.

Our data from the current work by predicting the T-cell peptide epitope cores for HEV may thus have application in development of a vaccine against this infection.

The peptides identified as putative T cell epitope cores will need verification in an *in vitro* or an *in vivo* system for T cell activation activity. If verified, short peptides containing these core sequences can be tried as additives to the recombinant HEV ORF2 protein based vaccines, at different stages of development. Alternatively, these peptides may be used in a prime-boost approach with the available recombinant proteins being used for initial priming and shorter peptides being used for boosting of immune response. The latter strategy may be particularly attractive for HEV vaccines since, the protective efficacy of recombinant protein based vaccine in primate was found to be short lasted. Use of short peptides containing T cell epitope cores may provide a cheaper booster vaccine for use

following the initial use of recombinant ORF2 protein based vaccine.

Distribution of HLA alleles varies widely among different racial and ethnic groups around the world (Imanishi *et al.*, 1991). Thus, a peptide which functions as a T-cell epitope in most members of one population may not be as effective in another population with a different HLA-DR allele distribution. The aim of vaccine design is to find one or a few immunogenic molecules that will be effective in inducing immune response in a large proportion (say >95%) of the population. Vaccines based on a single peptide thus may be limited by their inability to perform equally well in all individuals in a population due to variable extensive MHC polymorphism and peptide-MHC affinities. In fact, such vaccines may be ineffective in some populations, as was the case with the CS peptide based vaccine for malaria (Moriya and Fidal, 2001). Also, the use of a limited number of peptides may raise concerns about possible development of escape mutants. However, the combined use of a recombinant protein and immunogenic peptides, either simultaneously or as a prime-boost pair may reduce such risk.

Among the T-cell epitopes that we identified in HEV ORF2, one peptide (spanning amino acid 136-145) provided population coverage rate exceeding 90% for four ethnic groups and between 80% and 90% for four other ethnic groups among the 12 populations studied. Addition of two other peptides (aa 132-140 and 462-470) improved the coverage rates for most populations somewhat and provided population coverage exceeding 80% for 9 of the 12 populations. Thus, a vaccine based on these peptides should work well in these populations. However, these peptides still provided only 60-76% coverage for three populations, i.e. those in Europe, South-East Asia and Sub-Saharan Africa. Thus, a vaccine based on these peptides may have a limited application in these geographic areas. However, since the bulk of hepatitis E disease occurs in South Asia, the peptide-based vaccines based on T-cell epitopes predicted in the current work may have an application in this geographic region.

Attempts have been made to study cellular immune response to HEV using *in vitro* techniques. We and others have previously shown the presence of lymphocyte proliferation response to recombinant HEV ORF2 and ORF3 proteins and pools of peptides corresponding to these proteins (Aggarwal *et al.*, 2007). In another study, we mapped T-cell epitopes in HEV ORF2 protein using peptide pools (Aggarwal *et al.*, 2007). Two of the three peptides identified in the current work fall within the regions of ORF2 protein that showed reactivity in the previous epitope mapping study.

In silico data have their limitations, and these need validation in laboratory experiments. There is thus a need to validate putative T-cell epitope cores that have been predicted in this study; this can be done by demonstration of T-cell activation capability of peptides containing core sequences in experiments involving higher primates or humans. If these experiments confirm the *in silico* predictions made here, this knowledge may provide a useful lead for future immune studies on HEV infection and or for development of newer vaccination strategies for prophylaxis against hepatitis E.

Acknowledgement

Bhawna Rathi was supported by a fellowship from the Indian Council of Medical Research, New Delhi, India during this work (Grant number: P&I/BIC/1/1/2009). The authors thank Mr. AN Sarangi for helpful suggestions during this work. The Biomedical Informatics Center is supported by the Department of Biotechnology, Government of India and Indian Council of Medical Research.

References

- Aggarwal R and Naik SR (1994) Hepatitis E: interfamilial transmission versus waterborne spread. *J. Hepatol.*, **22**: 718-23.
- Aggarwal R and Naik S (2009) Epidemiology of hepatitis E: current status. *J. Gastroenterol. Hepatol.*, **24**: 1484-93.
- Agrawal S, Srivastava SK, *et al.* (2008) Genetic affinities of north and northeastern population of India: inference from HLA-based study. *Tissue Antigens*, **72**: 120-30.
- Aggarwal R, Shukla R, *et al.* (2007) T-cell epitope mapping of ORF2 and ORF3 proteins of human hepatitis E virus. *J. Viral Hepat.*, **14**: 283-92.
- Ali Azizi and Francisco Diaz-Mitoma. (2008) Viral peptide immunogens: current challenges and opportunities. *J. Peptide Sci.*, **13**: 776-86.
- Bui HH, Sidney J, *et al.* (2006) A Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics*, **7**: 153.
- Bui HH, Sidney J, *et al.* (2007) Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinformatics*, **8**: 361.
- Coursaget P, Buisson Y, *et al.* (1993) Mapping of linear B cell epitopes on open reading frame 2- and 3- encoded proteins of hepatitis E virus using synthetic peptides. *FEMS Microbiol. Lett.*, **109**: 251-5.
- Dawson GJ, Chau KH, *et al.* (1992) Solid-phase enzyme-linked immunosorbent assay for hepatitis E virus IgG and IgM antibodies utilizing recombinant antigen and synthetic peptides. *J. Virol. Methods*, **109**: 251-5.
- Favorov MO, Fields HA, *et al.* (1992) Serological identification of hepatitis E infection in epidemic and endemic settings. *J. Med. Virol.*, **36**: 246-50.

T-cell epitope mapping and HLA coverage in HEV

- Ferrari C, Missale G, *et al.* (2003) Immunopathogenesis of hepatitis B. *J. Hepatol.*, **39**: 67-73.
- Imanishi T, Akaza T, *et al.* (1991) Allele and haplotype frequencies for HLA and complement loci in various ethnic groups. In: HLA 1991. New York: Oxford University Press., 1065-1220.
- Kamar N, Selves J, *et al.* (2008) Hepatitis E virus and chronic hepatitis in organ-transplant recipients. *New Engl. J. Med.*, **358**: 811-7.
- Kaur M, Hyams KC, *et al.* (1992) Human linear B-cell epitopes encoded by the hepatitis E virus include determinants in the RNA-dependent RNA polymerase. *Proc. Natl. Acad. Sci. USA*, **89**: 3855-8.
- Khudyakov YE, Khudyakova NS, *et al.* (1993) Epitope mapping in proteins of hepatitis E virus. *Virology*, **194**: 89-96.
- Khudyakov YE, Lopareva EN, *et al.* (1999) Antigenic domains of the open reading frame 2-encoded protein of hepatitis E virus. *J. Clin. Microbiol.*, **37**: 2863-71.
- Meng J, Dai X, *et al.* (2001) Identification and characterization of the neutralization epitope(s) of the hepatitis E virus. *Virology*, **288**: 203-11.
- Moriya T and Fidal Z (2001) Peptide based sub-unit vaccine against pre-erythrocytic stages of malaria parasite. *Mol. Immunol.*, **38**: 433-42.
- Naik SR, Aggarwal R, *et al.* (1992) A large waterborne viral hepatitis E epidemic in Kanpur, *Ind. Bull. World Health Organ.*, **70**: 597-604.
- Naik S, Aggarwal R, *et al.* (2002) Evidence for activation of cellular immune response in patients with acute hepatitis E. *Indian J. Gastroenterol.*, **21**: 149-52.
- Nielsen M, Lundegaard C, *et al.* (2008) Quantitative predictions of peptide binding to any HLA-DR molecule of known sequence. *NetMHCIIpan. PLoS Comput. Biol.*, **24**: 4.
- Pal R, Aggarwal R, *et al.* (2005) Immunological alterations in pregnant women with acute hepatitis E. *J. Gastroenterol. Hepatol.*, **20**: 1094-101.
- Purcell RH and Emerson SU (2008) Hepatitis E: an emerging awareness of an old disease. *J. Hepatol.*, **48**: 494-503.
- Purdy MA, McCaustland KA, *et al.* (1993) Preliminary evidence that a trpE-HEV fusion protein protects cynomolgus macaques against challenge with wild-type hepatitis E virus (HEV). *J. Med. Virol.*, **41**: 90-94.
- Riddell MA, Li F, *et al.* (2000) Identification of immunodominant and conformational epitopes in the capsid protein of hepatitis by using monoclonal antibodies. *J. Virol.*, **74**: 8011-7.
- Robinson HL and Amara RR (2005) T cell vaccines for microbial infections. *Nature Med.*, **11**: 525-32.
- Shata MT, Barrett A, *et al.* (2007) Characterization of hepatitis E-specific cell-mediated immune response using IFN-gamma ELISPOT assay. *J. Immunol. Methods*, **328**: 152-61.
- Shrestha MP, Scott RM, *et al.* (2007) Safety and efficacy of a recombinant hepatitis E vaccine. *New Engl. J. Med.*, **356**: 895-903.
- Somani SK, Aggarwal R, *et al.* (2003) A serological study of intra-familial spread from patients with sporadic hepatitis E virus infection. *J. Viral Hepatitis.*, **10**: 446-9.
- Srivastava R, Aggarwal R, *et al.* (2007) Cellular immune responses in acute hepatitis E virus to the viral open reading frame 2 protein. *Viral Immunol.*, **20**: 56-65.
- Tam AW, Smith MM, *et al.* (1991) Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length viral genome. *Virology*, **185**: 120-31.
- Tamura K, Dudley J, *et al.* (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**: 1596-1599
- Wu T, Zhang J, *et al.* (2008) Specific cellular immune response in hepatitis E patients. *Intervirology*, **51**: 322-7.
- Zhang M, Emerson SU, *et al.* (2002) Recombinant vaccine against hepatitis E: duration of protective immunity in rhesus macaques. *Vaccine*, **20**: 3285-91.