Fakultät für Medizin

Institut für Virologie

Mechanism of immune escape from CD8+ T cell response in hepatitis delta virus infection

Hadi Karimzadeh

Vollständiger Abdruck der von der Fakultät für Medizin der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften
(Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender: Prof. Dr. Dirk Busch
Prüfer der Dissertation:

1. Prof. Dr. Ulrike Protzer
2. Prof. Dr. Dmitrij Frischmann

Die Dissertation wurde am 15.12.2016 bei der Technischen Universität München eingereicht und durch die Fakultät für Medizin am 17.05.2017 angenommen.

Table of Contents

Abstract

Hepatitis delta virus (HDV) is an RNA virus which causes delta hepatitis infection, the most severe form of chronic viral hepatitis in humans. HDV is a defective virus and needs hepatitis B virus (HBV), as a helper virus, for its life cycle. It is estimated that 15 to 20 million people are chronically infected with HDV worldwide. There are 8 different HDV genotypes of which genotype 1 is the most prevalent. To date, there is no efficient treatment or prophylactic vaccine against HDV infection. A minority of chronic carriers of HBV superinfected with HDV (about 10%) are able to resolve HDV infection. From vaccine studies in animal models, we learned that HDV-specific antibodies are not neutralizing; however, T cell response can control HDV infection. In humans, little is known about the HDV-specific CTL responses and respective epitopes. It is also not clear whether mutations within CTL epitopes may lead to viral evasion and adaptation to the host immune response. This may be the reason that the majority of patients are not able to recover from the disease. Therefore, the aims of this study were: characterization of HDV sequences in a large cohort of patients; evaluation of genetic variability of the only ORF of HDV encoding for the large hepatitis delta antigen (L-HD Ag); identification and characterization of MHC class I epitopes and CTL responses; identification of possible immune escape mutations within CD8+ epitopes in the L-HD Ag and their impact on T cell response; evaluation of HDV sequence variability in correlation with host HLA background as an approach to identify novel epitopes and possible immune escape mechanisms of these HDV epitopes.

A cohort of 142 patients with chronic HDV infection was recruited from 8 medical centers. The complete ORF of the L-HD Ag was amplified and sequenced. HDV genotyping was performed based on these sequences. HLA class I (Locus A and B) alleles were determined of 104 patients from this cohort. CTL epitopes were predicted for HDV-1 isolates using IEDB and SYFPEITHI databases. Predicted epitopes were synthesized and tested for binding affinity to the corresponding HLA alleles by UV-mediated exchange ligand MHC binding assay. Detection of T cell responses was performed on PBMCs isolated from patients with resolved HDV infection. Amino acid substitutions within the identified epitopes in the studied cohort were characterized, and the corresponding mutated peptides were synthesized. The impact of mutations on HDV-specific T cell response was evaluated by stimulation of T cells with ‘wild type’ and ‘mutant’ epitopes. Coincidence of molecular ‘footprints’ on L-HD Ag and particular HLA molecules was evaluated and applied for identification of T cell pressure and additional CTL epitopes.

Phylogenetic analysis of the sequences from this study indicated that all isolates belonged to the HDV genotype 1, the most prevalent HDV genotype worldwide. Longitudinal molecular

studies revealed a higher evolution rate during the early phase of infection than the late phase indicating virus adaptation to the host over time. Using CTL epitope prediction tools and MHC binding assay, we identified two HDV-specific epitopes restricted by HLA-B*27 allele, aa 99-108: RRDHRRRKAL and aa 103-112: RRRKALENKK which were also confirmed by stimulation of PBMCs and induction of interferon-γ secretion by HDV-specific CD8+ T cells. By sequence analysis, we found HLA-B*27-restricted immune selection pressure for two residues (R105K, K106M) within these two overlapping HDV epitopes in isolates from HLA-B*27 positive patients. Mutated epitopes impaired HDV-specific T cell recognition when compared to the 'wild type' epitopes, indicating that immune escape mutation occurs as an evasion mechanism in HDV infection.

By analysis of L-HD Ag sequence polymorphism in correlation with HLA alleles of patients, we detected a strong molecular footprint at the residue170 (S170N, $p=10^{-6}$) on L-HD Ag in the HLA-B*15 positive patients. In viral isolates of all 8 HLA-B*15 positive patients (100%), this substitution of serine (S) to asparagine (N) was observed, whereas only 11% (11/96) of HLA-B*15 negative group showed this substitution. This strongly indicated a possible immune escape mutation in L-HD Ag at this position. At this position, a novel HLA-B*15-restricted CD8+ T cell epitope was identified also by prediction (aa 170-179: SMQGVPESPF). This epitope was also confirmed by T cell assay in an HLA-B*15 positive patient who had resolved HDV infection. Further analysis of correlations between substitutions within all 214 residues of L-HD Ag and 35 HLA class I alleles (locus A and B) indicated a significant immune selection pressure restricted to HLA-B*13, -B*15, -B*37 and -B*41.

Conclusions: This is the first description of HLA-B*27 and HLA-B*15-restricted MHC class I HDV epitopes. Immune response to the identified HLA-B*27 and HLA-B*15 epitopes may contribute to final outcome of HDV infection. However, substitutions within these epitopes resulting in immune escape from CD8+ T cell response may cause evasion of HDV in HLA-B*15 and HLA-B*27 positive individuals. The high mutation rate after the onset of infection may be related to the virus-host interaction which leads to the adaptation to the new host's immune background and contributes to the evolution of circulating HDV isolates. Correlating mutations within the L-HD Ag with the HLA backgrounds seems to be a viable approach to identify new HDV-specific epitopes. HLA-B alleles seem to have a stronger selection pressure on HDV and, consequently, a greater impact on viral evolution than locus-A of HLA does. These results may be considered in HDV vaccine design studies.

Zusammenfassung

Das Hepatitis delta virus (HDV), ein RNA Virus, verursacht eine virale Hepatitis. Es ist die schwerste Verlaufsform einer viralen Hepatitis mit der höchsten Rate eines chronischen Verlaufs von über 80%. Circa 20 Millionen Menschen haben weltweit eine chronische Hepatitis D Virus Infektion. HDV ist ein defektes Virus und benötigt das Hepatitis B Virus (HBV) als Helfer-Virus um vollständige infektiöse Partikel zu generieren. Es werden acht verschiedene HDV Genotypen unterschieden. Genotyp 1 ist dabei weltweit dominant. Bis heute gibt es weder eine prophylaktische Impfung gegen eine HDV Infektion, noch eine effiziente Therapie der chronischen Verlaufsform. Nur etwa 10 % der infizierten Patienten sind in der Lage das Virus zu eliminieren. Aus verschiedenen Impfstudien in präklinischen Tiermodellen ist bekannt, dass die gebildeten Antikörper gegen das einzige HDV-Protein nicht neutralisierend sind und somit nicht zur Elimination des Virus beitragen. Aus Tiermodellen ist bekannt, dass eine gute T-Zell Antwort hingegen die Virusinfektion kontrollieren kann. Beim Menschen gibt es bisher wenig Kenntnisse über eine HDV spezifischen T-Zell Antwort und die dazugehörigen Epitope. Darüber hinaus ist nicht bekannt inwieweit Mutationen der T-Zell-Epitope durch Adaption des Virus an den Wirt zu einer Immunevasion führen. Eine solche Adaptation könnte der Grund sein, dass die Mehrzahl der Patienten nicht in der Lage sind das Virus zu eliminieren.

Die Ziele dieser Untersuchungen waren daher die Charakterisierung von HDV RNA Sequenzen in einer großen Kohorte von Patienten, die Evaluation der genetischen Variabilität des offenen Leserahmens (ORF) für das HDV-Protein (L-HD Ag), die Identifikation und Charakterisierung von MHC Klasse I Epitopen und CTL Antworten bei Patienten, Identifikation von möglichen Immunevasions-Mutationen in den CD8+ Epitopen des L-HD Ag und deren Bedeutung für die T-Zell Antwort, Evaluation der HDV Sequenz-Variabilität in Korrelation zum HLA Hintergrund des Wirtes als methodischen Ansatz, um bisher unbekannte Epitope zu detektieren und eine mögliche Immunevasion nachzuweisen.

Für diese Untersuchungen wurde eine Kohorte von 142 Patienten mit chronischer HDV-Infektion aus acht medizinischen Zentren rekrutiert. Das komplette ORF des L-HD Ag wurde amplifiziert und sequenziert. Eine HDV-Genotypisierung wurde auf der Grundlage dieser Sequenzen durchgeführt. HLA-Klasse I (Locus A und B) Allele wurden von 104 Patienten dieser Kohorte bestimmt. Mögliche CTL-Epitope wurden für HDV-1-Isolate mit den Datenbanken IEDB und SYFPEITHI vorausgesagt. Vorhergesagte Epitope wurden synthetisiert und deren Bindungsaffinität zu den entsprechenden HLA-Allelen mit dem UV-mediated exchange ligand MHC-Bindungstest getestet. Der Nachweis einer T-Zell-Antwort gegen die Peptide mit hoher Bindungsaffinität wurde mittels PBMCs von Patienten mit ausgeheilter HDV-Infektion erbracht. Aminosäuresubstitutionen innerhalb der identifizierten

Epitope in der untersuchten Kohorte wurden charakterisiert, und die entsprechenden mutierten Peptide synthetisiert. Die Auswirkungen der Mutationen auf HDV-spezifische T-Zellantwort wurden durch Stimulation von T-Zellen mit "Wildtyp" und "Mutanten" Epitopen untersucht. Eine Koinzidenz des molekularen "Footprint" auf den L-HD Ag und bestimmte HLA-Moleküle wurde zur Identifizierung von zusätzlichen CTL-Epitope angewendet.

Die phylogenetische Analyse der Sequenzen aus dieser Studie zeigte, dass alle 142 Isolate dieser Kohorte zum HDV-Genotyp 1 gehörten, der weltweit am weitesten Verbreitete ist. Longitudinale Studien mit HDV Isolaten von chronische HDV infizierte Patienten zeigten eine höhere Mutationsrate während der frühen Phase der Infektion im Vergleich zur späten Phase der Infektion. Dies ist ein Hinweis auf eine Virusanpassung an den Wirt über die Zeit.

Unter Verwendung von CTL-Epitope-Prediction-Tools und MHC-Bindungsassay identifizierten wir zwei HDV-Epitope, spezifisch für das HLA-B*27-Allel (aa 99-108: RRDHRRRKAL und aa 103-112: RRRKALENKK). Durch die Stimulation von PBMCs mit diesen Peptiden konnte die Induktion von IFN-γ-Sekretion induziert und damit die Spezifität dieser CD8+ T-Zell Epitope bestätigt werden. Durch Sequenzanalyse fanden wir einen HLA-B*27-restringierten Immunselektionsdruck für zwei Aminosäuren (R105K, K106M) für beide Epitope des HDV in Isolaten aus HLA-B*27-positiven Patienten. Mutierte Epitope wurden von HDV-spezifischen T-Zellen im Vergleich zu den "Wildtyp" Epitopen nicht erkannt. Die Immunevasions-Mutationen können als Ausweichmechanismus zur Erhaltung der HDV-Infektion gewertet wurde. Durch eine Footprint-Analyse von L-HD Ag-Sequenzen entdeckten wir einen Polymorphismus in Korrelation mit bestimmten HLA-Allelen von Patienten. Einen starken Footprint auf dem L-HD Ag wurde in den HLA-B*15-positiven Patienten nachgewiesen. In Virusisolaten aller 8 HLA-B*15-positiven Patienten (100%) unserer Kohorte wurde ein Serin (S) zu Asparagin (N) Substitution in diesem Bereich (S170N) beobachtet, während nur 11% (11/96) der HLA-B*15 negativen Gruppe diese Substitution aufwiesen. Dies ist deutlicher Hinweis auf eine Immunevasions-Mutation an dieser Position. In diesem Sequenzabschnittes L-HD Ag wurde ein neues HLA-B*15-restringierte CD8+ T-Zell-Epitop durch Prädiktion identifiziert (aa 170-179: SMQGVPESPF). Dieses Epitop wurde auch mit T-Zell-Assay in HLA-B*15-positiven Patienten mit ausgeheilter HDV-Infektion bestätigt. Eine weitere Analyse von Footprints innerhalb aller 214 Aminosäuren des L-HD Ag und 35 HLA-Klasse I-Allele zeigte einen signifikanten Immunselektionsdruck auf zusätzliche Bereiche für HLA-B*13, -B*37 und -B*41.

In dieser Arbeit wurden zum ersten Mal HLA-B*27 und HLA-B*15-beschränkte MHC-Klasse-I-Epitope für HDV beschrieben. Die Immunantwort gegen die identifizierten HLA-B*27 und HLA-B*15-Epitope könnte für eine T-Zell induzierte Ausheilung eine HDV-Infektion bei Patienten verantwortlich sein. Substitutionen innerhalb dieser Epitope resultieren jedoch in

einer Immunevasion der CD8+-T-Zell-Antwort bei HLA-B*15/27-positiven Individuen. Die hohe Mutationsrate zu Beginn der Infektion kann auf eine Virus-Wirt-Interaktion zurückzuführen sein, was zu einer Anpassung des Virus führt. Letztlich trägt diese zur Entwicklung von neuen zirkulierenden HDV-Isolaten bei. Die Korrelation von Mutationen in HDV Isolaten zu den HLA-Hintergründen scheint ein Ansatz zu sein, neue HDV-spezifische Epitope zu identifizieren. HLA-B-Allele scheinen einen stärkeren Selektionsdruck auf HDV auszuüben, als Lokus-A. Folglich haben Lokus-B Allele eine größere Wirkung auf die virale Evolution. Diese Ergebnisse sollten beim Design von Impfstrategien in Betracht gezogen werden.

Chapter 1: Introduction

Hepatitis delta virus (HDV) was discovered in 1977 by Mario Rizzetto following the identification of a novel antigen-antibody system in liver biopsies of hepatitis B virus (HBV) chronic carriers [1]. HDV is an RNA virus which causes delta hepatitis infection, known as the most severe form of chronic viral hepatitis. Later it was described that the presence of HBV is inevitable for transmission of this virus-like agent to chimpanzees [2] and was readapted in woodchuck models by using woodchuck hepatitis virus (WHV) as the helper [3]. In fact, HDV needs the hepatitis B surface antigen (HBs Ag) for its packaging and for the production of new virions; therefore, because of this essential need for a helper virus, HDV has been considered as a defective virus.

1.1. Structure and life cycle of HDV

HDV virions are very small in size; in average they are about 36 nm in diameter [4]. HDV genome is a circular single-stranded RNA with only ~1700 nucleotides (nt) which is the smallest described to infect human [5]. Since this circular conformation of HDV genome was not found amongst RNA viruses of animals, infectious subviral agents of plants (viroids/virusoids), with a small circular RNA genome, were taken as the closest relative to HDV. Based on these similarities, a double rolling circle mechanism was suggested as a model for viral RNA replication [6]. Nevertheless, there are a few differences between the plant subviral agents and HDV, for example, viroids' genome could be several folds smaller in size than that of HDV, viroids do not code for any protein and they do not need a helper virus. With respect to virusoids, some of them do need a helper virus which provides an RNA dependent RNA polymerase (RdRp) for the replication process; whereas, in the case of HDV, the helper virus (HBV) only provides the envelope proteins (HBs Ag) and HDV is indeed able to replicate in the absence of its helper, HBV. Another unique structural feature of HDV is that both genomic and antigenomic RNAs have a high degree (~70%) of intramolecular base-pairing (self-complementary) which shapes a semi-double-stranded

structure [7-9]. The intermolecular self-complementarity of a circular genome (rod-like structure) was also seen in viroids.

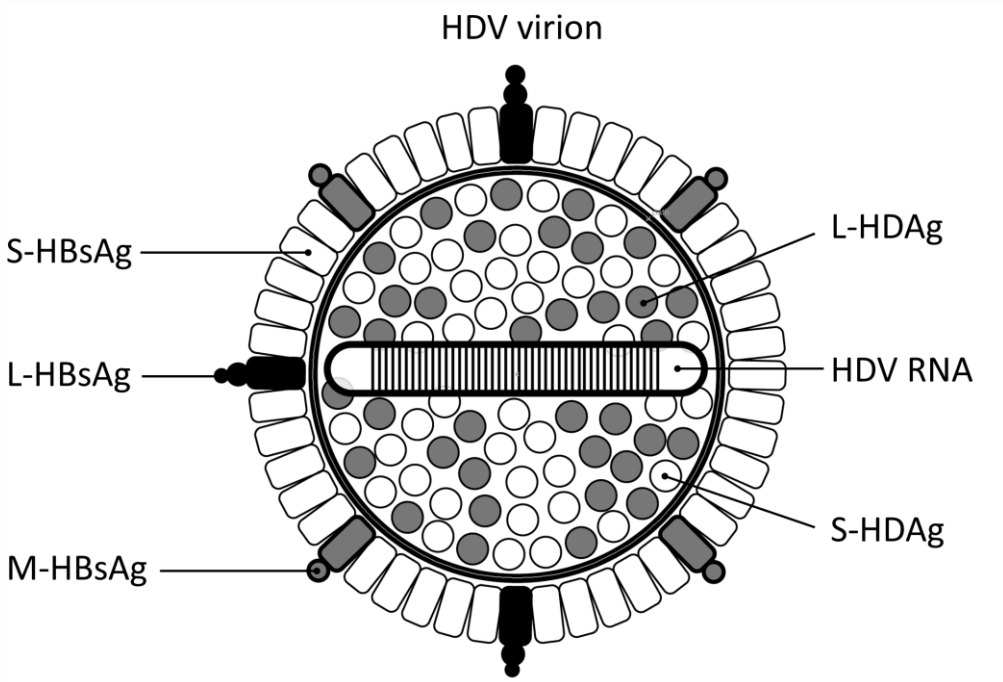


Figure 1.1 Schematic illustration of HDV virion (modified from [10]).

Genomic RNA is packed with multiple (70-200) copies of both L-HD Ag and S-HD Ag to form a ribonucleoprotein (RNP). The viral envelope, derived from HBV, comprises the three forms of HBV surface antigen: S-HBs Ag (S), M-HBs Ag (preS2-S) and L-HBs Ag (preS1-preS2-S).

Earlier studies indicated that the viral genome was not the only viral RNA present in the infected cells [11]. Further investigations described one ORF on the complementary viral genomic RNA, referred to as the 'anti-genome'. Therefore, HDV genome was considered as RNA of negative polarity. Although analysis of nucleotide sequence of anti-genomic RNA revealed the only ORF of HDV [8], this circular anti-genomic RNA cannot be translated to a viral protein directly. Therefore, further explorations revealed a linear mRNA of about 0.8 kb with a polyadenylated site at 3' and a unique 5'-cap which is transcribed from the genomic RNA [12]. This mRNA encodes for the only viral protein which appears as two isoforms: the small hepatitis delta antigen (S-HD Ag) of 195 amino acid (aa) residues and the large form (L-HD Ag) which has about 19 additional residues at its carboxyl terminus. The unique phenomenon of RNA editing as a part of post-transcriptional modification allows switching from S-HD Ag (195aa) with a molecular weight of 24 kDa to L-HD Ag (214aa), 27 kDa. The S-HD Ag mediates RNA replication and RNA editing; whereas, L-HD Ag down-regulates the RNA replication and triggers virion assembly [13-16]. Synthesis of mRNA encoding L-HD Ag is the result of RNA editing process on ORF of HD Ag at 196th codon, so called amber/W

site. The term “RNA editing” mostly refers to nucleotide changes via deamination process. As the Figure 1.2 shows, the antigenomic RNA, exclusively, undergoes RNA editing at the amber/W site [17], that is, a UAG, the stop codon (amber) for S-HD Ag, changes to UIG by Adenosine deaminase (ADAR) [18, 19]. Deaminase activity of ADAR on the adenosine at the amber/W site of antigenomic RNA results in an intermediate antigenomic RNA with an inosine (I) at this site which then would be used as a template for transcription of a second genomic RNA and since this inosine in UIG codon acts as a guanosine (G), results in the synthesis of genomic RNA with an ACC codon at this position [20]. The genomic RNA with an ACC codon at the amber/W site would be the template for transcription of a second antigenomic RNA and the L-HD Ag mRNA with UGG codon at this position which encodes for tryptophan (W) and allows the synthesis of an additional 19-aa to form L-HD Ag [20]. It was shown in separate studies, when cells or hepatocytes of chimpanzees were transfected with a cDNA clone encoding S-HD Ag only, unexpectedly, both S-HD Ag and L-HD Ag were produced after initiation of replication [21, 22].

1.2. Ribozyme activity of the HDV genome

Early studies indicated a self-cleaving activity on HDV genome, so called ribozyme [23, 24]. This self-cleavage was related to a ~100nt-fragment present on both genomic and anti-genomic RNAs. HDV ribozyme plays a central role in replication cycle. Ribozyme activity was first considered a very unique feature of HDV genome and a limited number of viroids. However, after identification of HDV ribozyme structure by crystallization [25], a homolog of HDV ribozyme was found in a human messenger RNA, the cytoplasmic polyadenylation element binding protein 3, CPEB3 [26]. A more recent study, proved that HDV-like ribozymes are detectable widely in various types of organisms [27]. Based on a double-rolling circle model, using cellular RNA polymerases, HDV circular genome is replicated into a linear multimeric of transcripts which serves as the substrate of ribozyme to produce monomeric circular RNAs by self-cleavage and then self-ligation [28]. Although circularization of monomeric RNAs was initially considered to be, exclusively, the revers reaction of HDV ribozyme, host RNA ligases might be also involved [29]. Discovery of an HDV-like ribozyme in human genome, CPEB3, was consistent with another earlier report, where the authors introduced a cellular homolog for HD Ag [30], implying that HDV might have been evolved from a human transcriptome. However, this precursor mRNA, CPEB3, was also preserved among other mammalians; in addition, HDV-like ribozymes were later identified in a various number of organisms [27]. In contrast with the S-HD Ag, the L-HD Ag is not only a self-limiting factor for RNA replication, but also, because of its additional 19 aa at the carboxyl terminal, is essential for virus assembly [14, 31]. A series of studies showed that a

heterogeneous number of HD Ag molecules (between 70 and 200), including both small and large isoforms, are packed inside one HDV virion along with the genomic RNA (and not anti-genomic RNA, for some unknown reasons) [32, 33].

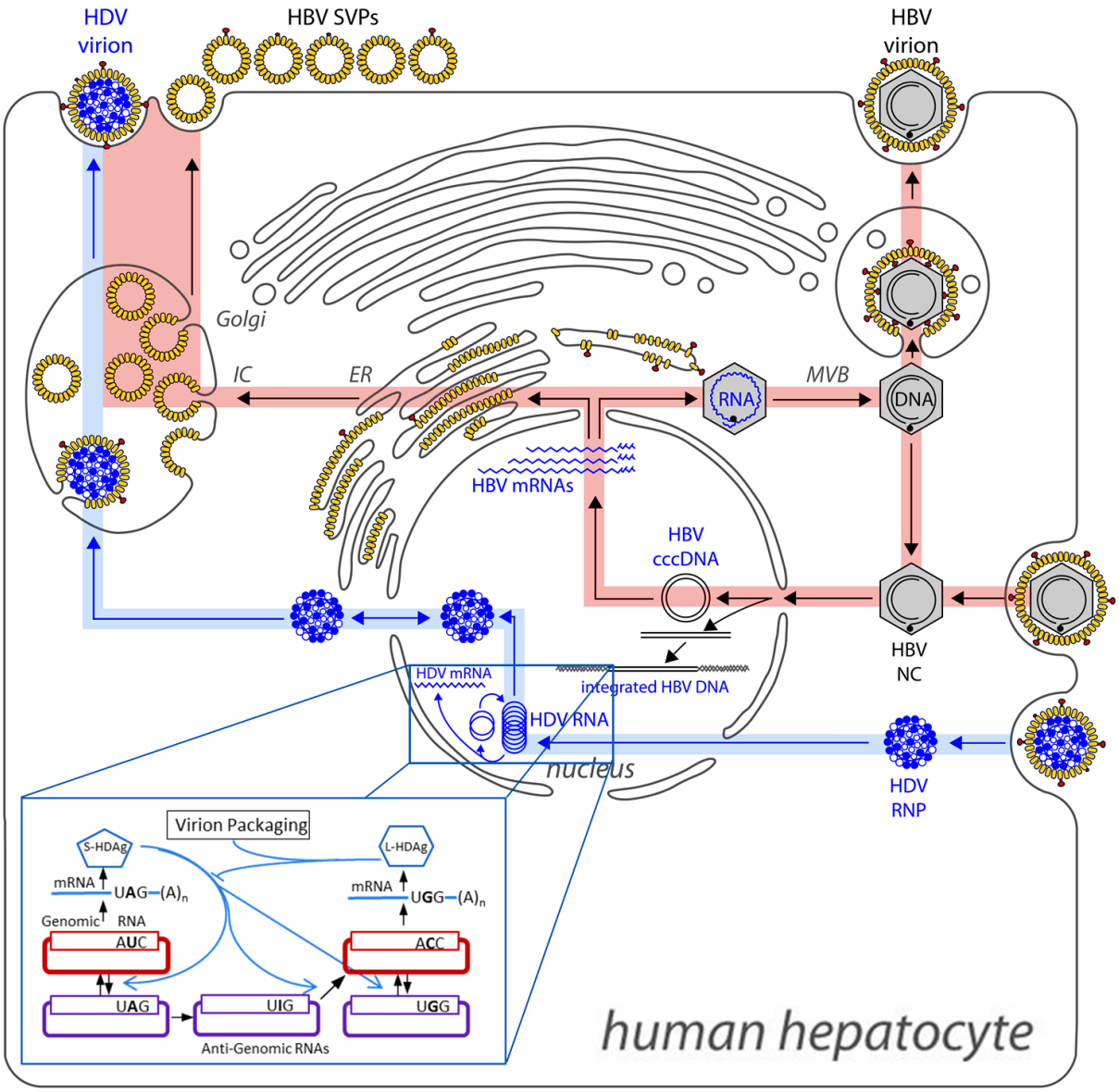


Figure 1.2 proposed HBV/HDV replication cycle in hepatocytes.

HDV genome replication is depicted in the lower-left corner with some proposed details on HDV RNAs replication/transcription cycle and the role of RNA editing. These illustrations are adapted from [20, 34].

Ultimately, HDV inquires its helper virus, HBV, to get the envelope and release from the cells. HBV encodes three related envelope proteins: small (S), middle (M) and large (L) HBs Ag. It is necessary for HBV virion formation to achieve all three isoforms of HBs Ag in its envelope. In contrast, HDV assembly may not require all SML forms which is closer to formation of empty subviral particles (SVPs). However, L protein is necessary for the

infectivity of both HBV and HDV virions [35-37]. In fact, a region of pre-S1 inside L-HBs Ag acts as the ligand of the recently discovered receptor of HBV and HDV [38].

1.3. Animal models for HDV

Animal models available to study the natural history of HDV are quite restricted simply because that tentative model must be able to host two viruses, HDV and HBV. Primate models including chimpanzees, that are susceptible to HBV, are the best models for studying HDV natural history and developing vaccine and therapeutic agents. Infection of chimpanzees with serum of patients with chronic HBV/HDV infection played a central role in the characterization of HDV and its pathogenesis. However, utility of these animals as a model is extremely restricted by the ethical concerns. Although duck hepatitis B virus (DHBV) has been used as a model for HBV, ducks cannot support infection and virion formation of HDV in the duck hepatocytes. The very next successful model, after chimpanzees, has been woodchucks. Chronically infected woodchucks with woodchuck hepatitis virus (WHV) can be superinfected with HDV, experimentally. The early studies showed that WHV infected woodchucks can develop HDV superinfection after inoculation with HDV derived from chimpanzees [3]. This conclusion was later extended to show that woodchuck hepatitis surface antigen (WHsAg) is playing the role of HBs Ag in this setting and can efficiently support the HDV packaging [37]. Although woodchucks are able to develop HDV infection when they are inoculated experimentally, there is no data showing that HDV infection occurs in woodchucks naturally. Woodchucks have been broadly used in HDV studies since. Nevertheless, there are several restriction factors and challenges regarding these animals to be a perfect model for studying HDV.

Interestingly, an earlier study showed that the cellular machinery of mouse hepatocytes is able to support HDV replication but not virus production, obviously due to the absence of its helper, HBV [39]. Humanized chimeric mouse models, uPA/SCID mice, have been developed very recently; these models are able to produce infection of both viruses, HBV and HDV [40].

1.4. Genetic variability and epidemiology of HDV

Viral genome was first described back in 1986 [7, 8, 11] as a single-stranded circular RNA which is quite unique among infectious agents infecting animals; the plant viroids are the closest known relatives. As an RNA virus, genetic variability of HDV is very complex. Beside the nature of developing the quasispecies by RNA viruses, the unique HDV genome replication and ADAR dependent RNA editing will add more to this complexity. Since the first description of HDV genome sequence in 1986 [8], the proposed HDV nomenclature was updated once from 3 genotypes [41] to 8 major clades, HDV-1 to HDV-8 [42]. HDV-1 found

to be the most prevalent HDV type in the world including North America, central Europe, Africa, some parts of Asia and the western pacific region [43-46]. HDV-2, initially isolated from Japanese and Taiwanese patients [47, 48], is prevalent in far East and some parts of North Asia [49], while HDV-4 is the next most prevalent type in Asia, mostly in Taiwan and Okinawa island [50]. HDV-3, the most divergent one among all types [51], occurs exclusively in South America [41]. HDV-5 to HDV-8 were initially discovered in Africa [9, 42]; however, HDV-8 was recently reported from some patients in Brazil [52]. There are several studies, including experimental infection in woodchucks [53], investigated the association between HDV (/HBV) genotypes and the severity of the disease. Several studies showed that infection with HDV-1, the worldwide distributed type, can lead to a wide range of disease outcomes from a very mild to a severe form of fulminant hepatitis [54]. HDV-3 (along with HBV-F) was shown to be related to a more severe acute infection than other types [55, 56]. On the contrary, HDV-2 and -4 reported to be associated with a milder disease than HDV-1 and -3 [54]. Ultimately, it would be important to understand how these different HDV genotypes have an impact on clinical outcomes and response to therapy.

As a defective virus, prevalence of HDV should be dependent on the prevalence of its helper virus, HBV; however, it does not seem to be that straightforward, for example, in China, where HBV prevalence is high, HDV is low prevalent. During the late 80s, there was an estimation that about 5% of all HBs Ag carriers (~15 million patients) are superinfected with HDV. Later studies reported a very heterogeneous prevalence of HDV in different regions of the world, from 0% up to 80% of HBs Ag positive individuals. Several epidemiological studies have underestimated the prevalence of HDV. Target groups, for instance, is obviously the most important source of data; therefore, study subjects must be chosen carefully. Healthy HBs Ag positive carriers and HBV positive blood donors were targeted very often to achieve the HDV prevalence rate in given regions. For obvious reasons, these groups are not appropriate representatives of all HBV carriers for studying HDV prevalence. Due to the fact that HBV/HDV infected individuals are those with severe symptoms unable to donate blood samples and, therefore, will be excluded in these studies. There was a suggestion that hepatitis D is disappearing and is completely controlled in the near future [57, 58]. This assumption is not correct and resulted in an unrecognized increase of HDV in recent years [59, 60] due to immigrations from areas endemic for HDV and also drug abusers in Western European countries [61, 62]. Ultimately, HDV remains a potential threat and a major health problem especially for healthy careers of HBs Ag.

1.5. Natural course and pathogenicity

Two classical forms are described for HDV/HBV infection; simultaneous infection with both HBV/HDV and HDV superinfection of a patient with a preexisting and persisting HBV infection. Simultaneous infection is transient and self-limiting with over 95% recovery rate in adults, similar to that of HBV monoinfection. A high titer of IgM anti-HBc is a key marker indicating HBV/HDV simultaneous infection. Obviously, acute simultaneous HDV infection occurs in susceptible individuals (*i.e.*, anti-HBs antibody negative) and demonstrates the same characteristic as for HBV acute infection. Superinfection, the worse scenario, is defined by HDV infection of an individual with an established HBV infection. In this setting, a patient may undergo a severe acute phase of infection which, in most cases (80-90%), is followed by a chronic HDV infection [63]. Acute phase of infection is characterized by an elevation of liver enzymes (ALT/AST) and bilirubin in the serum, indicating the liver damage. There is a common period of 1-2 months before the symptoms of HDV acute infection appears. Another key feature of an acute phase of both types of HDV infection is the transient appearance of HD Ag which could be detected by repeated testing during the first 2 weeks of illness [64]. This quick vanishing of HD Ag shortly after the onset of infection is highly related to immune response. Consequently, in HIV patients, having some levels of immunodeficiency, HD Ag lasts longer [65].

Apart from the above mentioned two types of infections, simultaneous infection and superinfection, there is a third pattern, in which HDV is able to infect the hepatocytes but cannot leave the cells simply because of absence of HBV and HBs Ag. This was initially described as a possible model happening in the transplantation settings [66, 67] and very recently was proven experimentally in humanized mice where even 6 weeks after HDV monoinfection, HBV superinfection was able to rescue HDV virion assembly and virus production [68]. This study confirms that HBV is not needed for viral replication but only for packaging and viral release [69]. Clinical and virological significance of HBV-independent HDV infection is not well understood.

HDV superinfection in the most cases, but not always [70], leads to the suppression of HBV replication and consequently, absence of HBV DNA [71]; therefore, liver damage is considered to be caused by HDV.

In summary, acute simultaneous HDV infection, mostly transient and self-limiting, is not going to be the main concern. However, HDV superinfection accelerates liver fibrosis and consequently leads to early and frequent development of cirrhosis [72, 73]. A European study reported that the 5- and 10-year probability of survival of HDV-related cirrhotic patients estimated to be 49% and 40%, respectively [74]. Chronic HDV infection can increase

liver-related mortality by faster induction of cirrhosis and HCC development when compared with HBV or hepatitis C virus (HCV) mono-infection [75, 76].

1.6. Treatment of HDV infection

A few years after the discovery of HDV, interferon-alpha (IFN-α) was used as the first approved therapy for hepatitis delta [77, 78]. In the early studies, it was shown that IFN is able to inhibit HDV in short-term, but relapse may occur 1-2 years after discontinuation of therapy [77]. IFN was later replaced with pegylated-interferon (PEG-IFN). Modification of proteins, in general, by adding poly ethylene glycols can lead to an increase in protein's half-life and stability and a decrease in immunogenicity of this therapeutic protein [79, 80].

Nucleos(t)ide analogs, widely used against chronic hepatitis B (CHB) and C (CHC), were also tested against chronic hepatitis D (CHD) either alone or in a combination with interferon (Conventional or pegylated) assuming a better virologic response. The first attempt was the administration of ribavirin against CHD; however, no effective reduction of biological (*e.g.* ALT) or virologic markers was observed [81]. Ribavirin was later also tested in a combination with IFN-α-2a for a period of 2 years in patients with chronic HDV, no increase in virologic response was observed over IFN-α monotherapy, however [82]. Likewise, lamivudine [83] or adefovir [84] alone or in combination with IFN demonstrated no improved virologic response when compared with IFN monotherapy. Entecavir (ETV), a potent NA for treatment of HBV, was also given to 13 HBV/HDV patients in a clinical trial for a year. At the end of treatment with ETV, HBV DNA, which was positive in 8 out of 13 patients, was not detectable anymore; however, ETV showed no significant effect on HDV RNA or alanine aminotransferase (ALT) [85]. Overall, NAs alone could not enhance the virologic response in hepatitis delta patients, nor could they in combination with IFN.

Apparently, as shown in Table 1.1, interferon therapy does lead to the viral response in a large proportion of patients and the consequent improvement of some biochemical markers, as long as therapy is not broken, can be detected. Unfortunately, the relapse rate, after discontinuation of therapy, is relatively high. Obviously, interferon has no HDV-specific antiviral effects; moreover, overall response rate of delta hepatitis patients to interferon is lower than in other viral infections. There have been vigorous efforts to develop new drugs against HDV, yet the progress is quite slow. Designing any new antiviral agent is very dependent upon a well-characterized viral target which is reflected in virus life cycle. HDV highly relies on the cell machinery for its replication, and, as a defective virus, HDV needs HBV, but only for packaging and virus assembly. There was a suggestion that targeting HBV replication would affect and control, presumably indirectly, the HDV infection. However, since HDV only requires HBs Ag from HBV to be able to complete its life cycle and, thus far

approved, anti-HBV agents do not affect HBs Ag production, no beneficial effects can be expected from HBV treatment on HDV infection.

Table 1.1 Summarized important studies and controlled clinical trials treating patients with (PEG-) IFN alone, IFN combined with NAs, and NAs alone.

Agent(s)	N	Duration (Months)	Response EOT[a]→EOFu[a]	Reference
IFN-α-2	6		First report	[78]
IFN-α-2a t.i.w, 9MU/m^2	14	12mo	71% → 43%	[86]
IFN-α-2a t.i.w, 3MU/m^2	14	12mo	29% → 14%	
IFN-α-2a t.i.w, 9MU/m^2	10	24mo	45% → 20%	[82]
IFN-α-2a t.i.w, 9MU/m^2+ ribavirin	21	24mo	57% → 24%	
lamivudin (LAM)	17	12mo	11% → 11%	[83]
IFN-α-2a t.i.w, 9MU/m^2+LAM	14	12mo	50% → 38%	
IFN-α-2a t.i.w, 9MU/m^2	8	12mo	50% → 50%	
PEG-IFN-α-2b	12	12mo	17% → 17%	[87]
adefovir	30	12mo	0%	[84]
PEG-IFN-α-2a	31	12mo	23%	
PEG-IFN-α-2a + adefovir	29	12mo	24%	
PEG-IFN-α-2	104	12mo	42% → 23%	[88]

[a] EOT: End of therapy; EOFu: End of follow-up

Entry inhibitor

The notion that HBV and HDV share the envelope protein (HBs Ag), led to an assumption that both viruses use the same mechanism/cellular-receptor to enter the hepatocytes. In addition, the observation that only those HDV virions comprising the L-HBs Ag were infectious [35, 36], predicted that a domain in preS1 must be responsible for HBV/HDV attachment and entry to the cells, which turned out to be the case [89]. An earlier study had showed that, the myristylation of the preS1 domain in L-HBs Ag is absolutely necessary for HBV infectivity but not for the virus production [90]. The myristoylated N-terminal preS1 domain of the L protein is involved in attachment to the very recently discovered receptor, the sodium taurocholate co-transporting polypeptide (NTCP) [38]. The discovery of HBV (HDV) receptor, NTCP, has revolutionized the field of HBV/HDV research [91]. Furthermore, preventing the HBV/HDV entry by blocking this receptor has offered a new approach to fight HBV and HDV infections [92]. Studies in mice models proved that blocking NTCP by a peptide, which mimics the myristoylated N-terminal of preS1 domain (*e.g.* Myrcludex B), was able to effectively abolish HBV infection [40, 93, 94]. To date, Myrcludex B is in clinical trial

phase 2a, and preliminary results have been reported on the proof-of-concept in patients with chronic HDV infection. Although Myrcludex B was well tolerated [95], a daily administration for 6 months led to only 1-2 $\log_{10}$ reduction in viral load [96] which presumably will be reversed after discontinuation of therapy.

Prenylation inhibitor

Prenylation of L-HD Ag, a post-transcriptional modification, is a necessary step in HDV life cycle and virus assembly [15]. Several studies, *in vitro* and *in vivo*, confirmed the antiviral efficacy of prenylation inhibitors against HDV [97, 98]; therefore, inhibition of prenylation was considered as an antiviral strategy for treatment of HDV infection in the last decade. In a very recent study, lonafarnib, a prenylation inhibitor, was given to chronic HDV patients in double-blind, randomized, placebo-controlled phase 2a trial [99]. At this phase, drug was well tolerated, and it did show a reduction of virus level which was significantly correlated with the serum level of this compound [99]. It would at least take a couple of years to develop the above mentioned new therapies and make them available for treatment of chronic hepatitis delta.

Liver transplantation

Besides all restrictions and complications, liver transplantation continues to be the ultimate treatment option to rescue HBV/HDV infected patients with end-stage liver damage. To date, pegylated interferon remains the only effective treatment option in chronic HDV infection, despite the low response rate and its unfavorable side effects.

While interferon induces MHC class I and II antigen presentation pathways, the low response rate in patients might be due to the lack of immunogenic viral epitopes being able to trigger a protective immune response in majority of patients after interferon therapy. Therefore, it is important to understand the significance role of immune response to HDV in viral clearance and final outcome of the disease.

1.7. Immune response to viral infections

Viruses are intracellular infectious agents which are able to stimulate both arms of immune system, innate and adaptive immunity. Innate immunity, besides its all physical and mechanical barriers, which prohibit entering of pathogens at the first place, consists of phagocytic cells (*e.g.* macrophages) which fight a wide range of pathogens including viruses without requiring prior exposure. This distinguishes such responses form adaptive immunity, defined by developing a specific immune response against a particular pathogen, which occurs as an adaptation to the new infections. This response involves the production of specific antibodies whose presence reflects the infection to which the given person has been exposed.

As for innate immunity, single-stranded DNA (*e.g.* from HBV) or double-stranded RNA (*e.g.* from HDV), which are seen as a part of viruses' life cycle, can be recognized by, so called, pattern recognition receptors (PRRs). This is able to activate a series of signaling cascades leading to releasing type I IFN, among which IFN-α [100]. Type I IFNs induce transcription of a large group of genes with antiviral activities leading to the host resistance to viral infections. Moreover, IFNs also activate some key components of the innate and adaptive immune systems such as professional antigen presenting cells (APCs), *e.g.* dendritic cells (DCs). Activation and maturation of DCs by type I IFNs (IFN-α/ß) include upregulation of MHC molecules (mostly MHC class I), costimulatory molecules and chemokines which in turn leads to efficient CD4+ or CD8+ T cell responses (adaptive immunity) and enhancement toxicity of natural killer (NK) cells (innate immunity) [101, 102].

Lymphocytes are the main cellular arm of adaptive immunity with two major types of the cells, B cells and T cells. Both cells are originated from the bone marrow. Maturation of the B cells occurs in bone marrow, whereas, thymus is the maturation place for the T cells. B cells, after further developments to plasma cells, produce antibodies which, in most cases, are protective by neutralizing the pathogens. For instance, anti-HBs antibodies are produced after acute HBV infection or vaccination. T helper (T_h) cells (known as CD4+ T cells) and cytotoxic T cells (known as CD8+ T cells or cytotoxic T lymphocytes, CTLs) are the main classes of the T cells. CD4+ T cells stimulate B cells and help them to develop to plasma cells which are able to produce antibodies. Cytotoxicity of CD8+ T cells is also enhanced by the help of CD4+ T cells. Ultimately, CD8+ T cells or CTLs are known as the most effective cells against intracellular pathogens such as viruses. T cell receptors of CD8+ and CD4+ T cells are able to recognize the cytosolic antigens (epitopes) presented by MHC classes I and II, respectively. After antigen recognition, CD4+ T cells produce a series of cytokines, among which interleukin-4 (IL-4) is the main inducer of B cells to develop plasma cells (antibody producing cells) and memory B cells. CD8+ T cells or CTLs, by contrast, have a direct cytotoxic effect on the infected cells. Cytotoxic T cells, after activation and recognition of an infected cell, release cytotoxic proteins, perforin and granzymes, and lead to cell lysis and cell death. While immune responses mediated by the cytotoxic T cells are believed to be the central response against intracellular pathogens, failure of these cells is a very common feature. There are several reasons, why T cells may fail in mounting a proper response when it comes to intracellular pathogens like viruses. This may be due to primary or secondary failure of the T cell response such as exhaustion of the T cells due to the overload amount of antigens which happens often in viral infections. Exhaustion phenomenon is defined by expression of inhibitory molecules such as cytotoxic T-lymphocyte-associated Protein 4 (CTLA-4) and programmed death-1 (PD-1.) Immune escape mutation (substitution) is

another mechanism by which viruses may evade immune response and establish a persistent infection [103].

1.7.1. Impact of immune response and immune escape to hepatitis viruses

HAV

Hepatitis A virus (HAV), a non-enveloped RNA virus, was first identified in 1973 [104]. HAV is acquired through fecal-oral transmission and, obviously, infects the liver. Hepatitis A is highly endemic in Central and South America, Africa, the Middle East, Asia, and the Western Pacific. Upon administration of a single dose of highly immunogenic hepatitis A vaccines, more than 95% of adults will develop protective antibodies. Such an efficient vaccination has led to a significant decline in prevalence of HAV [105]. Despite vaccine-dependent control of this disease, acute hepatitis A is known as a self-limiting liver disease. In order to determine the mechanism of immune response to hepatitis A, by which HAV resolution is reached, our group conducted an investigation to identify the T cell targets on HAV [106]. Schulte and colleagues characterized several novel CTL-epitopes against which corresponding and strong T cell responses were detected in acute HAV patients [106]. Nevertheless, no significant immune escape mutations were observed within those identified epitopes; which implies that development of immune escape substitutions is more likely to occur during a long lasting chronic infection than in a self-limiting acute phase. In fact, by introduction of immune escape substitutions, viruses may undergo fitness cost through developing not so replication-efficient variants. Therefore, introducing amino acid substitutions may not be preferred by viruses.

HBV and WHV

To date, about 248 million individuals are estimated to be HBV carriers around the world [107]. Over 90% of adults are able resolve the acute HBV infection. Protective (neutralizing) antibodies are produced after a natural infection or HBV vaccination. Despite an efficient vaccine, which is able to induce protective antibodies in over 95% immunocompetent individuals, millions of people are still suffering from a chronic infection. Chronic HBV infected patients, most of whom transmitted prenatally, are not able to achieve seroconversion. Therapeutic vaccination was developed in recent years as an immunomodulatory tool to restore the immune function in those patients. In a recent study of WHV (a close model to HBV) by Kosinska and colleagues, it was shown that treatment of chronic WHV infected woodchucks with a combination of a DNA prime-adenovirus (AdV) boost vaccination and entecavir led to induction of strong CD4+ and CD8+ T cell responses against WHV and sustain virologic response achieved in half of animals [108]. Moreover, combination of therapeutic vaccine and blockage of PD-L1 (T cell exhaustion marker)

resulted also in restoration of immune response [109]. These data, collectively, emphasize on the importance of immune response in final outcome of such infections. Recovery from acute HBV infection was strongly associated with CD8+ T cell responses directed against the HBV core protein [110]. In contrast with B cells, recognition of epitopes by T cells is exclusively restricted to the human leukocyte antigen (HLA) molecules and their ability (affinity) to present a given antigen. The types of HLA molecules are defined by numbers, *e.g.*, HLA-A*01, HLA-A*02, HLA-A*03 and so forth. HLA-A*02 is one of the most prevalent alleles around the world, and, conspicuously, most studies have focused on a set of HLA-A*02-restricted CD8+ T cell epitopes for analyses of HBV immunology. Earlier studies reported that natural amino acid substitutions within a dominant HLA-A*02-restricted epitope (HBV 18-27 core epitope) result in failure of CD8+ T cells to recognize this epitope; in addition, the altered epitope may act as antagonist for antiviral CTLs [111, 112]. Another study recently showed a correlation between specific HLA alleles and HBV variations. In this study, specific sequence variants within identified epitopes were often observed in patients carrying the relevant HLA alleles [113]. A very recent study identified novel residues on HBV genome (isolated from chronic patients) under selection pressure. This study proved that selected substitutions are able to reproducibly impair the recognition by HBV-specific CD8+ T cells [114].

HCV

HCV is a small positive-stranded RNA virus, and was first known as Non-A/Non-B Hepatitis. In contrast with HAV and HBV, after an acute infection by HCV, a persistent infection can be established in 55% to 85% of cases [115], which increases a risk of developing cirrhosis and hepatocellular carcinoma. The global prevalence of viraemic HCV infections is estimated to be 80 million cases [116]. The role of immune system in viral clearance or viral persistence in HCV infection has been vastly investigated. HCV-specific CD4+ and CD8+ T cell responses are believed to be associated with viral clearance during the early phase of the infection [117-119]. Since HCV-specific CD8+ T cells can be detected in both chronic and recovered patients, there must be some other mechanisms involved in failure of immune response, such as exhaustion of the T cells. It was reported that the failure of T cell response in HCV chronic patients may be due to the impairment of proliferative, cytokine, and cytotoxic effector functions [120]. This impaired function of the T cells is defined as T cell exhaustion due to the continuous exposure of the T cells to the overload of viral antigens. Exhaustion of the T cells can be correlated with the high expression of some inhibitory markers such as PD-1, and, interestingly, by blocking these molecules the dysfunction of those specific T cells can be restored [121-123].

In recent decades, a large number of HLA-restricted (CD4+/CD8+) T cell epitopes of HCV have been identified. The role of immune escape mutations (substitutions) has been well studied in HCV infected individuals. Early studies of HCV in chimpanzees suggested that escape mutations in CTL epitopes may play a role in evasion of the relevant immune responses [124, 125]. Later studies in human HCV infection, revealed also a series of CTL escape mutations which may support establishing a persistent infection in humans [126].

HDV

Superinfection of HBV carriers with HDV leads to a flare of serum aminotransferases (ALT and AST) indicating the liver damage. Due to the fact that HDV itself is not cytopathic [127], this observation may indicate that the destruction of hepatocytes is mediated by cytotoxic T lymphocytes; however, the role of CTL response in HDV infection has not been well studied. Overall, the immune response to HDV is far less characterized than that of HAV, HBV or HCV.

There is evidence that HDV may interfere with IFN-α signaling *in vitro* by blocking the activation of the Tyk2, which is part of the Janus kinase-signal transducer and activator of transcription (JAK-STAT) pathway [128]. Inhibition of IFN-α by HDV may represent an important mechanism of viral persistence and patients' resistance to IFN therapy.

Antibodies against HD Ag are present during acute (at low titer) and chronic (at high titer) HDV infection [129]. Early IgM anti-HDV antibody, after an acute HDV infection, is followed by IgG antibodies which may last for years. Notably, IgM anti-HDV antibody persists at high titer in patients with progressive liver disease [130, 131]. Experimentally, anti-HDV antibodies have not shown neutralizing/protective activities in chimpanzees, reinfected with HDV in presence of high titer of anti-HDV antibodies [132], or in woodchucks with vaccine-induced anti-HDV antibodies [133].

Animal models (mice and woodchucks) were immunized using different protocols to induce a protective immune response against HDV. In these studies, different approaches were used including immunization directly with HD Ag (expressed in yeast or E-coli) or HD Ag expression vectors (DNA or vaccinia virus). However, not all of the vaccinated animals, regardless of the applied protocol, were protected against HDV superinfection [133-135].

To understand the role of T cell immunity in HDV infection, it is crucial to identify and characterize the immunodominant epitopes on the single delta protein, HD Ag. However, to date, a very limited number of studies were able to detect those epitopes (Figure 1.3). In an early study by Nisini *et al.* regarding the role T helper cells in HDV infection, 4 MHC class II-restricted epitopes were identified [136]. In this study those patients with inactive HDV infection (anti-HDV positive/HDV RNA negative) responded to the synthetic HD Ag, whereas the same responses could not be detected in active (HDV RNA positive) patients [136]. This

observation may explain the severity of HDV-induced liver damage in HBV/HDV infected patients when compared with healthy carriers of HBV. Experimental studies suggested that protective CTL responses against HDV can be generated [137]; however, little is known about the immunodominant CTL epitopes and their contribution in virologic response in patients. Utilizing 12 predicted HLA-A*02:01-restricted HDV peptides, Huang *et al.* were able to identify two CD8+ T cell HDV epitopes, aa 26-34 and aa 43-51 [138]. Specific CD8+ T cells to these two identified HLA-A*02:01-restricted epitopes were also detected in immunized HLA-A*02-transgenic mice with a DNA construct coding for HD Ag. In addition, epitopes-specific CTLs were detected in peripheral blood mononuclear cells (PBMCs) from 2 out of 4 HLA-A*02 positive patients with resolved HDV infection. In this study, HDV-specific CTL response was detected by stimulation of PBMCs with tetramers loaded with the mentioned HLA-A*02-restricted epitopes. Huang and colleagues stated that HDV-specific CTL response was not detectable in HLA-A*02 positive patients with active HDV infection reflecting a possible mechanism of failure of HDV-specific CD8+ T cells in controlling the virus in these patients. It was not addressed which possible mechanisms may be involved in failure of immune response in the studied patients.

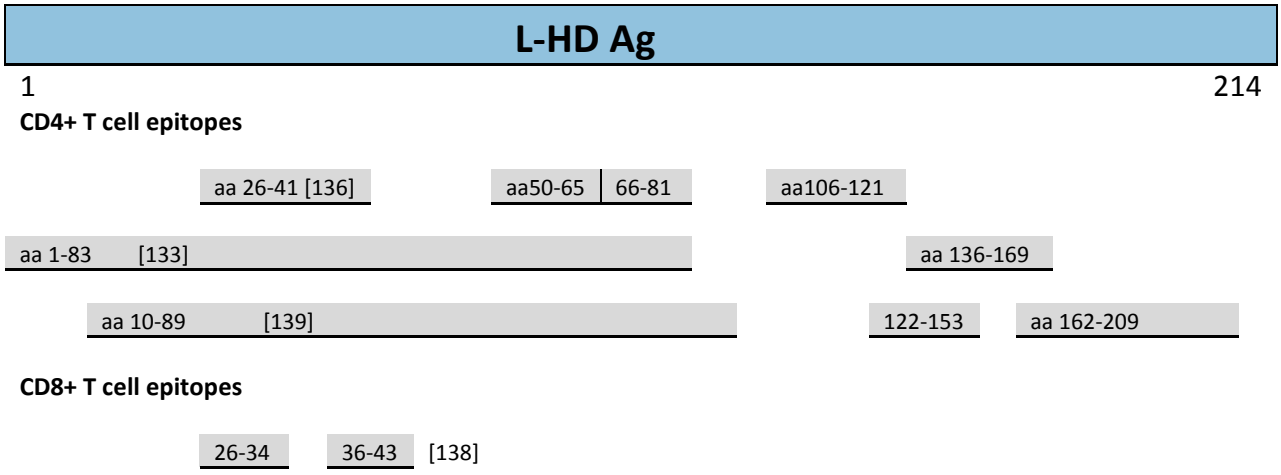


Figure 1.3 Localization of CD4+ and CD8+ T cell epitopes on ORF of HD Ag (adapted from [140]).

Taken together, HDV-specific CD8+ T cell response is not well characterized in patients with HBV/HDV co-infection, and there are open questions regarding HDV immunobiology in human which will be addressed throughout this dissertation.

1.8. Aims of this study and open questions

Hepatitis delta virus infection is considered as the most severe form of viral hepatitis. Up to date, there is no efficient treatment of chronic HDV infection, nor is there a prophylactic vaccine available. Despite being a very severe disease, HDV infection has been one of the most neglected forms of viral hepatitis. Our knowledge about immunopathogenesis and virus-host interactions is also very limited. In fact, HDV-specific T cell response and its possible epitopes on the single HDV protein have been poorly investigated. Only a small fraction (about 10%) of infected HDV patients is able to recover from the disease. Since antibodies to HDV are not neutralizing, HDV clearance, in these patients, is probably mediated by CTL response. It is not clear which immunological defects lead to the failure of immune response to HDV. Development of escape mutants within CD8+ epitopes, similar to HIV, HBV or HCV, may be an important reason why majority of HDV infected patients fails to eliminate the virus and establish a persistent infection. In many studies, of these viruses, it could be shown clearly that substitutions within protective epitopes can lead to failure of the corresponding CD8+ T cell response and consequently establishing a chronic and long-lasting infection.

In the light of these observations, the aims of this study were as follows:

A) First we aimed to **investigate the genetic variability and evolution of HDV** especially within L-HD Ag coding region of isolates from HDV-1 (the most prevalent genotype in the world).

B) Next, we aimed to **identify and characterize HDV-specific CD8+ T cell response and the corresponding epitopes**. For this purpose, we applied prediction tools and MHC binding assay using UV-mediated ligand exchange method and finally confirm the predicted epitopes in the patients with resolved HDV infection by T cell assay.

C) To address the possible mechanism(s) of viral evasion from specific T cell responses, we wanted to **evaluate the contribution of viral amino acid substitutions on failure of CTL response and consequently persistency of HDV.**

D) Finally, if this is an escape mechanism for HDV and resistant mutants are indeed selected under immune pressure by CTL responses, the ultimate goal of this thesis would be to **address the question whether HDV sequence analysis would allow identification of so far unknown HLA class I-restricted CTL epitopes.** These data together with clinical information may help to understand which HLA alleles are probably protective against HDV infection.

Chapter 2: Methods

2.1. Patients and samples

The study population, derived from patients with chronic HBV/HDV co-infection, was recruited from the following collaborating centers:

1. Institute of Virology, University of Duisburg-Essen, Germany
2. Institute of Virology, Technical University of Munich (TMU), Munich, Germany
3. Hannover Medical School, Hannover, Germany
4. Vall d'Hebron Hospital, Barcelona, Spain
5. Department of Gastroenterology and Hepatology, Molinette Hospital, Turin, Italy
6. Tehran Hepatitis Center, Tehran, Iran
7. Institute of Immunology, Ludwig-Maximilians-University (LMU), Munich, Germany
8. Department of Gastroenterology and Hepatology, University Hospital Freiburg, Germany

The local ethics committees of the University Hospital Essen have approved this study according to the 1975 Declaration of Helsinki guidelines. All patients were interviewed and informed about the study before they agreed to take part in this study.
We collected the serum samples of those individuals who were tested HDV positive at the above mentioned centers. The viral RNA was extracted from those serum samples and in some cases extracted RNA was provided by the centers.
PBMCs were isolated from the whole blood samples of selected patients by density gradient centrifugation (Biocoll; Biochrom, Germany). We also used whole blood, PBMCs or serum in some cases (when there was no blood sample available) to extract the genomic DNA for HLA typing. All serum samples and extracted materials were stored at -80°C.

2.2. Viral RNA and genomic DNA isolation

To study the HDV sequences of our subjects, we isolated viral RNA from patients' serum. RNA extraction was performed on automated DNA/RNA extractor (MagNA Pure 96 System, Roche, Basel, Switzerland). For manual method, viral RNA was extracted from 200 µl of serum. Extracted material was either used freshly or stored at -80 C for long-term storage.
Genomic DNA was isolated from 200-400 µl whole (EDTA) blood samples using QIAamp® DNA Mini and Blood Mini Kit (QIAGEN®, Germany) and in some cases when there was no

access to the fresh blood samples, a relatively large volume (up to 2ml) of serum was used to extract genomic DNA. According to the fact that genomic DNA of destroyed cells is released to the serum, there is a chance of isolation of DNA from serum samples. To increase the probability of DNA isolation from serum samples, a large volume of serum (up to 2ml) was centrifuged at the maximum speed (14000 rpm) and 200-400 µl was taken from the bottom of each tube for extraction, exactly according to the manufacturer's instruction (Appendix I). For yielding higher concentration of DNA after the elution, only 50µl of buffer AE (Elution Buffer) was added and incubated for 1 min before centrifugation at 8000 rpm for 1 min. The DNA concentration was measured using NanoDrop™ (Thermo Scientific, Waltham, Massachusetts, USA). Finally the sample proceeded for HLA typing.

2.3. Peripheral blood mononuclear cells (PBMCs) isolation

PBMCs were obtained by standard density gradient centrifugation (Biocoll; Biochrom, Germany). The EDTA or heparinized blood was 1:2 diluted with phosphate buffered saline (PBS: pH 7.45). Leucosep® tubes were filled with 15 ml separating solution (density 1.077 g/ml) and centrifuged for 1 min at 700g, in room temperature (the Biocoll goes under the barrier). Diluted blood was carefully transferred to the 50 ml Leucosep® tube. Diluted blood makes a separate layer above the barrier. To proceed with separation of blood's component, the Leucosep® tubes were centrifuged at 800g for 15 minutes in room temperature. At this step we should see 4 different layers that are erythrocytes, separation solution (Biocoll), PBMCs and plasma (from bottom to top). To reach PBMCs, the plasma layer was removed and the PBMCs layer (directly above the barrier) was transferred to a new 50 ml Falcon prefilled with 10ml PBS and centrifuged at 250g for 10 minutes in room temperature in order to wash the cells. The supernatant was discarded. The pellet of the cells were re-suspended with 10 mL of RPMI and filled up to 50 ml with RPMI and centrifuged 250g for 10 minutes in room temperature. Supernatant was discarded and in order to count the cells, the pellet was re-suspended in 5 ml of RPMI and 50µl of this suspension was diluted with 450µl of PBS for automated cell counter. To count manually, 10µl of this suspension was mixed with 90µl of trypan blue. After counting, cells were transferred to CryoTubes® and placed in Mr. Frosty freezing container which was then put into the -80 C freezer for 24 hours before transferring them to the liquid nitrogen container for longer storage.

2.4. MHC (HLA) class I typing

The extracted genomic DNA was applied to perform a two-digit HLA-class-I typing by a Luminex™ Polymerase Chain Reaction-Sequence Specific Oligonucleotide Probe (PCR-SSO) using the LABType™ SSO Kits (One Lambda, Canoga Park, CA, USA) as described

elsewhere in details [141]. A high resolution assay was done on some patients to define a four-digit HLA type or sub-groups.

2.5. HDV sequence identification

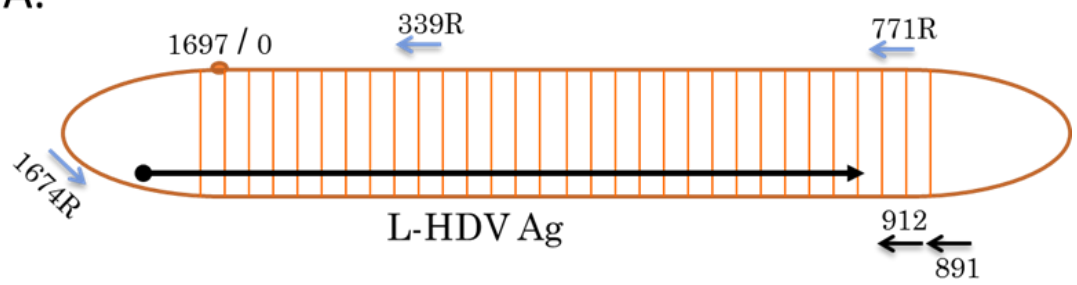
2.5.1. L-HD Ag open reading frame amplification

Reverse Transcriptase: Ten µl of extracted RNA was used to make the complementary DNA (cDNA). First the RNA was denatured in the presence of 1 µl reverse primer (771R, Figure 1B) at 98°C for 5 min and then at 62°C for another 5 min, and then the cDNA was synthesized by adding 4 µl 5X reaction buffer, 4 µl dNTPs (2.5mM), 1 µl MMLV-RT polymerase (Promega, USA) and 0.5 µl of RNAse inhibitor in a total volume of 20 µl at 37°C for 1 hour.

First PCR: A reaction mix of 45 µl containing 23.75 µl ddH2O, 10 µl 5X reaction buffer, 5 µl primer 891(10pmol), 5 µl primer 339R(10pmol), 1 µl dNTPs (10mM) and 0.25 µl GoTaq (Promega, USA) was made and added to 10 µl of cDNA obtained from RT-PCR. Then the reaction was done under the following thermal profile: denaturation at 94°C for 10 Min. and a 35 cycle of 94°C for 30 Sec, 54°C for 45 Sec., and 72°C for 90 Sec. and then an elongation step at 72°C for 7 Min. and finally stop at 4°C.

Nested PCR: The same reaction mixture as for the 1st PCR was prepared but with primers 912 and 1674R and then 5 µl of the 1st PCR product was added to each tube to undergo a nested PCR applying the following thermal conditions: denaturation at 94°C for 10 Min. and a 35 cycle of 94°C for 30 Sec, 56°C for 45 Sec., and 72°C for 90 Sec. and then an elongation step at 72°C for 7 Min. and finally stop at 4°C.

A.



B.

Primers	Sequence	Location
771R	5'-CGGTCCCCTCGGAATGTTG-3'	753-771 nt
339R	5'- GCTGAAGGGGTCCTCTGGAGGTG-3'	319-341 nt
1674R	5'-AGAAAAGAGTAAGAGYACTGAGG-3'	1652-1674 nt
912	5'-GAGATGCCATGCCGACCCGAAGAG-3'	912-936 nt
891	5'-AGGTCGGACCGCGAGGAGGT-3'	891-910 nt

Figure 2.1(A) Schematic representation of HDV genome and the location of PCR amplified fragments. (B) List of primers used for reverse transcription (RT), amplification of L-HD Ag open reading frame and sequencing.

2.5.2. PCR product purification and preparation of positive samples for sequencing

The final positive PCR products were purified using either QIAquickSpin kit or *QIAquick Gel Extraction Kit* (see Appendix section for the detailed instruction). To avoid interfering the sequencing by unspecific PCR products which were shorter or longer than the region of interest (L-HD Ag coding region), Gel extraction protocol were performed. Briefly, the total volume of the PCR product of each sample was loaded on a 1% agarose gel and run for an hour and then the correct band of each sample was cut out. The PCR product of this piece of gel was extracted and purified exactly according to the instruction of *QIAquick Gel Extraction Kit* (see Appendix II). The concentration of the purified samples was measured by NanoDrop® and finally they were prepared and sequenced at LGC Company on the ABI 3730 XL platforms for traditional Sanger sequencing.

2.5.3. Phylogenetic analysis and genotyping of HDV

From each patient/time-point both forward and reverse sequences were evaluated using Geneious program version 7.0.6 (Biomatters, Auckland, New Zealand) and exported as FastA file. A master FastA file of full-length L-HD Ag sequences from this study and 53 reference sequences (HDV genotypes 1-8) from NCBI GenBank were used to perform multi sequence alignment using ClustalX2 method [142] integrated to the Geneious program. A Maximum Likelihood (ML) phylogenetic tree was constructed under the Tamura-Nei substitution model using MEGA software v6 [143]. To address the reliability of pairwise

comparison and phylogenetic tree analysis, bootstrap test was performed with 1,000 replicates. All sites containing alignment gaps and missing information had been removed before the calculation started by selecting “Complete-deletion” option under “Gaps/Missing Data Treatment”. In order to improve the likelihood of the phylogenetic tree, “Nearest-Neighbor-Interchange” was applied as a ML heuristic method. Genotypes of strains were determined by comparison to the full-length of HD Ag reference sequences, representing all HDV genotypes.

2.5.4. Longitudinal study

On closer examination of sequencing data, we observed a few number of sites showing single nucleotide polymorphisms (SNPs) (Figure 2.2). Therefore, we were interested to know whether these changes are correlated with immune pressure at these sites promoting evolution of the virus in those HDV infected patients over time. To investigate this, we analyzed longitudinal samples from 24 patients with chronic HDV infection for varying time lengths, from 3 months up to 6 years. In this study also the only ORF of HDV encoding for L-HD Ag was amplified and sequenced. Then the post-infection mutation rate was determined over time. Evolutionary distance was also calculated within the individual patients and within the whole group of the patients.

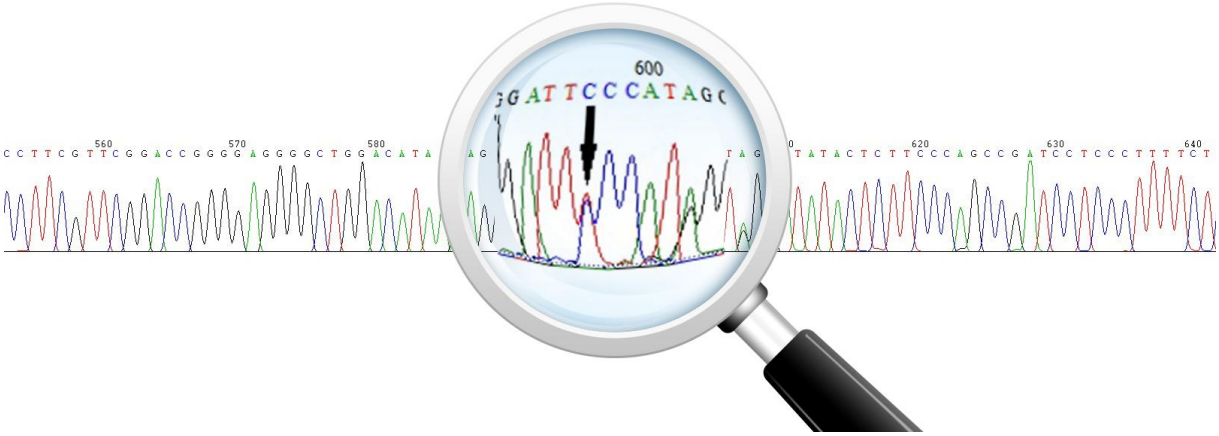


Figure 2.2 Schematic illustration of single nucleotide polymorphisms (SNPs) within the L-HD Ag coding region.

2.5.5. HDV genome database

To facilitate molecular analysis of this study, a local database was developed using Geneious program version 7.0.6 (Biomatters, Auckland, New Zealand). This database consists of all L-HD Ag sequences from this study and all thus far experimentally identified and submitted HDV isolates retrieved from GenBank available at https://www.ncbi.nlm.nih.gov/genbank. This database also includes of all thus far studied partial L-HD Ag sequences as well as HDV full genome sequences.

2.5.6. Deep sequencing by pyrosequencing

In order to identify variations at lower detection level of direct sequencing technology, we applied next generation sequencing by ultra-deep pyrosequencing (UDPS), to pin down quasispecies in specific region.

All nucleotide (nt) numbering positions are according to HDV sequence ID: dbj|D01075.1 | HPDCGDA. One µl of the first PCR product obtained by primers 891fw and 339rv was processed by a second run of PCR by primers flanking positions 956 and 1360 which included M13 fagus universal tails in 5' ends followed by specific HDV sequences:

M13HDV956forward:

5'-GTTGTAAAACGACGGCCAGTTCACTGGGGTCGACAACTCTG-3'

M13HDV1360reverse:

5'-CACAGGAAACAGCTATGACCGTAGACTCCGGACCTAGGAAGA-3'

To identify every patient, the final product was subjected to 15-cycles of re-amplification using primers composed by a complementary universal M13 primer (either upstream or downstream) followed by a Roche's Validated Multiplex IDentifier (MID, indicated in italics), and with oligoA or B at 5' or 3' end of the upstream or downstream primer respectively (in bold) 5'-oligoA/oligoB+MID+M13-3':

OligoAMIDM13fw

5'-**CGTATCGCCTCCCTCGCGCCATCAG***MID*GTTGTAAAACGACGGCCAGT-3'

OligoBMIDM13rv

5'-**CTATGCGCCTTGCCAGCCCGCTCAG***MID*CACAGGAAACAGCTATGACC-3'

OligoAMIDM13fw and OligoBMIDM13rv primers were preloaded and lyophilized in 96 wells ready to use PCR plates. Therefore, the use of M13 universal primers allow the incorporation of MID sequences by means of these PCR plates preloaded with OligoAMIDM13fw and OligoBMIDM13rv primers, avoiding sample identification errors.

The final 515 bp amplicon included a 405 nt HDV specific sequence in which, once discarding primer sequences, 362 nt HDV sequences were analyzed (positions 977 to 1338) by UDPS. This sequence covered HD Ag amino acids 89 to 208. Further general details on this method are discarded elsewhere [144]. Since there was no NGS facility available in our institute, this part of experiments was designed and performed together with our collaborator at Vall d'Hebron Hospital, in Barcelona, Spain.

2.6. Epitope binding prediction

Epitope binding prediction was performed using 2 different prediction methods/databases (IEDB, SYFPEITHI). This epitope prediction on the L-HD Ag, from HDV genotype 1, was done for frequent MHC class-I alleles in European population [141]. Based on recommended

algorithm of the IEDB online tool and the SYFPEITHI algorithm, we selected those peptides with the best predicted scores to be tested for a possible binding affinity *in vitro*. Further details on the predicted methods applied in this study are described elsewhere [145, 146]. Overall 18 predicted peptides with the highest ranking including 2, 3, 2, 3, 7 and 2 peptides restricted to HLA-A*01, A*02, A*03, A*24, HLA-B*07 and B*27, respectively, were chosen to be tested in an HLA binding assay.

2.7. MHC binding affinity evaluation by peptide exchange assay

We applied UV-mediated peptide exchange assay using PeliChange™ p*HLA-A*01:01, A*02:01, A*03:01, A*24:02, B*07:02 and B*27:05 kits (Sanquin, Amsterdam, The Netherlands) and PeliScreen™ HLA class I ELISA kit (M1924, Sanquin) to determine the binding affinity of the predicted epitopes as described elsewhere [106] and illustrated below.

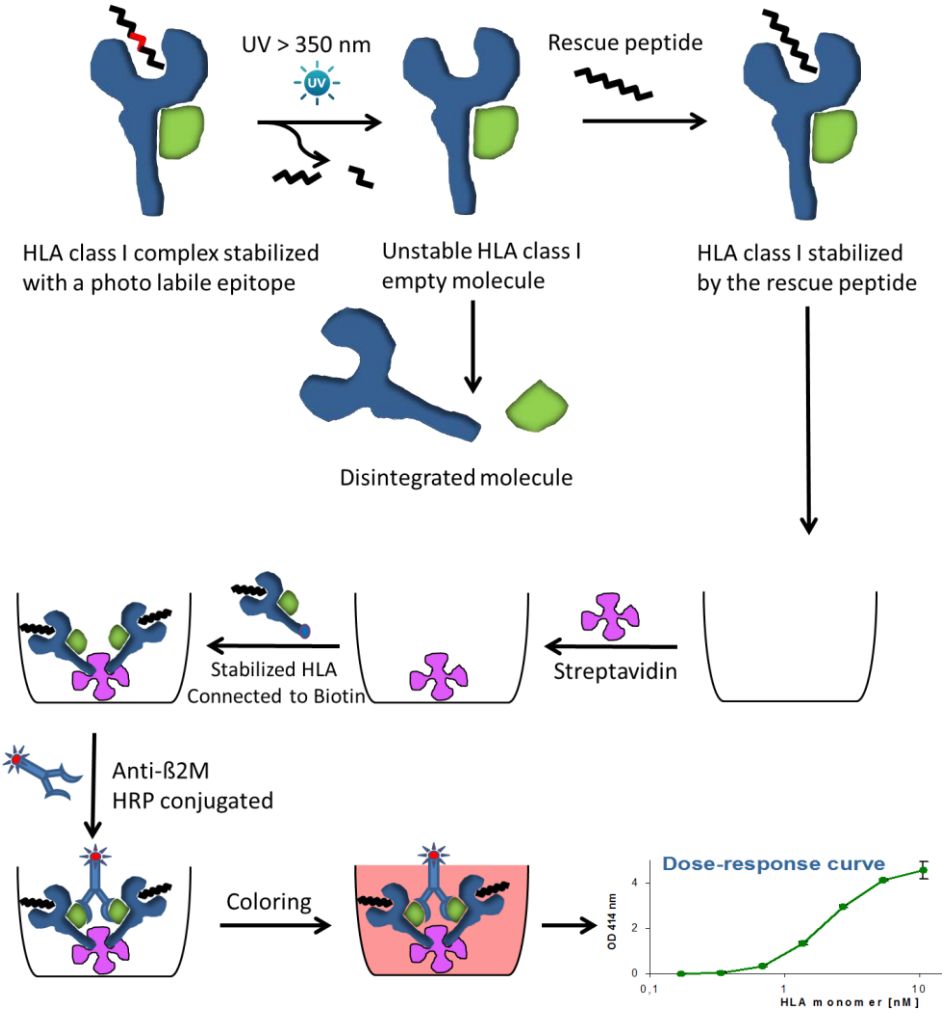


Figure 2.3 principle of MHC binding assay as explained elsewhere [147].

Briefly, peptide-exchange reactions were performed by the exposure of conditional pHLA

complexes to long wavelength UV using a 366-nm UV lamp (CAMAG, Muttenz, Switzerland), in the presence or absence of the indicated peptide. Subsequently, peptide-exchange efficiency was analyzed using the PeliScreen™ HLA class I ELISA kit. This ELISA determines the presence of intact HLA class I complexes in an exchange–reaction mixture (Figure 2.3). A known HLA allele-specific ligand with high affinity was used as positive control for each corresponding allele. The absorbances of all peptides were evaluated relative to that of the high-affinity binder which was set at 100%. Figure 2.3 demonstrates schematically the process and principle of the MHC binding assay.

2.8. Detection of HDV specific CD8+ T cells

2.8.1. Polyclonal antigen-specific expansion of T cells

Isolated PBMCs were cultured in 1 ml (2×10^6 per 1 ml) of complete medium (RPMI 1640 supplemented with 10% fetal calf serum, 100 U/ml penicillin, 100 g/ml streptomycin) containing anti-CD28 antibody (0,5 µg/ml; BD Biosciences). Cells were stimulated with 8 pools of 31 synthetic overlapping 16-mer peptides (EMC Microcollections, Tübingen, Germany) corresponding to the complete sequence of L-HD Ag (Table 2.1), added to a final concentration 1µg/ml per peptide. On day two, 1 ml of complete medium and interleukin-2 (10 U/ml; Roche) was added. On day 10 the cells were tested for IFN-γ secretion by intracellular cytokine staining after re-stimulation with the peptide pools (day 10). Positive cells from the pools were then re-stimulated with the individual peptides from the respective pools and tested for IFN-γ (day 12). For additional PBMCs with predictive T cell response to a specific peptide, this relevant overlapping peptide was used for T cell stimulation instead of the whole library. For further characterization, the minimal 9-mer or 10-mer peptides within the corresponding overlapping 16-mer were synthesized and applied for T cell assay.

2.8.2. Intracellular cytokine staining (ICS)

Intracellular cytokine staining was performed as described elsewhere [148] with some modifications. Prior to intracellular cytokine staining, cells were cultured for 5 h in the presence of peptide pools, or individual peptides (10 µg/ml), and 5 µg/ml of brefeldin A (BFA; Sigma-Aldrich). Cell surface staining was performed using allophycocyanin (APC)-conjugated anti-CD8 and phycoerythrin (PE)-conjugated anti-CD4 (BD Biosciences) at 4 °C for 15 min. Dead cells were excluded from analyses using Viaprobe reagent (BD Biosciences). Permeabilization of cells was performed with Beckton Dickinson permeabilizing solution (BD Biosciences), according to the manufacturer's protocol. After washing, cells were stained with fluorescein isothiocyanate (FITC)-conjugated anti-IFN-γ mAb (BD

Biosciences). Data were acquired on a FACS Calibur flow cytometer. Data files were analyzed with FlowJo X.0.7 software (Tree Star Inc., Ashland, Oregon).

Table 2.1 list of overlapping peptides applied for T cell stimulation.

Nr.	Name	Position	Sequences
1	1	1-16	MSRSESRKNRGGREEL
2	2	2-17	SRSESRKNRGGREELL
3	3	10-23	RGGREELLEQWVAGRK
4	4	18-33	EQWVAGRKKLEELERD
5	5	26-41	KLEELERDLRKTKKKL
6	6	34-49	LRKTKKKLKKIEDENP
7	7	42-57	KKIEDENPWLGNIKGI
8	8	50-65	WLGNIKGILGKKDKDG
9	9	58-73	LGKKDKDGEGAPPAKR
10	10	66-81	EGAPPAKRARTDQMEV
11	10a	66-81	EGAPPAKRARTDRMEV
12	11	74-89	ARTDQMEVDSGPGKRP
13	11a	74-89	ARTDRMEVDSGPGKRP
14	12	82-97	DSGPGKRPLRGGFTDK
15	13	90-105	LRGGFTDKERRDHRRR
16	14	98-113	ERRDHRRRKALENKKK
17	15	106-121	KALVNKKKQLSAGGKN
18	16	114-129	QLSAGGKNLSKEEEEE
19	17	122-137	LSKEEEEELRRLTEED
20	18	130-145	LRRLTEEDERRERRVA
21	19	138-153	ERRERRVAGPPVGGVN
22	20	146-161	GPPVGGVNPLEGGSRG
23	21	154-169	PLEGGSRGAPGGGFVP
24	22	162-177	APGGGFVPNLQGVPES
25	23	170-185	NLQGVPESPFSRTGEG
26	23a	170-185	NLQGVPESPFARTGEG
27	24	178-193	PFSRTGEGLDIRGNQG
28	24a	178-193	PFARTGEGLDIRGNQG
29	25	186-201	LDIRGNQGFPQDTLFP
30	26	194-209	FPQDTLFPADPPLSPQ
31	27	199-214	LFPADPPLSPQSCRPQ

2.9. Identification of amino acid residues under immune selection pressure and discovery of novel epitopes using "SeqFeatR" program

In order to identify which protein residue is under selective pressure and also at the same time predict novel epitopes, we analyzed the association of L-HD Ag polymorphism with particular HLA alleles. In large dataset of isolates, such as this study, manual searching for residues under statistically significant selection pressure may be challenging in large datasets. Therefore, powerful computational tools may facilitate analyzing the association between amino acid polymorphism and some phenotypes (features) of the patient, *e.g.* HLA types. One powerful tool for analyzing of association between **Seq**uence and a **Feat**ure (*e.g.* HLA) is called **SeqFeat**R which is now also available online at https://seqfeatr.zmb.uni-due.de [149]. SeqFeatR has been very recently introduced and successfully applied in several immunomonitoring laboratories. This tool indicates the association of each single amino acid residue with particular HLA allele by constructing 2×2 contingency table analyzing the number of coincident HLA allele and amino acid residues and calculating the respective *p* values. Two main approaches have been introduced for calculation of this association, Frequentist and Bayesian. In the Frequentist approach, Fisher's exact test is used to make contingency tables and calculate the *p* values for each amino acid-HLA allele correlation [150]. For the analysis, SeqFeatR is able to read multiple sequence alignments in FastA format. Technically, SeqFeatR does not make sequence alignment; this should be performed elsewhere. There are several methods for performing a multiple sequence alignment. We applied Clustal program inside Geneious® software licensed version 7.0.6. It is important to lable the sequence with the feature of interest, HLA alleles in this case, and to keep the order and location of each letter and digit of the feature (HLA alleles) equal through the whole sequences. The sequence alignment can then be uploaded to the server for calculations. After analysis of the alignment, this tool provides a large table of all amino acid positions and the relevant *p* values to those HLA alleles given in the name of each sequence. SeqFeatR also provides for each HLA allele a Manhattan plot where *p* values of each amino acid residue are plotted logarithmically along the whole ORF or the selected protein sequence.

Chapter 3: Results

3.1. HDV sequence identification, phylogenetic analysis and genotyping

3.1.1. Description of cohort of studied subjects

One hundred and forty two subjects with persistent HDV/HBV infection from 8 different health care centers were enrolled into this study (87 male, 31 female and gender of 22 patients were not submitted; mean age 47 years, range 21-75 years). The collaborating centers were located in Essen, Germany (n=38); Hannover, Germany (n=42); Munich, Germany (n=4); Barcelona, Spain (n=17) Turin, Italy (n=20) and Tehran, Iran (n=21). All 142 patients were tested positive for anti-HDV antibody (the main inclusion criteria). Although HDV RNA was not detected in all patients at the diagnostic section of our collaborating centers, they were included in the study to be tested with the sensitive PCR method established for this study. In total, 123 out of 142 patients were successfully HLA typed for locus A and B alleles (Table 3.1). Based on the quality of the samples and/or viral load, the whole ORF encoding L-HD Ag of 645nt was amplified, sequenced and analyzed from a total number of 116 patients. Obviously, further patient materials were not provided because the follow-up with those patients was interrupted due to several reasons.

Table 3.1 demographic, HLA types, and virologic information of the subjects enrolled in this study.

Patient	Code	Sex	Age	Center	HLA Type	HDV RNA	Anti-HDV
1	**E01**	F	45	Essen	HLA-A: *01, *02 HLA-B: *15(62), *57	+	+
2	**E02**	M	38	Essen	HLA-A: *23, *24 HLA-B: *35, *62	+	+
3	**E03**	M	42	Essen	HLA-A: *01, *02 HLA-B: *07, *08	+	+
4	**E04**	F	32	Essen	HLA-A: *02, *03 HLA-B: *35	+	+
5	**E05**	M	46	Essen	Not determined	+	+
6	**E06**	M	48	Essen	HLA-A: *01, *03 HLA-B: *18, *50	+	+
7	**E07**	M	35	Essen	HLA-A: *24 HLA-B: *44, *51	+	+
8	**E08**	M	52	Essen	HLA-A: *02, *24 HLA-B: *35, *52	+	+
9	**E09**	F	43	Essen	HLA-A: *24, *68 HLA-B: *14(65), *35	+	+
10	**E10**	M	75	Essen	HLA-A: *02 HLA-B: *41, *51	+	+
11	**E11**	M	38	Essen	HLA-A: *02, *68 HLA-B: *15(62)	+	+
12	**E12**	M	21	Essen	HLA-A: *01, *24 HLA-B: *35, *58	+	+
13	**E13**	M	60	Essen	HLA-A: *01, *03 HLA-B: *08, *35	+	+
14	**E14**	M	36	Essen	HLA-A: *02, HLA-B: *55, *60	+	+
15	**E15**	M	50	Essen	HLA-A: *02, *26 HLA-B: *38, *51	+	+
16	**E16**	F	57	Essen	HLA-A: *02, *11 HLA-B: *51, *55	+	+

17	**E17**	M	50	Essen	HLA-A: *01, *11 HLA-B: *35, *51	+	+
18	**E18**	M	44	Essen	HLA-A: *03, *11 HLA-B: *35, *51	+	+
19	**E19**	M	53	Essen	HLA-A: *11, *30 HLA-B: *18, *35	+	+
20	**E20**	M	55	Essen	HLA-A: *02,*03 HLA-B: *18, *41	+	+
21	**E21**	M	64	Essen	HLA-A: *02, HLA-B: *50, *52	+	+
22	**E22**	M	36	Essen	Not determined	+	+
23	**E23**	M	26	Essen	HLA-A: *24, *31 HLA-B: *35, *51	+	+
24	**E24**	M	75	Essen	HLA-A: *3, *24 HLA-B: *18, *35	+	+
25	**E25**	M	50	Essen	Not determined	+	+
26	**E26**	M	58	Essen	HLA-A: *03, *32 HLA-B: *18, *51	+	+
27	**E27**	M	61	Essen	HLA-A: *1, *26 HLA-B: *13, *38	-	+
28	**E28**	F	32	Essen	HLA-A: *02, *24 HLA-B: *38, *51	+	+
29	**E29**	F	64	Essen	HLA-A: *3 HLA-B: *07, *40(61)	+	+
30	**E30**	M	35	Essen	HLA-A: *02, *29 HLA-B: *15(62), *44	+	+
31	**E31**	M	37	Essen	HLA-A: *01, *11 HLA-B: *35, *40(61)	+	+
32	**E32**	M	31	Essen	HLA-A: *66, *68 HLA-B: *52, *57	+	+
33	**E33**	M	48	Essen	HLA-A: *24 HLA-B: *18, *51	+	+
34	**E34**	M	74	Essen	HLA-A: *00, *00 HLA-B: *14	+	+
35	**E35**	F	45	Essen	HLA-A: *02, *26 HLA-B: *00, *00	+	+
36	**E36**	M	48	Essen	HLA-A: *02 HLA-B: *44	+	+
37	**E37**	M	41	Essen	Not determined	+	+
38	**E38**	F	44	Essen	HLA-A: *3, *32 HLA-B: *15(62), *52	+	+
39	**H1**	F	35	Hannover	HLA-A: *02, *03 HLA-B: *07, *35	+	+
40	**H2**	M	40	Hannover	HLA-A: *02, *24 HLA-B: *38, *73	+	+
41	**H3**	M	40	Hannover	HLA-A: 34, *74 HLA-B: *15(72), *35	+	+
42	**H4**	M	30	Hannover	HLA-A: *30, *32 HLA-B: *13, *35	+	+
43	**H5**	F	55	Hannover	HLA-A: *11, *23 HLA-B: *49, *50	+	+
44	**H6**	M	40	Hannover	HLA-A: *03, *30 HLA-B: *13, *51	+	+
45	**H7**	N/A	N/A	Hannover	HLA-A: *01, *30 HLA-B: *38, *40	+	+
46	**H8**	N/A	N/A	Hannover	HLA-A: *26, *29 HLA-B: *07, *44	+	+
47	**H9**	N/A	N/A	Hannover	HLA-A: *01, *29 HLA-B: *27, *44	+	+
48	**H10**	F	25	Hannover	HLA-A: *02, *68 HLA-B: *14, *44	+	+
49	**H11**	N/A	N/A	Hannover	HLA-A: *24 HLA-B: *07, *51	-	+
50	**H12**	N/A	N/A	Hannover	HLA-A: *68, *69 HLA-B: *35, *51	+	+
51	**H13**	F	47	Hannover	HLA-A: *02, *26 HLA-B: *07, *15(62)	+	+
52	**H14**	F	56	Hannover	HLA-A: *03, *24 HLA-B: *08, *51	+	+
53	**H15**	F	51	Hannover	HLA-A: *02 HLA-B: *40(60), *40(61)	+	+
54	**H16**	F	59	Hannover	HLA-A: *02, *11 HLA-B: *07, *51	-	+
55	**H17**	M	37	Hannover	HLA-A: *03, *30 HLA-B: *13, *51	+	+
56	**H18**	N/A	N/A	Hannover	HLA-A: *03, *68 HLA-B: *35, *53	-	+
57	**H19**	N/A	N/A	Hannover	HLA-A: *29, *68 HLA-B: *15(71), *47	+	+
58	**H20**	F	56	Hannover	HLA-A: *03, *11 HLA-B: *27, *52	±[a]	+
59	**H21**	M	41	Hannover	HLA-A: *02, *32 HLA-B: *07, *18	+	+
60	**H22**	N/A	N/A	Hannover	HLA-A: *02, *03 HLA-B: *07, *55	+	+

61	**H23**	M	55	Hannover	HLA-A: *01, *02 HLA-B: *41, *44	+	+
62	**H24**	M	54	Hannover	HLA-A: *02, *32 HLA-B: *07, *18	+	+
63	**H25**	M	59	Hannover	HLA-A: *03, *24 HLA-B: *35, *38	+	+
64	**H26**	N/A	N/A	Hannover	HLA-A: *11, *69 HLA-B: *35, *52	+	+
65	**H27**	M	43	Hannover	HLA-A: *01, *02 HLA-B: *07, *08	+	+
66	**H28**	M	60	Hannover	HLA-A: *01, *68 HLA-B: *35, *55	+	+
67	**H29**	N/A	N/A	Hannover	HLA-A: *01, *02 HLA-B: *51	+	+
68	**H30**	N/A	N/A	Hannover	HLA-A: *02, *32 HLA-B: *14(64), *18	+	+
69	**H31**	N/A	N/A	Hannover	HLA-A: *02, *68 HLA-B: *07, *35	+	+
70	**H32**	N/A	N/A	Hannover	HLA-A: *02, *24 HLA-B: *50, *51	+	+
71	**H33**	M	44	Hannover	HLA-A: *02, *03 HLA-B: *35, *51	+	+
72	**H34**	N/A	31	Hannover	HLA-A: *02, *23 HLA-B: *35, *44	+	+
73	**H35**	M	37	Hannover	HLA-A: *02, *25 HLA-B: *18, *39	+	+
74	**H36**	M	61	Hannover	HLA-A: *02, *24 HLA-B: *27, *50	±[a]	+
75	**H37**	N/A	39	Hannover	HLA-A: *01, *02 HLA-B: *40(61), *57	+	+
76	**H38**	M	53	Hannover	HLA-A: *02, *26(08) HLA-B: *39, *51	+	+
77	**H39**	F	47	Hannover	HLA-A: *01 HLA-B: *08	-	+
78	**H40**	M	36	Hannover	HLA-A: *02, *11 HLA-B: *08, *44	+	+
79	**H41**	F	38	Hannover	HLA-A: *01, *02(05) HLA-B: *35, *51	+	+
80	**H42**	F	61	Hannover	HLA-A: *01, *02 HLA-B: *35, *51	-	+
81	**It01**	F	37	Italy	HLA-A: *02, *03 HLA-B: *13, *35	+	+
82	**It02**	M	64	Italy	HLA-A: *02, HLA-B: *35, *52	+	+
83	**It03**	M	59	Italy	HLA-A: *01, *68 HLA-B: *44, *57	+	+
84	**It04**	M	28	Italy	HLA-A: *03, *25 HLA-B: *18, *50	+	+
85	**It05**	M	44	Italy	HLA-A: *02, *68 HLA-B: *18, *37	+	+
86	**It06**	M	60	Italy	HLA-A: *26, *30 HLA-B: *27, *57	+	+
87	**It07**	M	43	Italy	HLA-A: *11, *24 HLA-B: *07, *44	+	+
88	**It08**	M	69	Italy	HLA-A: *02, *30 HLA-B: *18, *58	+	+
89	**It09**	F	53	Italy	HLA-A: *01, HLA-B: *35, *37	+	+
90	**It10**	M	41	Italy	HLA-A: *01, HLA-B: *35, *37	+	+
91	**It11**	M	58	Italy	HLA-A: *32, *68 HLA-B: *35	+	+
92	**It12**	M	43	Italy	HLA-A: *01, *30 HLA-B: *13, *50	+	+
93	**It13**	F	53	Italy	HLA-A: *11, *68 HLA-B: *35, *39	+	+
94	**It14**	M	58	Italy	HLA-A: *23, *30 HLA-B: *42, *49	+	+
95	**It15**	F	68	Italy	Not determined	-	+
96	**It16**	M	48	Italy	HLA-A: *01, *24 HLA-B: *49, *51	+	+
97	**It17**	M	58	Italy	HLA-A: *01, *11 HLA-B:*08, *15(75)	+	+
98	**It18**	M	65	Italy	Not determined	+	+
99	**It19**	M	27	Italy	Not determined	+	+
100	**It20**	M	30	Italy	HLA-A: *01, *02 HLA-B: *50, *51	+	+
101	**T01**	F	63	Tehran	Not determined	+	+
102	**T02**	M	71	Tehran	HLA-A: *01, *24 HLA-B: *44, *57	+	+
103	**T03**	F	64	Tehran	HLA-A: *26, *32 HLA-B: *35, *38	+	+
104	**T04**	M	54	Tehran	HLA-A: *01, *24 HLA-B: *35, *49	+	+

105	**T05**	F	53	Tehran	HLA-A: *32, *33 HLA-B: *35, *38	-	+
106	**T06**	M	50	Tehran	Not determined	+	+
107	**T07**	M	N/A	Tehran	HLA-A: *26, *33 HLA-B: *35, *51	+	+
108	**T08**	M	53	Tehran	HLA-A: *11, *33 HLA-B: *52	+	+
109	**T09**	M	45	Tehran	Not determined	-	+
110	**T10**	N/A	54	Tehran	HLA-A: *02, *24 HLA-B: *40(61), *50	+	+
111	**T11**	M	34	Tehran	HLA-A: *03, *24 HLA-B: *35, *50	+	+
112	**T12**	M	49	Tehran	HLA-A: *24, *33 HLA-B: *14(65), *35	+	+
113	**T13**	M	46	Tehran	Not determined	-	+
114	**T14**	N/A	68	Tehran	Not determined	-	+
115	**T15**	F	56	Tehran	HLA-A: *01 HLA-B: *49, *51	+	+
116	**T16**	M	26	Tehran	HLA-A: *01, *11 HLA-B: *44, *51	+	+
117	**T17**	M	45	Tehran	Not determined	-	+
118	**T18**	N/A	N/A	Tehran	Not determined	+	+
119	**T19**	N/A	N/A	Tehran	Not determined	+	+
120	**T20**	M	37	Tehran	Not determined	+	+
121	**T21**	N/A	N/A	Tehran	Not determined	+	+
122	**B01**	F	42	Barcelona	Not determined	+	+
123	**B02**	M	51	Barcelona	HLA-A: *26, *30 HLA-B: *14(64), *41	+	+
124	**B03**	M	52	Barcelona	HLA-A: *03, *33 HLA-B: *08, *14(65)	+	+
125	**B04**	M	45	Barcelona	HLA-A: *02, *30 HLA-B: *13, *44	+	+
126	**B05**	M	45	Barcelona	HLA-A: *24 HLA-B: *35, *44	+	+
127	**B06**	M	46	Barcelona	HLA-A: *03, *11 HLA-B: *07, *35	+	+
128	**B07**	F	22	Barcelona	Not determined	-	+
129	**B08**	M	48	Barcelona	HLA-A: *01, *30 HLA-B: *37, *44	+	+
130	**B09**	M	36	Barcelona	HLA-A: *24, *32 HLA-B: *35, *51	+	+
131	**B10**	M	48	Barcelona	HLA-A: *01, *03 HLA-B: *35, *57	+	+
132	**B11**	F	45	Barcelona	HLA-A: *01, *11 HLA-B: *35, *40	-	+
133	**B12**	M	44	Barcelona	HLA-A: *02, *33 HLA-B: *14(65), *49	+	+
134	**B13**	M	63	Barcelona	HLA-A: *01, *02 HLA-B: *15(62), *56	+	+
135	**B14**	M	50	Barcelona	HLA-A: *30, *31 HLA-B: *18, *44	+	+
136	**B15**	M	46	Barcelona	HLA-A: *02, *24 HLA-B: *44, *53	+	+
137	**B16**	M	52	Barcelona	HLA-A: *02, *23 HLA-B: *49, *50	+	+
138	**B17**	F	37	Barcelona	HLA-A: *00, *00 HLA-B: *18, *27	+	+
139	**M01**	N/A	30	Munich	HLA-A: *02, *68 HLA-B: *13, *35	+	+
140	**M02**	N/A	53	Munich	HLA-A: *02, *11 HLA-B: *35, *39	+	+
141	**M03**	N/A	N/A	Munich	HLA-A: *02, *29 HLA-B: *27, *45	+	+
142	**M04**	N/A	N/A	Munich	HLA-A: *02, *11 HLA-B: *35, *39	+	+

[a] some patients lost the HDV RNA in the follow-up and are indicated with ±

3.1.2. HDV sequence analysis

The complete coding region for the L-HD Ag from each studied subject was amplified and sequenced by direct sequencing technology. The forward and reverse sequences of each isolate/time-point were aligned and verified. Verified sequences were submitted at a local

database using Geneious program version 7.0.6 (Biomatters, Auckland, New Zealand). Overall, molecular analysis at nucleotide and amino acid levels was performed on 116 verified sequences from this study. An extremely high GC content was observed within the HD Ag ORF (58% in average). The difference between the maximum GC content, 60.6% (isolate T21) and the minimum, 56.9% (isolate B10), was only 3.7% which is quite small. Further analysis of sequences showed that Guanosine (G) with a total number of 26,333 (35.2%) had the highest frequency and Thymine (T) with overall number of 9,239 (12.3%) was the lowest frequent nucleotide. Because of single nucleotide polymorphism at some sites, other IUPAC codes (See Appendix III), reflecting the ambiguity at those sites, were also applied in construction of the final consensus. Regarding these hot spots, R (representing A or G) and Y (for C or T) were the most used codes at 48 and 14 sites, respectively (Table 3.2).

Table 3.2 Frequency of nucleotides of the studied isolates.

Single-letter Code	Nucleotide/s	Frequency	Percentage
A	A	21,899	29.3%
C	C	17,284	23.1%
G	G	26,333	35.2%
T	T	9,239	12.3%
R	A or G	48	0.1%
Y	C or T	14	0.0%
W	A or T	5	0.0%
S	C or G	1	0.0%
M	A or C	1	0.0%
K	G or T	2	0.0%

A homogeneous length of 645 nt, encoding for 214 aa, was obtained from almost all sequences. Interestingly, one codon insertion was observed in two isolates, E02 and T20, so that, these two isolates presented an ORF with 648 nt, encoding 215 aa, which is rather unique not only between HDV genotype 1 isolates, but also among all other types.

As explained in the introduction section, HDV genome has only one ORF which benefits from two stop codons in the same frame. While the second stop codon (TGA), corresponding to the L-HD Ag, was observed in 100% of isolates, the first one (TAG), at the amber/W site which is associated with the termination of S-HD Ag translation, was detected in 91% of isolates (at the direct sequencing level). Although in over 90% of isolates a stop codon was reported at the amber/W site, a closer look at the chromatograph data shows that there is an obvious ambiguity at this site (Figure 3.1).

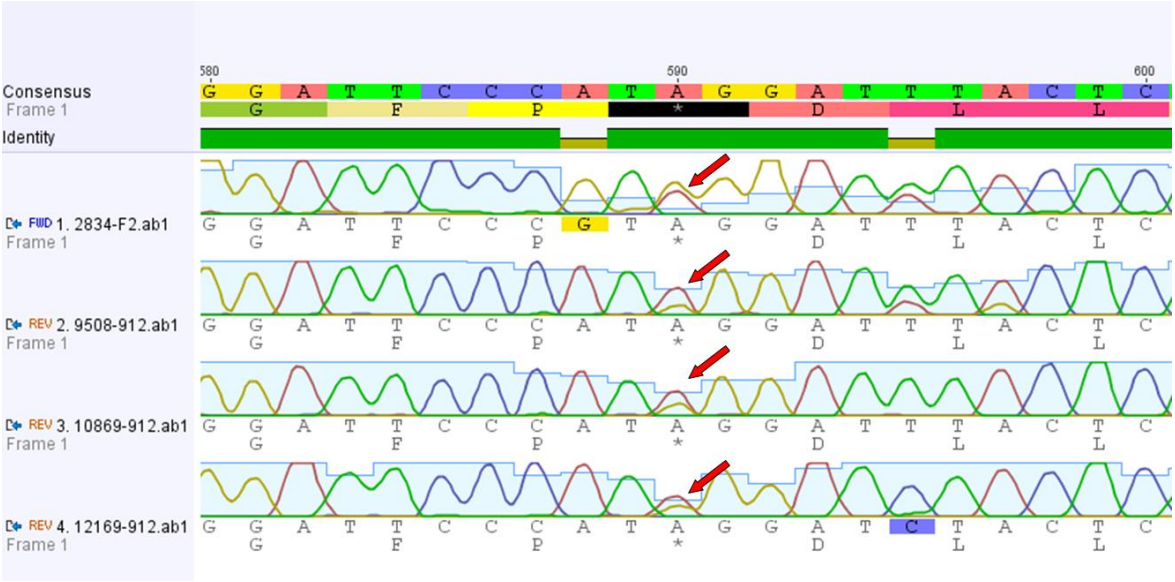


Figure 3.1 Sequence alignment at the amber/W site of 4 random isolates from this study.

To evaluate the variability of L-HD Ag from the studied isolates, the percentage of changes throughout the whole ORF of 214 aa of all isolates was calculated. Figure 3.2 demonstrates the frequency of changes in different positions on HD Ag at the amino acid level. While amino acid residues at 91 (out of 214) positions showed no variability (100% conserved) throughout all isolates, 17 residues were the most variable spots on HD Ag (with > 50% divergence). It is important to note that Figure 3.2 demonstrates residues of L-HD Ag with cumulative polymorphisms based on all isolates studied in this project. This analysis reflects past actions of fixation of substituted residues within the open reading frame among all studied individuals. In order to study the current polymorphism, variation among sequences sampled from individual subjects must be studied either by deep sequencing or in a longitudinal manner. These data are presented in the following sections as well.

A pairwise comparison of all studied isolates at the amino acid level, as well as, nucleotide level, indicated that the minimum percentage of identity was between the isolate T20 and M03 with only 82.6% identity (163 identical residues out of 214aa or 535 nucleotides out of 645), whereas, the maximum identity was observed between M02 and M04 with 99.9% identity and 212 identical residues (644 identical nucleotides). The next most identical isolates belonged to patients E12 and E23 with 99.5% identity and 212 identical residues. Notably, patients E12 and E23 were not only from the same center in Essen, Germany, but also they are both members of one family which may explain their highly identical isolates.

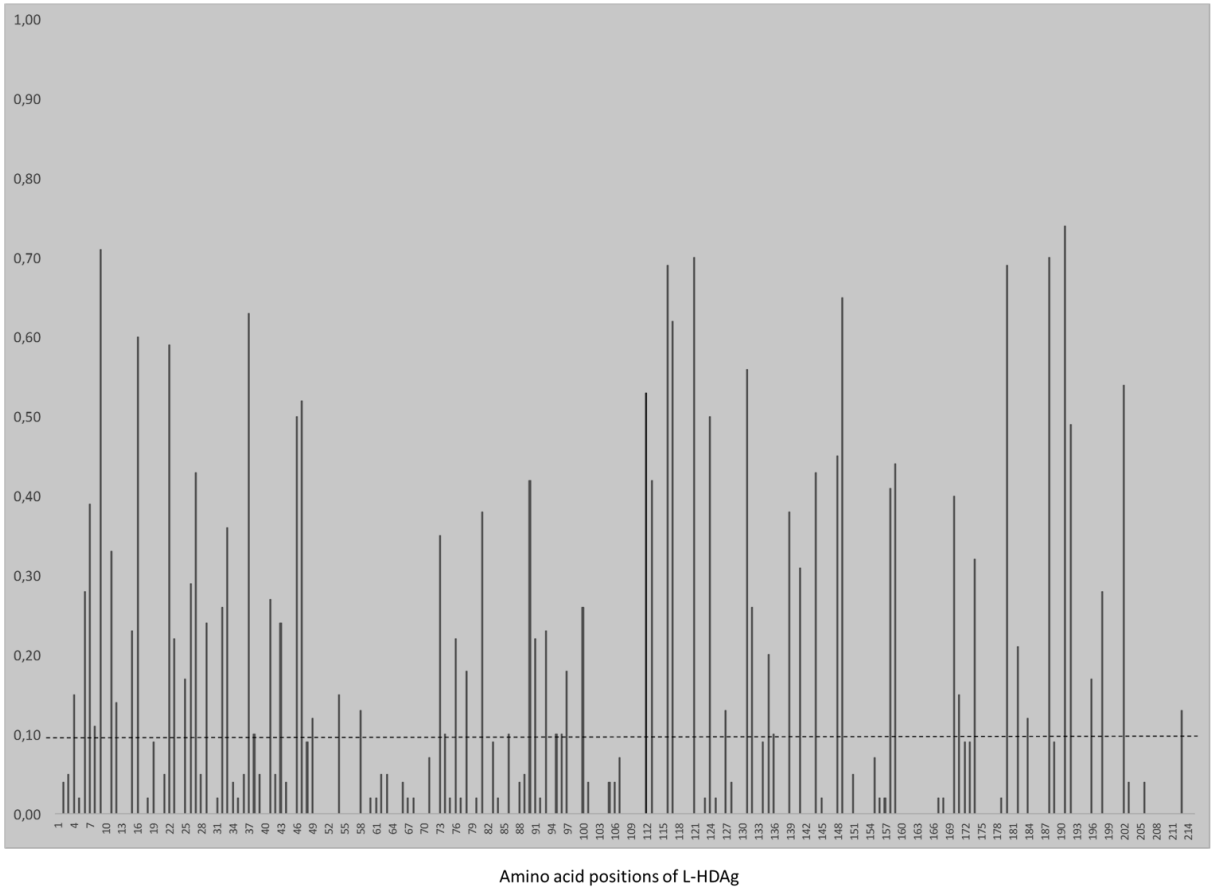


Figure 3.2 Frequencies and location of sequence changes of HD Ag.

Variability is demonstrated at the amino acid level within the whole open reading frame for all 214 aa of 120 isolates of this study. The residues: 1, 10, 13, 14, 17, 20, 24, 30, 40, 45, 50, 51, 52, 53, 55, 56, 57, 59, 64, 65, 69, 70, 72, 79, 82, 85, 87, 94, 98, 99, 102, 103, 104, 108, 109, 110, 111, 114, 115, 118, 119, 120, 122, 126, 129, 130, 133, 137, 138, 140, 142, 143, 146, 147, 150, 152, 153, 154, 160, 161, 162, 163, 164, 165, 166, 169, 175, 176, 177, 178, 181, 183, 185, 186, 187, 190, 193, 194, 195, 197, 199, 200, 201, 204, 205, 207, 208, 209, 210, 211, 212, 214 were conserved, whereas 17 residues (191, 9, 188, 116, 121, 180, 149, 37, 117, 16, 22, 131, 202, 112, 47, 124 and 192) were highly variable.

Further molecular analysis of the sequences at amino acid level indicated that Glycine (Gly; G) and Arginine (Arg; R) with 12.5% and 12%, respectively, were the most prevalent amino acids, whereas Tyrosine (Tyr, Y) with a total number of 4 (0.01%) was the least prevalent amino acid residue. This phenomenon may be explained by the high GC-content nature of HDV genome, because, as shown in Table 3.3, Glycine (G) and Arginine (R) have the most GC-rich codons.

Table 3.3 Frequency of amino acids and their codons of the studied sequences.

Codon	AA	% of AA	Freq	Total Freq	Total %AA
GCA	A	7.2%	68		
GCC	A	47.9%	454		
GCG	A	32.3%	306		
GCT	A	12.7%	120	948	3.8%
TGC	C	0.0%	0		
TGT	C	100.0%	116	116	0.5%
GAC	D	59.1%	871		
GAT	D	40.9%	602	1473	5.9%
GAA	E	37.1%	1,067		
GAG	E	62.9%	1,807	2,874	11.6%
TTC	F	87.8%	610		
TTT	F	12.2%	85	695	2.8%
GGA	G	51.1%	1,594		
GGC	G	14.3%	447		
GGG	G	27.1%	845		
GGT	G	7.4%	231	3,117	12.5%
CAC	H	92.7%	152		
CAT	H	7.3%	12	164	0.7%
ATA	I	29.4%	182		
ATC	I	47.9%	296		
ATT	I	22.7%	140	618	2.5%
AAA	K	16.4%	433		
AAG	K	83.6%	2,206	2,639	10.6%
CTA	L	4.2%	78		
CTC	L	56.4%	1,038		
CTG	L	17.8%	328		
CTT	L	9.1%	168		
TTA	L	5.4%	99		
TTG	L	7.0%	128	1,839	7.4%
ATG	M	100.0%	348	348	1.4%
AAC	N	75.7%	541		
AAT	N	24.3%	174	715	2.9%

CCA	P	10.7%	221		
CCC	P	50.0%	1,037		
CCG	P	22.3%	463		
CCT	P	17.1%	354	2,075	8.3%
CAA	Q	16.2%	144		
CAG	Q	83.8%	747	891	3.6%
AGA	R	23.7%	708		
AGG	R	44.9%	1,341		
CGA	R	13.9%	416		
CGC	R	4.6%	136		
CGG	R	12.3%	366		
CGT	R	0.6%	18	2,985	12.0%
AGC	S	32.6%	457		
AGT	S	8.9%	125		
TCA	S	0.6%	9		
TCC	S	30.7%	430		
TCG	S	17.9%	251		
TCT	S	9.3%	130	1,402	5.6%
ACA	T	10.4%	66		
ACC	T	60.3%	381		
ACG	T	16.6%	105		
ACT	T	12.7%	80	632	2.5%
GTA	V	22.7%	199		
GTC	V	29.4%	258		
GTG	V	29.6%	260		
GTT	V	18.2%	160	877	3.5%
TGG	W	100.0%	237	237	1.0%
TAC	Y	75.0%	3		
TAT	Y	25.0%	1	4	
TAA	*	0.0%	0		
TAG	*	47.7%	106		
TGA	*	52.3%	116	222	0.9%

3.1.3. Developing a specialized database for HDV genome

To facilitate molecular analysis of HDV isolates a comprehensive database was developed. This database consists of sequences from all thus far studied and submitted HDV isolates as well as the studied isolates from this cohort. Beside nucleotide and amino acid sequences of each isolate submitted in this database, additional information including accession number, origin and date of collection/publication is also provided. The database is going to be accessed globally online and will offer generic and specific tools, *e.g.*, genotype identifier, phylogenetic tree maker, primer compatibility checker and siRNA target advisor. This database could provide opportunities for the discovery of new isolates and would be a reliable platform for comprehensive analysis of HDV polymorphism. Local access to database including extra features would be possible through institute of virology of Technical University of Munich.

3.1.4. Evolutionary distance analysis between the isolates of HDV-1 and other types

To evaluate the distribution of evolutionary distances within HDV genotypes, 170 isolates from genotype 1 to 8 including the ones from this study (presumably genotype one) were compared in a pairwise manner. The total number of 13695 comparisons was made. Interestingly, the distribution of distances was confined to three separate and non-overlapping groups (Figure 3.3). This calculation implies that there may be a need to define some subtypes which are closer than the types but not as close as isolates within each subtype. The most divergent group consists of HDV-3 isolates with average evolutionary distance of 0.38 (Figure 3.3) from other genotypes. Most of other types show an average distance of 0.28 from each other (excluding HDV-3); these isolates belong to the group in the middle. Finally, the isolates within each type or clade show the lowest diversity and drop into the first group. Obviously, the low frequency of the most divergent group is due to the very low number of available isolates from HDV-3. In order to avoid selection bias, the number of isolates from HDV-1 included in this analysis was intentionally reduced. HDV-1 is the majority of available sequences because first of all, it is the most prevalent genotype in the world and, secondly, a high average number of research projects and accordingly HDV studies are carried out in western countries (where HDV-1 is prevalent) and not in South America (with HDV-3 isolates).

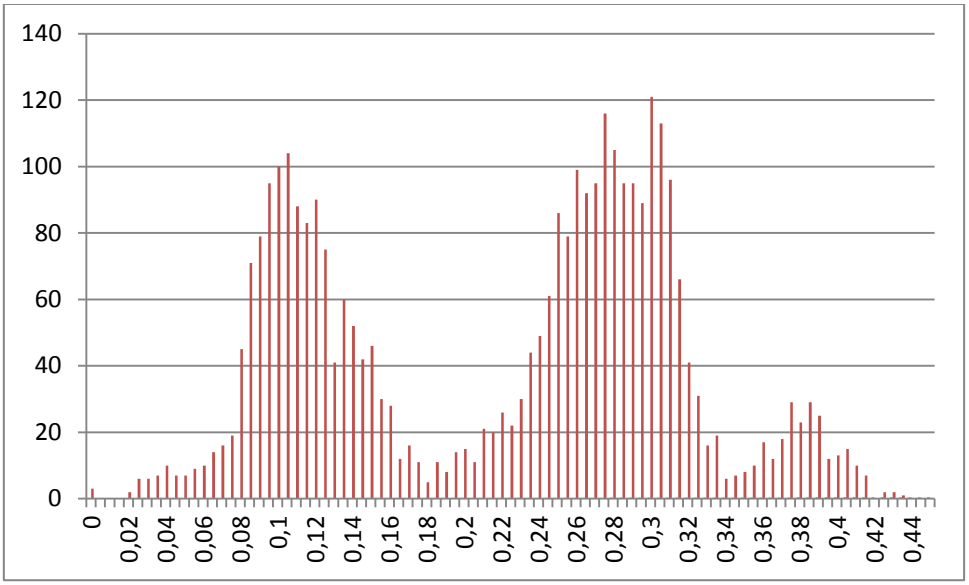
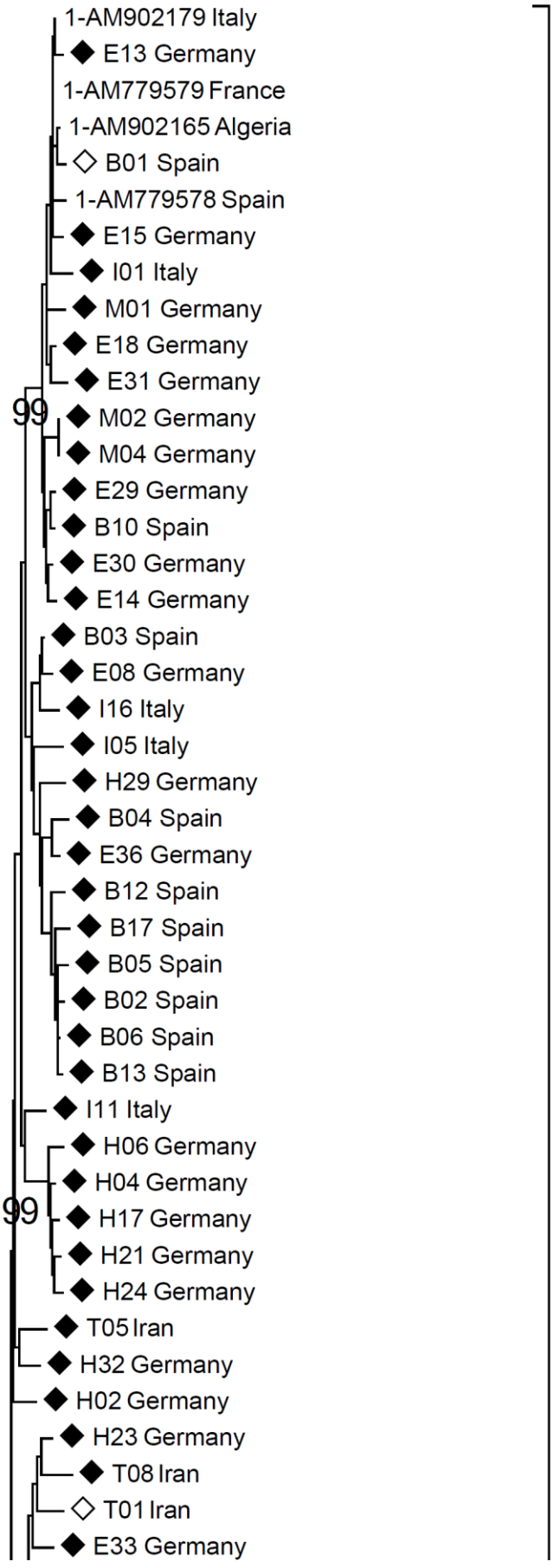
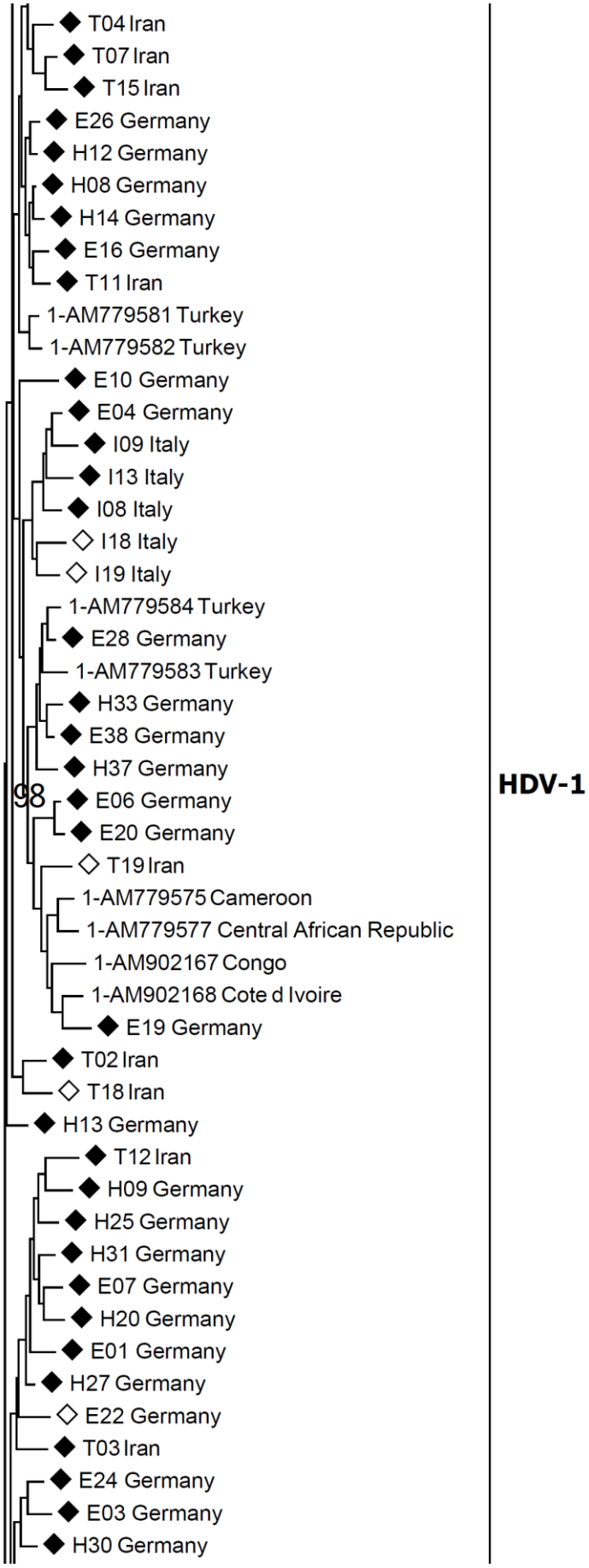


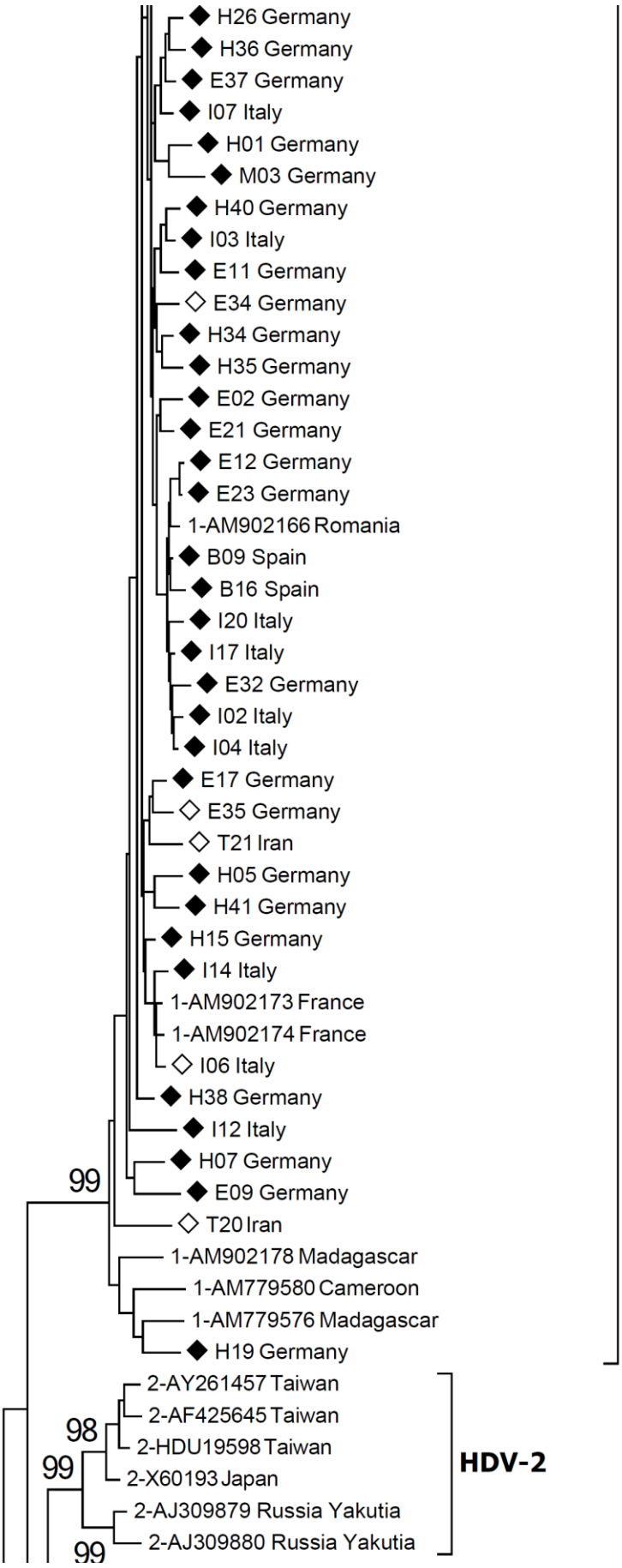
Figure 3.3 Distribution of evolutionary distances between HDV isolates.

3.1.5. Phylogenetic analysis to identify genotypes of all studied isolates

Based on the aims of this study, only isolates from HDV genotype 1 (the most prevalent genotype) were allowed to be included. To investigate whether all these isolates meet this main inclusion criterion, phylogenetic analysis was performed using genetic distance and maximum-likelihood (ML) phylogenetic reconstructions. For these analyses, the sequence corresponding to the full-length L-HD Ag was used. A total number of 116 isolates of this cohort were included. Eighteen full-length L-HD Ag sequences of HDV-1 and 36 sequences corresponding to HDV-2 to HDV-8 from diverse geographical regions were included as reference sequences in the phylogenetic analysis (Figure 3.4).


1-AM902179 Italy
E13 Germany
1-AM779579 France
1-AM902165 Algeria
B01 Spain
1-AM779578 Spain
E15 Germany
I01 Italy
M01 Germany
E18 Germany
E31 Germany
99
M02 Germany
M04 Germany
E29 Germany
B10 Spain
E30 Germany
E14 Germany
B03 Spain
E08 Germany
I16 Italy
I05 Italy
H29 Germany
B04 Spain
E36 Germany
B12 Spain
B17 Spain
B05 Spain
B02 Spain
B06 Spain
B13 Spain
I11 Italy
H06 Germany
H04 Germany
99
H17 Germany
H21 Germany
H24 Germany
T05 Iran
H32 Germany
H02 Germany
H23 Germany
T08 Iran
T01 Iran
E33 Germany


T04 Iran
T07 Iran
T15 Iran
E26 Germany
H12 Germany
H08 Germany
H14 Germany
E16 Germany
T11 Iran
1-AM779581 Turkey
1-AM779582 Turkey
E10 Germany
E04 Germany
I09 Italy
I13 Italy
I08 Italy
I18 Italy
I19 Italy
1-AM779584 Turkey
E28 Germany
1-AM779583 Turkey
H33 Germany
E38 Germany
H37 Germany
98
E06 Germany
E20 Germany
T19 Iran
1-AM779575 Cameroon
1-AM779577 Central African Republic
1-AM902167 Congo
1-AM902168 Cote d Ivoire
E19 Germany
T02 Iran
T18 Iran
H13 Germany
T12 Iran
H09 Germany
H25 Germany
H31 Germany
E07 Germany
H20 Germany
E01 Germany
H27 Germany
E22 Germany
T03 Iran
E24 Germany
E03 Germany
H30 Germany
HDV-1


H26 Germany
H36 Germany
E37 Germany
I07 Italy
H01 Germany
M03 Germany
H40 Germany
I03 Italy
E11 Germany
E34 Germany
H34 Germany
H35 Germany
E02 Germany
E21 Germany
E12 Germany
E23 Germany
1-AM902166 Romania
B09 Spain
B16 Spain
I20 Italy
I17 Italy
E32 Germany
I02 Italy
I04 Italy
E17 Germany
E35 Germany
T21 Iran
H05 Germany
H41 Germany
H15 Germany
I14 Italy
1-AM902173 France
1-AM902174 France
I06 Italy
H38 Germany
I12 Italy
H07 Germany
99
E09 Germany
T20 Iran
1-AM902178 Madagascar
1-AM779580 Cameroon
1-AM779576 Madagascar
H19 Germany
2-AY261457 Taiwan
2-AF425645 Taiwan
98
2-HDU19598 Taiwan
99
2-X60193 Japan
HDV-2
2-AJ309879 Russia Yakutia
2-AJ309880 Russia Yakutia
99

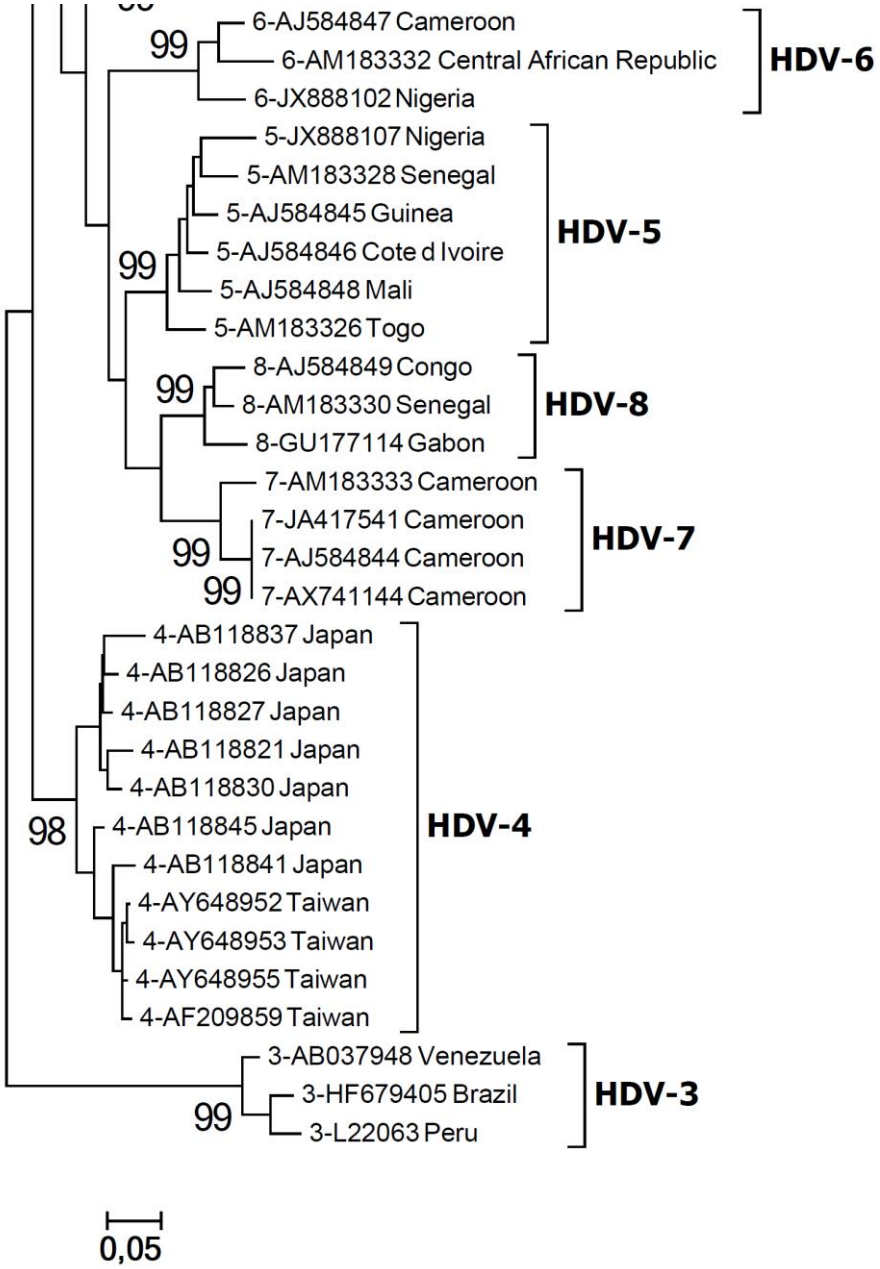


Figure 3.4 Reconstruction of phylogenetic tree obtained from distance analysis.

Maximum Likelihood phylogenetic tree based on the whole large delta antigen sequences was reconstructed. 116 strains from this study are marked by ◆ or ◇(12 isolates are indicated by ◇ are from the patients who were not HLA typed) and 54 reference sequences from GenBank representing HDV genotype 1 to 8. For the reference sequences of HDV 1-8, the genotypes, the accession number and their origins (collection places) are indicated in the name of these isolates. The numbers at the nodes represent the percentages of the bootstrap support values (1,000 replicates).

The phylogenetic and evolutionary distance analyses confirmed that all of the studied isolates in this project branched together under HDV genotype 1 (the most prevalent genotype in the world). Although all isolates from this study branched together with the reference sequences of HDV-1, some of them reproducibly clustered together indicating a closer evolutionary relations between some of the isolates. Based on the evolutionary distances as well as morphology of the phylogenetic tree, these clusters can be precisely determined. These clusters are strongly supported by the evolutionary distances and reflect geographical or even familial relationships of the patients. For example, Spanish cluster consists exclusively of Spanish isolates such as B12, B17, B05, B02, B6 and B13. This phenomenon may introduce a founder effect into sequence analysis. However, in this multi centric study, this may not be of great concern.

Phylogenetic analysis also indicated that one of the isolates of this study (T20) was the most divergent isolate among all isolates of HDV-1. Isolate T20 showed the lowest identity with M03 (82.6%). AM779577 and AM779575 with 82.7% and 82.9% were the next least identical isolates to T20. In conclusion, T20 with longest L-HD Ag of 215 amino acids (when compared to the prototypical isolates with 214 aa) and most diverse from other HDV-1 isolates, may be defined as a novel cluster or clade of HDV in the future.

Based on the analysis of the distribution of evolutionary distances carried out in this study on L-HD Ag, redefining of the HDV types seems to be of dire need. For example, the isolate AM779577 from Cameron and AM902178 from Madagascar with 83.6% identity branch under HDV-1; whereas, the isolate with higher identity, *e.g.*, X60193 from Japan and AJ584846 from Africa with 84.3% identity, are classified under HDV-2 and HDV-5, respectively.

3.1.6. Evolution of HDV in patients (Longitudinal studies)

In order to investigate viral evolution and mutation pattern inside the encoding region during the chronic infection, 20 patients from this study were followed for a varying time lengths, from 3 months up to 6 years (Table 3.4). The post-infection mutation rate was determined over time for each patient. Evolutionary distance was also calculated within the individual patients as well as within the whole group of patients.

Table 3.4 Patient characteristics studied longitudinally for HDV evolution at the L-HD Ag coding region.

#	Patient	Sex	Age	HLA-Types		Follow-up (months)	AA substitutions
1	E03	M	39	HLA-A: *01, *02	HLA-B: *07, *08	15	7
2	E06	M	46	HLA-A: *01, *03	HLA-B: *18, *50	9	0
3	E08	M	49	HLA-A: *02, *24	HLA-B: *35, *52	14	2
4	E10	M	73	HLA-A: *02	HLA-B: *41, *51	19	4
5	E11	M	37	HLA-A: *02, *68	HLA-B: *15 (62)	14	8
6	E12	M	18	HLA-A: *01, *24	HLA-B: *35, *58	15	4
7	E14	M	35	HLA-A: *02	HLA-B: *55, *60	11	3
8	E16	F	54	HLA-A: *02, *11	HLA-B: *51, *55	11	0
9	E17	M	48	HLA-A: *01, *11	HLA-B: *35, *51	13	1
10	E18	M	41	HLA-A: *03, *11	HLA-B: *35, *51	7	0
11	E19	M	52	HLA-A: *11, *30	HLA-B: *18, *35	16	0
12	E21	M	62	HLA-A: *02,	HLA-B: *50, *52	20	0
13	E22	M	35	Not determined		2	3
14	E23	M	23	HLA-A: *24, *31	HLA-B: *35, *51	3	3
15	E24	M	72	HLA-A: *3, *24	HLA-B: *18, *35	50	0
16	E32	M	28	HLA-A: *66, *68	HLA-B: *52, *57	14	12
17	E33	M	45	HLA-A: *24	HLA-B: *18, *51	6	4
18	E36	M	47	HLA-A: *02	HLA-B: *44	72	5
19	E37	M	40	Not determined		5	9
20	E38	F	43	HLA-A: *3, *32	HLA-B: *15, *52	3	6

A total number of 74 samples from 20 patients with chronic HDV infection were recruited retrospectively at the diagnostic section of the institute of virology, university hospital of Essen. The L-HD Ag encoding region of all 74 isolates was analyzed. Interestingly, the evolution rate of HDV varied in different patients probably due to their disease status. For

example, patients over 48 years old (n=7) showed almost no amino acid substitutions; whereas, the younger patients (n=13) demonstrated up to 12 changes at the amino acid level.

In this study, 4 sample/time-points from a patient (E32) with acute HDV infection were analyzed. Although this patient was studied for a period of less than a year, the number of amino acid substitutions was vast (Figure 3.5). This probably implies the viral evolution and consequently adaptation during the acute phase of infection. In separate studies, we were able to show how fast fixation of mutations occurs short after onset of infection (data not shown).

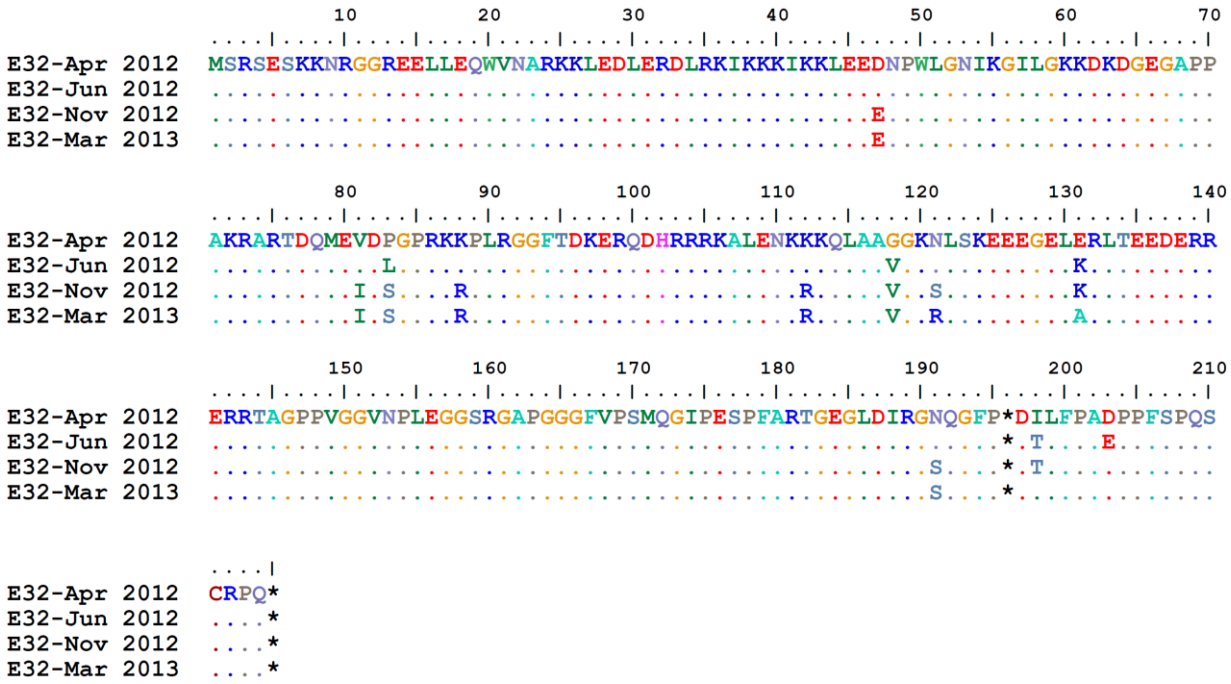


Figure 3.5 HD Ag variability of the isolated variants from patient E32 after acute infection (1 year follow-up).

In contrast with the above mentioned patient, several other patients (Table 3.4) did not show accumulation of mutations over time. For instance, as illustrated in Figure 3.6, longitudinal sequence analysis of the samples from patient E24 with HLA-A*03, *24 and HLA-B*18, *35, demonstrated no mutations when studied in a period of 50 months (over 4 years). According to the records, this patient had been already tested positive for HDV for more than 15 years. This probably indicates adaptation of HDV after long term infection in a single host.

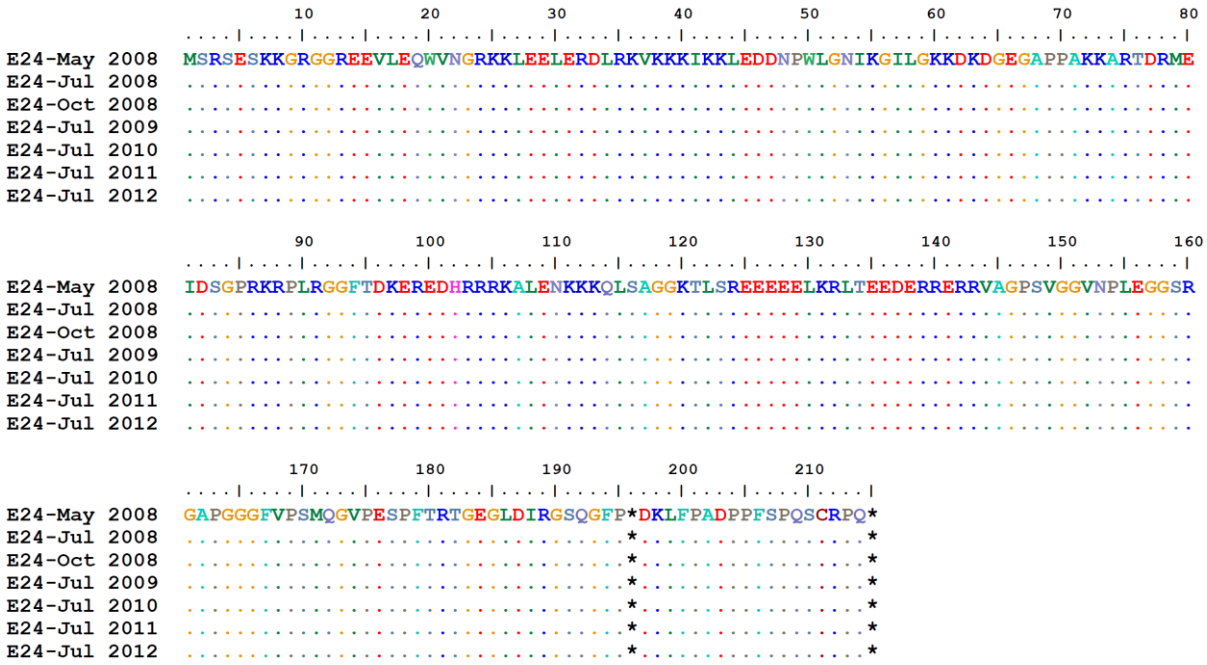


Figure 3.6 HD Ag variability of the isolated variants from patient E24 with chronic HDV infection (4 years follow-up).

Phylogenetic and evolutionary distance analyses indicated that the isolates from each patient of this longitudinal study cluster perfectly together (not shown) but two patients (E23 and E12), whose isolates formed a single cluster together first then evolved into two closely related clusters (Figure 3.7).

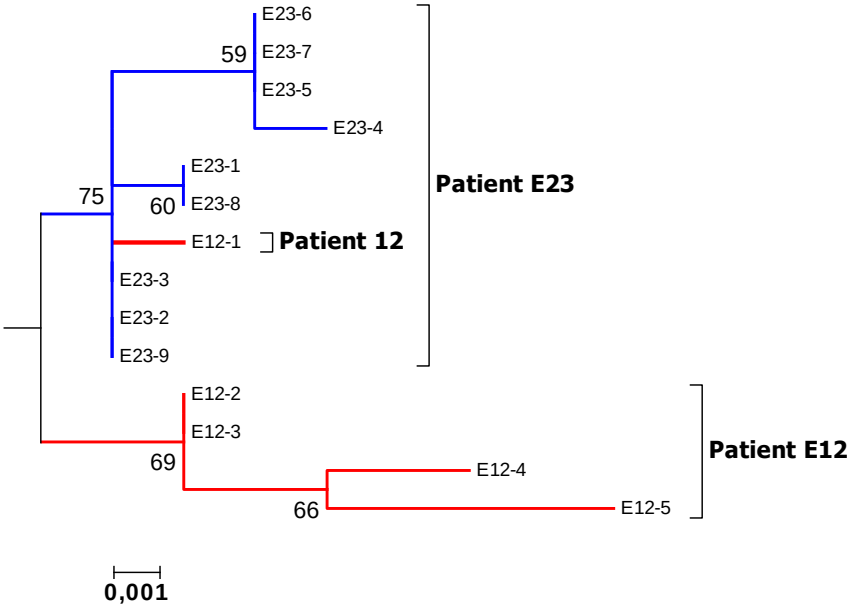


Figure 3.7 Phylogenetic tree construction for patients E23 and E12.

Branches related to isolates from patient E23 and E12 are indicated in blue and red, respectively. The sampling order of isolates is indicated after the code of each patient. This construct is obtained from the phylogenetic tree of all 74 isolates studied longitudinally.

Further information revealed that these patients are members of the same family with clinical records of a preexisting chronic HBV infection. E23 was first co-infected with HDV and then transmitted to the younger member of this family, E12.

As demonstrated in Figure 3.7, isolate E12-1 is the very first sample from patient E12 which, evolutionary, is still very close to those of patient E23. However, in later phase, isolate E12-5 which is the last tested sample from patient E12, shows the highest diversity with E12-1 isolate and other ones from the patient E23. This analysis nicely shows that patient E23 was the first infected patient and then later transmitted to the younger member. These two patients share 50% of their HLA-A and B alleles (E12: HLA-A:*01, ***24** / HLA-B:***35**, *58 and E23: HLA-A:***24**, *31 / HLA-B:***35**, *51), therefore, one might speculate that further evolution of HDV in the second patient can be the results of adaptation to the HLA alleles of the new host (*i.e.* HLA-A:*31 and HLA-B:*51) which may indicate T cell pressure restricted to these alleles. This hypothesis needs to be investigated in a separate study with larger data set.

3.1.7. Evidence of positive selection by calculation of the ratio of non-synonymous to synonymous mutations

Longitudinal studies clearly showed that there are some sorts of evolution of HDV towards adaptation. Now to address whether positive selection occurs during viral adaptation to the host, d_N/d_S ratio was calculated using all studied isolates. Essentially, when d_N/d_S ratio is greater than one, then there must have been positive selection pressure. The result of this calculation on studied isolates indicated positive selection in various codons in L-HD Ag (some are listed in Table 3.5). These results are strongly supported by statistical *p* values. For instance, codon 149 (nt 445-447) with d_N/d_S ratio of 5.7 indicated to be under positive selection pressure ($p=1 \times 10^{-6}$).

Table 3.5 List of some residues on L-HD Ag under possible positive selection.

Codon	Nucleotide Position	Codon (amino acid)	Normalised d_N/d_S	*p*-value
41	121-123	ATC (I), GTC (V), CTC (L)	2.215523593	0.03873
75	223-225	AGG (R), AAG (K)	3.048700221	0.00141
81	241-243	ATA (I), GTA (V)	2.594473455	0.0031
90	268-270	TCC (S), CTC (L), CAC (H), CCC (P), GCC (A), TTC (S)	3.812963368	0.00015
124	370-372	AAG (K), AGG (R)	4.325139341	0.00033
149	445-447	GTT (V), ATT (I), ACT (T), GGT (G), CCT (P)	5.739511247	1.19E-06

3.2. Characterization of HDV specific CD8+ T cell epitopes and the corresponding response in patients with resolved infection

In order to characterize the CTL response in HDV infected patients, it is necessary to know which parts of HDV protein (HD Ag) are targeted by these cells. There are different approaches for identification of CD8+ T cell targets (epitopes). One approach, used also in our lab, is to stimulate the PBMCs, ideally isolated from a patient with resolved HDV infection, using a library of overlapping peptides spanning the whole aa sequence of L-HD Ag. In this method, positive PBMCs will be tested further to identify the optimal epitope. Another approach is to use prediction algorithms to locate the potential epitopes and testing the binding affinities of these predicted epitopes to their corresponding HLA molecules and finally test whether they are functional epitope in patient by stimulation of PBMCs. The latter approach was mainly used for this study and the corresponding data are presented in the following sections.

3.2.1. HDV epitope prediction and MHC binding affinities

Since this study was focused on HDV-1 (the most prevalent HDV type in the world), epitope prediction was performed with respect to the HDV-1 isolates only. There are several prediction tools available online; Immune Epitope Database (IEDB) is one of the most comprehensive databases to date which offers a quite broad range of Analysis Resources including T cell epitope prediction tools and the relevant updated algorithms. IEDB offers a recommended algorithm which searches for the best available prediction method from which predictive performances for the HLA allele of interest have been observed. Selected predictors may vary based on the selected HLA allele and probably other characteristics of the peptide. IEDB recommended applies a consensus method including NetMHC, SMM, NetMHCpan and CombLib based on their availability for the molecule of interest. The peptides with the lowest percentile scores are considered as the possibly good binders to the corresponding HLA molecules. Therefore, the predicted peptides with the lowest scores (among all predicted peptides in the same group) were selected (Table 3.6). However, all these methods are *in silico* approaches which need to be supported by experimental data, *e.g.* evaluation of the binding affinities of these predicted epitopes to their corresponding HLA molecules *in vitro*. Therefore, the prediction was included for the HLA alleles for which there was an established binding assay. Another point was to include those HLA alleles with high prevalence or biological significance. Therefore, HLA-A*01, A*02, A*03, A*24, HLA-B*07 and B*27 were selected for this analysis.

Table 3.6 Predicted epitopes restricted to selected alleles from HLA class I locus A and B using two algorithms.

HLA:	Position:	Sequence:	Length	Score (IEDB)	Score SYFPEITHI
A*01	75-83	RTDQMEVDS	9	3.2	17
	94-102	FTDKERQHD	9	1.3	17
A*02	43-51	KLEEDNPWL	9	1.3	23
	143-152	RVAGPPVGGV	10	4.2	23
	198-206	ILFPSDPPF	9	0.8	16
A*03	53-61	NIKGILGKK	9	2.7	22
	89-97	PLRGGFTDK	9	2.1	28
A*24	49-57	PWLGNIKGI	9	3.9	13
	192-200	RGFPWDILF	9	1.1	13
	198-206	ILFPSDPPF	9	2.4	11
B*07	68-76	APPAKRART	9	1.9	21
	71-79	AKRARTDQM	9	3.1	12
	73-81	RARTDQMEV	9	1.1	9
	86-94	RKRPLRGGF	9	3.2	11
	100-108	QDHRRRKAL	9	2.4	12
	168-176	VPSMQGVPE	9	3.0	16
	194-202	FPWDILFPS	9	1.4	11
B*27	99-108	RRDHRRRKAL	10	0.35	24
	103-112	RRRKALENKK	10	0.20	29

Overall 18 predicted peptides with the highest ranking including 2, 3, 2, 3, 7 and 2 peptides restricted to HLA-A*01, A*02, A*03, A*24, HLA-B*07 and B*27, respectively, were chosen to be tested in an HLA binding assay. As demonstrated in Table 3.6, the selected peptides were among the top listed peptides for each HLA allele; however, this does not necessarily mean that these peptides have reached the percentile scores of well-known characterized epitopes. The selected peptides were also analyzed using SYFPEITHI, and the corresponding prediction scores are listed in the very right column of Table 3.6.

The sequences at positions 99-108 (RRDHRRRKAL) and 103-112 (RRRKALENKK) restricted by HLA-B*27 showed the best prediction scores for the selected HLA types within L-HD Ag (Table 3.6). Presumably, such prediction data retrieved and calculated according to the data bases may slightly vary due to the new updates.

Although predicted epitopes restricted by the most of common alleles did not reach an appropriate prediction score, those with the best predicted scores were selected to be tested

for a possible binding affinity. As explained in section 2.7. MHC stabilization assay was applied in order to test binding affinities of the predicted ligands listed in Table 3.6 to the corresponding HLA class I molecules. For each set of assay one positive epitope (a well-characterized and binding epitope to the corresponding allele) was used as the positive control. An irrelevant peptide was also used as the negative control. For each assay, one reaction was included without any rescue peptide (no peptide reaction) in order to test the background signal.

RTDQMEVDS and FTDKERQHD were selected for HLA-A01:01 allele, and showed the binding of 4% and 17%, respectively, when compared to the positive control (CTELKLSDY) as the 100% binder. For HLA-A*02, one of the most prevalent HLA alleles, three ligands were tested: KLEEDNPWL, RVAGPPVGGV and ILFPSDPPF achieved the average binding of 9%, 8% and 8%, respectively. A well-known CMV epitope (NLVPMVATV) restricted to HLA-A*02:01 was used as the positive control. CTELKLSDY, positive control epitope for HLA-A*01:01, was used as the irrelevant epitope (negative control) for HLA-A*02:01. NIKGILGKK and PLRGGFTDK were tested for binding to HLA-A*03:01 resulted in binding of 12 and 14%, respectively, to the HLA-A*03 molecules. These binding affinities were about the range of irrelevant epitope or when no epitope was applied; therefore, these selected epitopes for HLA-A*03 were considered as non-binding and irrelevant epitopes (at least to HLA-A*03:01). For HLA-A*24, three possible ligands were selected, PWLGNIKGI, RGFPWDILF and ILFPSDPPF, which were able to show binding affinities between 10 and 15% as compared to the positive control (LYSACFWWL, a known HLA-A*24 epitope). For HLA-B alleles, B*07 and B*27 were analyzed. A rather larger number of predicted ligands were selected for HLA-B*07 with percentile scores (ranks) of under 4 (Table 3.6). These ligands, with an average binding of 20% to HLA-B*07:02, demonstrated 2 times more affinity than that of the irrelevant ligand (LIYRRRLMK) used as the negative control (Figure 3.8). Finally, two predicted ligands for HLA-B*27, aa 99-108 (RRDHRRRKAL) and 103-112 (RRRKALENKK), were tested for binding affinities to the corresponding HLA molecule, where a well-characterized HLA-B*27-restricted epitope from HIV (KRWIILGLNK) was used as the positive control (Figure 3.8). Interestingly, RRDHRRRKAL and RRRKALENKK showed high binding affinities of 120 and 111%, respectively, as compared to the positive control. These findings were in line with the percentile scores of these to predicted ligands.

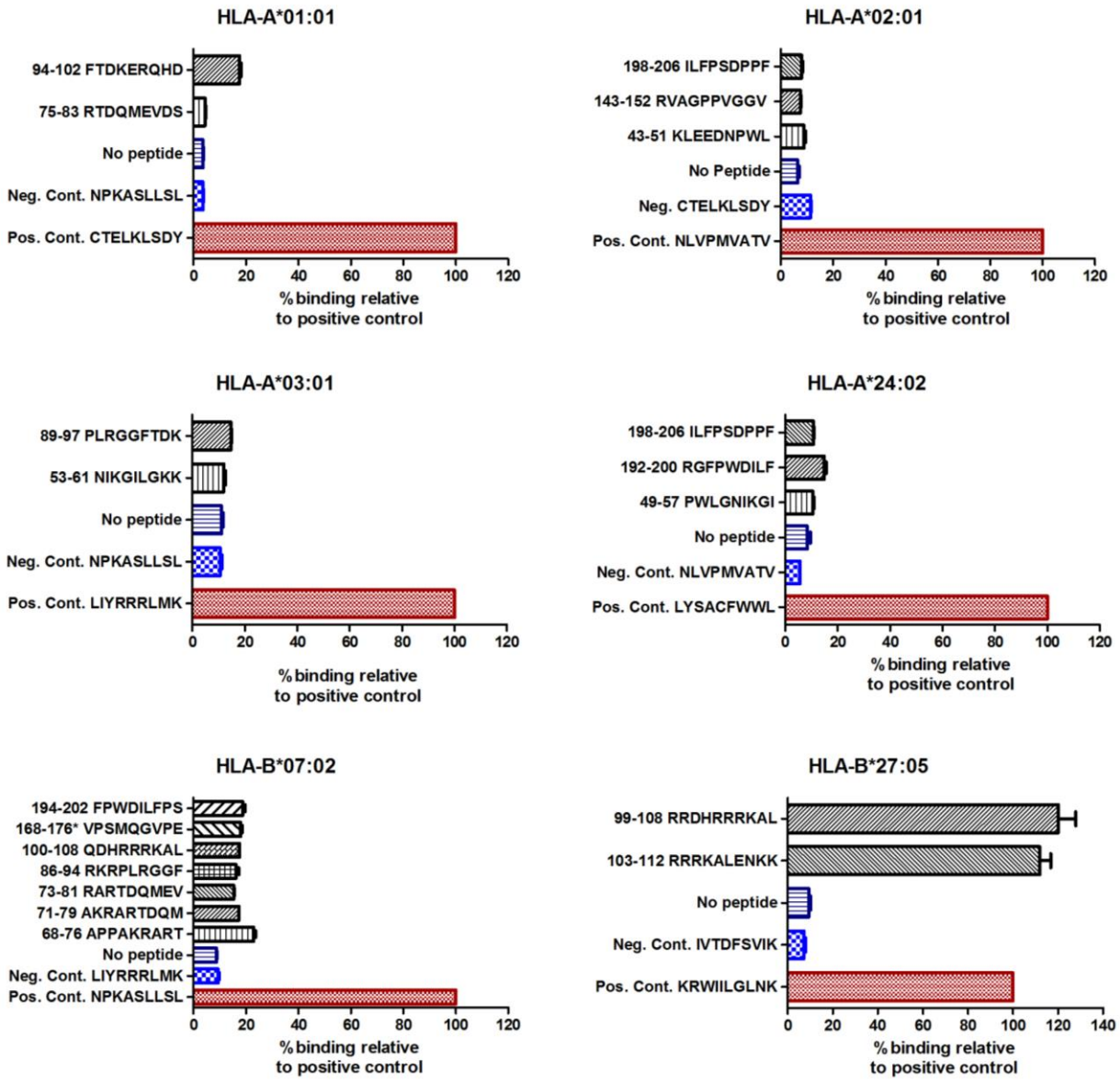


Figure 3.8 Binding affinities of predicted synthetized epitopes listed in Table 3.6.

Binding affinities of 18 candidate peptides tentatively restricted to 6 different MHC molecules. The MHC alleles include HLA-A*01 (number of tested peptide=2), HLA-A*02(n=3), HLA-A*03(n=2), HLA-A*24(n=3), HLA-B*07 (n=7) and HLA-B*27(n=2). The affinities are calculated based on UV-induced ligand exchange. A well-known epitope was used as a positive control for each assay indicated by red and an irrelevant peptide (mismatching epitope) was used as the negative control indicated by blue color. Testing peptides are labeled with their positions on the L-HD Ag followed by the amino acid sequences.

3.2.2. Detection of HDV specific CD8+ T cells in resolved HDV infection

Induction of antiviral state (*e.g.* production of IFN) upon antigenic exposure is one of the most significant functions of the T cells. This important characteristic of the lymphocytes is

used for epitope discovery. In order to identify HDV-specific CD8+ T cell epitopes, therefore, IFN-γ production was measured upon stimulation of PBMCs with HDV specific peptides. Obviously, PBMCs were isolated from a donner with resolved HDV infection. This approach was also used to confirm the findings of novel HLA-B*27 epitopes (described in previous section, 3.2.1.). With respect to the HLA-B*27, the first evidence was found in a HLA-B*27 positive patient (Patient A) with resolved HDV infection (HDV RNA negative but anti-HDV antibody positive). The PBMCs of this patient was stimulated with library of overlapping 16-mer peptides spanning L-HD Ag (kindly provided by Dr. Melanie Fiedler). In this preliminary study, HDV peptide library was divided into 8 pools of peptides (A to H) and applied for a 10-day stimulation of the isolated PBMCs from recovered patients. After re-stimulation of the expanded PBMCs on day 10 with the respective pools of HDV peptides, IFN-γ staining was performed and the percentage of IFN-γ-positive CD8+ T cells was measured in each well. Interestingly, IFN-γ-positive CD8+ T cells were detected in PBMCs from a HLA-B*27 positive patient with resolved HDV infection. In this patient, IFN-γ was induced in PBMCs stimulated with peptide pool D (Figure 3.9). This response was not observed in the same cells stimulated with other pools. In fact, the percentage of IFN-γ-positive CD8+ T cells (0.54%) was approximately 12 times higher than that in the cells stimulated with other peptide pools or with unrelated peptide control (mean 0.045%).

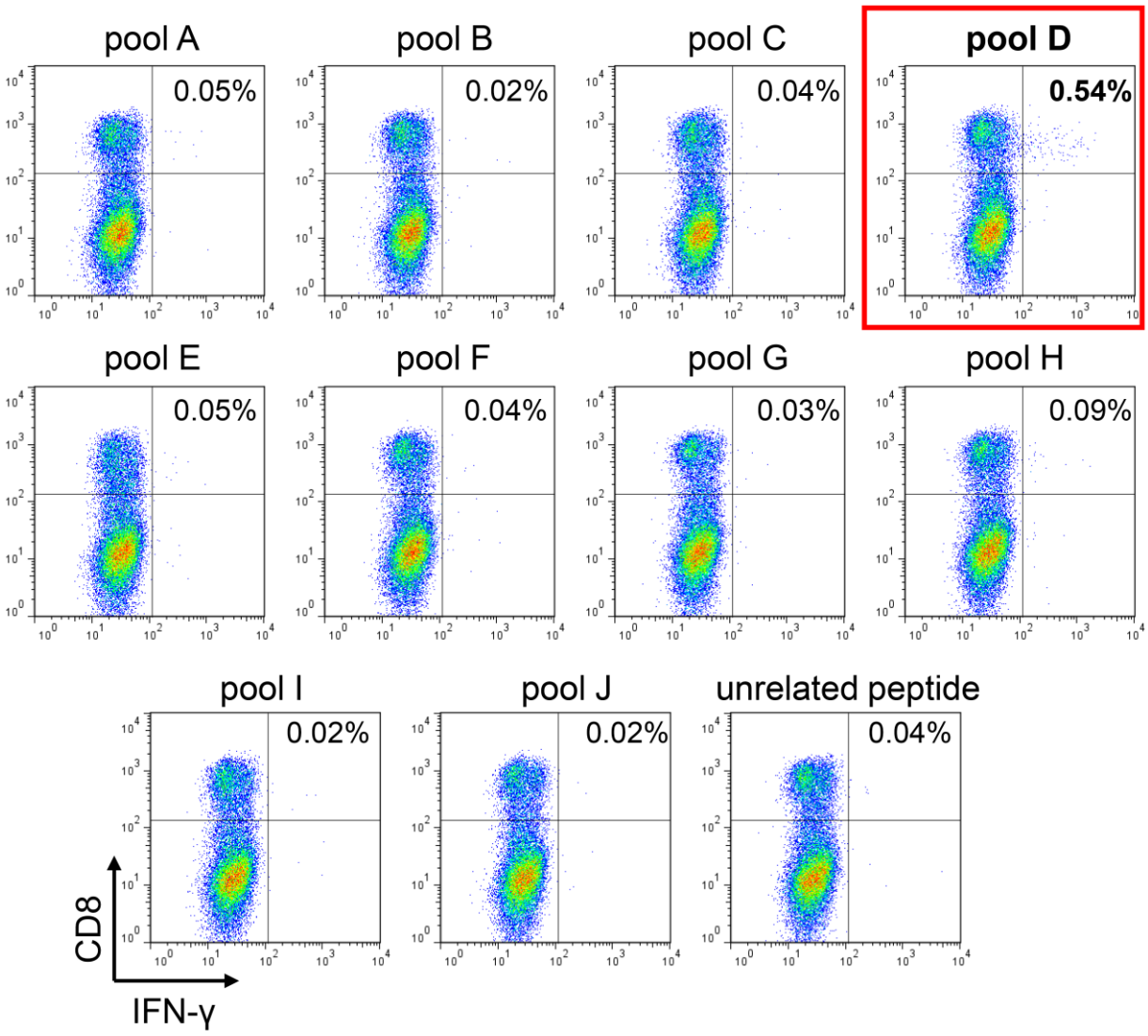


Figure 3.9 FACS staining for IFN-γ production after stimulation of PBMCs (Patient A) by overlapping 16-mer peptides listed in

Table 2.1.

PBMCs from a HLA-B*27 positive patient with resolved HDV infection were stimulated with overlapping peptides spanning the ORF of L-HD Ag with 214 amino acids. The percentage of IFN-γ+/CD8+T cells upon stimulation with peptide pools A to J is given in the upper right corner of each panel.

Next, to investigate which peptide(s) within the pool D could trigger this specific T cell response, the cultured cells were re-stimulated in the presence of single peptides from pool D. As shown in Figure 3.10, peptide D14 was able to induce a proportion of CD8+ T cells to produce IFN-γ (0.25% IFN-γ+ CD8+ T cells vs 0.01-0.02% detected for other peptides or un-re-stimulated control). Interestingly, the 16-mer peptide, D14 (98-113 ERRDHRRRKALENKK K), includes the sequences of both predicted epitopes, aa 99-108 RRDHRRRKAL and aa 103-112 RRRKALENKK, which were able to bind to the HLA-B*27 molecule with high affinities (Figure 3.8).

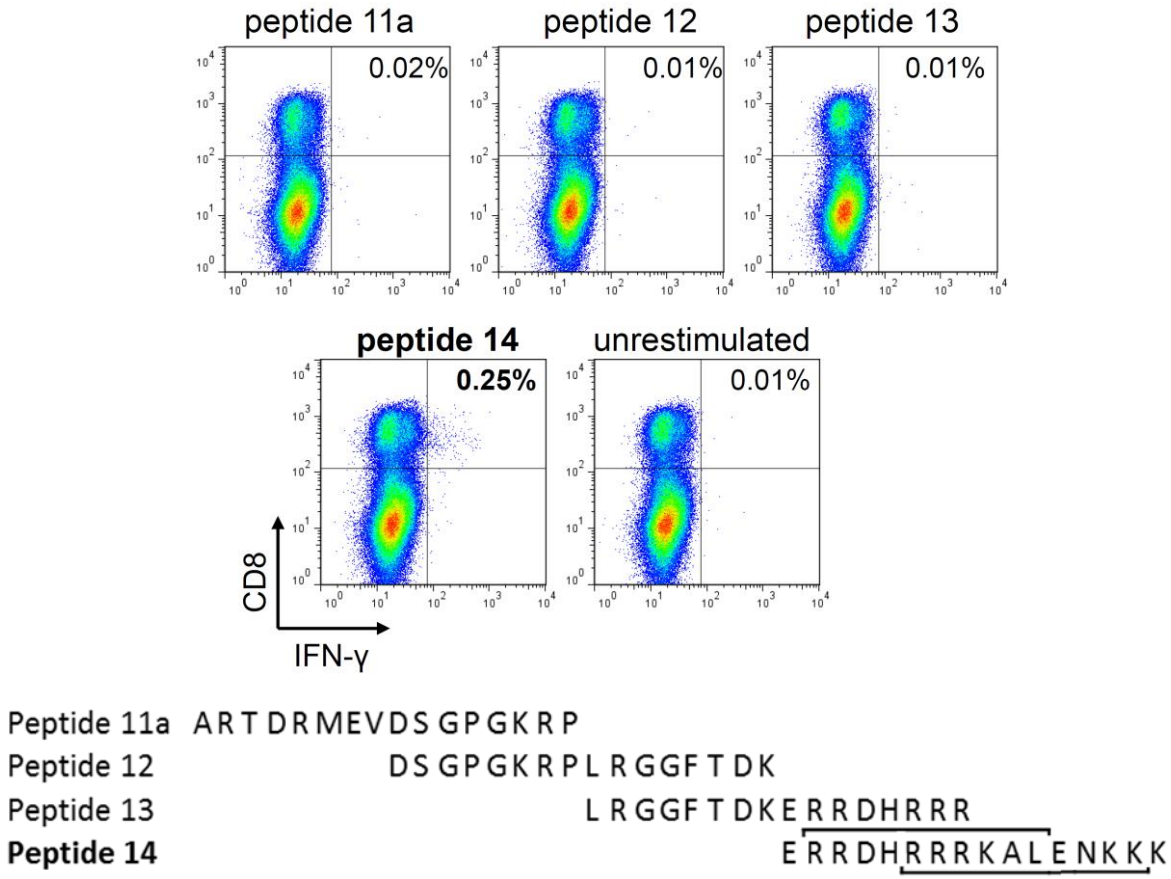


Figure 3.10 FACS staining after stimulation with single peptides.

(A) FACS staining for IFN-γ production after stimulation of PBMCs (from patient A) with single peptides from the pool D (B) Amino acid sequences of all single peptides within the pool D (peptide 11a, 12, 13 and 14) are listed and the sequence of HLA-B*27 restricted epitopes are marked with lines in peptide 14.

As demonstrated in Figure 3.10, the two predicted and binding epitopes are only present in one of the overlapping peptides (peptide D14) of this pool which indicates that the IFN-γ induction is probably restricted to, at least, one of these two epitopes. Since the PBMCs of this patient were no longer available, we were not able to pinpoint which one of these two epitopes within this 16-mer (peptide D14) was indeed the main inducer of CD8+ T cells to produce IFN-γ. It is possible that sometimes one epitope may appear in more than one peptide of a library of overlapping peptides. Obviously, this depends on how many amino acid is the library overlapped by. In which case, it would be possible to recognize the minimal epitope after detection of response upon stimulation with two peptides of the library. This was not applicable here in this study.

We, in collaboration with the university hospital of Freiburg, tested PBMCs of another HDV recovered HLA-B*27 positive patient (Patient B) in order to determine the exact minimal CD8+ T cell epitope, against which the T cell response was directed, and check the

reproducibility of the results obtained from the first patient, whose PBMCs were stimulated with overlapping peptides.

In this assay, PBMCs were isolated and cultured in the presence of peptide D14 (aa 98-113 ERRDHRRRKALENKKK), and after 10 days, the cells were re-stimulated with both predicted epitopes: aa 99-108 RRDHRRRKAL and aa 103-112 RRRKALENKK as well as the 16-mer (peptide D14). Re-stimulation of the PBMCs from patient B with 16-mer (98-113 ERRDHRRRKALENKKK) resulted in IFN-γ induction in 0.50% of the CD8+ T cells (Figure 3.11) which was consistent with the findings in patient A after stimulation with the same peptide (peptide D14). Interestingly, re-stimulation of the cells with the two predicted binding epitopes within this 16-mer revealed that indeed RRDHRRRKAL may be the HDV-specific HLA-B*27-restricted CD8+ T cell epitope targeted in the patient B. In fact, re-stimulation of the cells with the synthetic peptide RRDHRRRKAL (L-HD Ag aa 99-108) was able to induce the same percentage of CD8+ T cells to produce IFN-γ as for the 16-mer, peptide D14; however, aa 103-112 RRRKALENKK did not induce any response in this patient (patient B). This analysis indicates that despite binding capability of aa 103-112 RRRKALENKK to its corresponding HLA molecule (HLA-B*27) *in vitro*, this peptide cannot be considered as a CTL epitope at least based on the data obtained from patient B from Freiburg. The 10-mer RRDHRRRKAL, in contrast, was able to induce a specific T cell response in this HLA-B*27 positive patient who had resolved HDV infection. Therefore, the results from epitope prediction analysis, MHC binding assay, and, most importantly, functional T cell assay confirm that RRDHRRRKAL is a HDV-specific T cell epitope.

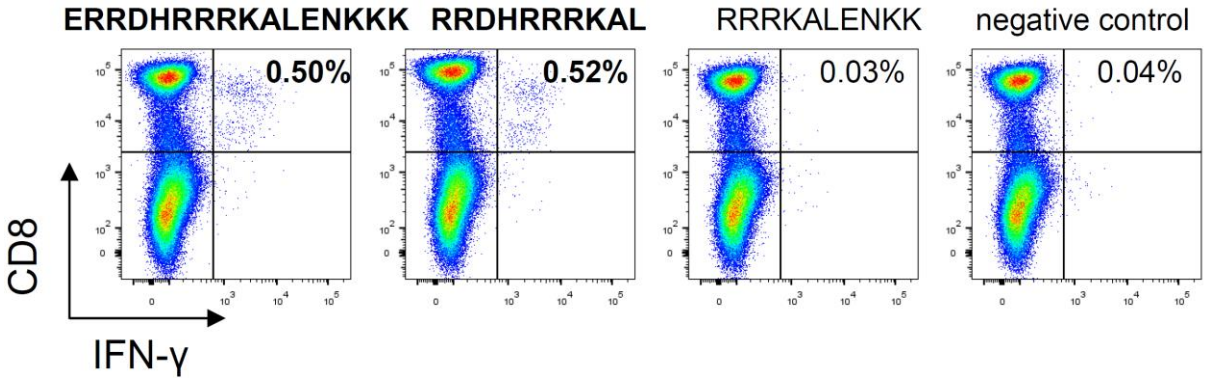


Figure 3.11 CD8 T cell assay in patient B to distinguish the exact prototype HDV-specific HLA-B*27-restricted epitope.

Flow cytometry staining of the PBMCs from Patient B (HLA-B*27 positive / HDV resolved patient) stimulated with the 16-mer (first panel from the left) and the 10-mers. The frequency of IFN-γ+/CD8+ T cell determined by ICS is indicated as percentage in right upper corner of each panel.

One of the chronically HDV infected subjects who lost HDV RNA during the follow-up was HLA-B*27 positive (Patient H36). Therefore, the PBMC sample from this subject was evaluated by T cell assay. The PBMCs applied for the T cell assay were, obviously, collected from the time-point when the patient was HDV RNA negative. According to the previous data achieved from patient B, we did not use the library of L-HD Ag anymore for this subject (patient C), and performed the assay with focus on overlapping 16-mer peptide (aa 98-113 ERRDHRRRKALENKKK) covering both predicted epitopes as well as single epitopes (aa 99-108 RRDHRRRKAL and aa 103-112 RRRKALENKK) to detect the prototype epitope in this patient. A 10-day culture and stimulation of the cells with the peptides were performed as described for patient B.

Flow cytometry analysis showed that the CD8+ T cells of this patient are induced for IFN-γ production after stimulation with overlapping 16-mer peptide (aa 98-113 ERRDHRRRKALENKKK). This was consistent with the data from patient A and B. However, analysis of single epitopes revealed that the second predicted and HLA-B*27-ligand (aa 103-112 RRRKALENKK) is indeed the minimal epitope targeted by the T cells in this patient and not the previously identified epitope in patient B (aa 99-108 RRDHRRRKAL). As demonstrated in Figure 3.12, 0.55% and 0.75% of the cells were positive for CD8+ and IFN-γ after stimulation with 16-mer peptide (aa 98-113 ERRDHRRRKALENKKK) and (aa 103-112 RRRKALENKK), respectively, whereas, 0.02% of the cells were positive for these two markers after stimulation with the first epitope, (aa 99-108 RRDHRRRKAL).

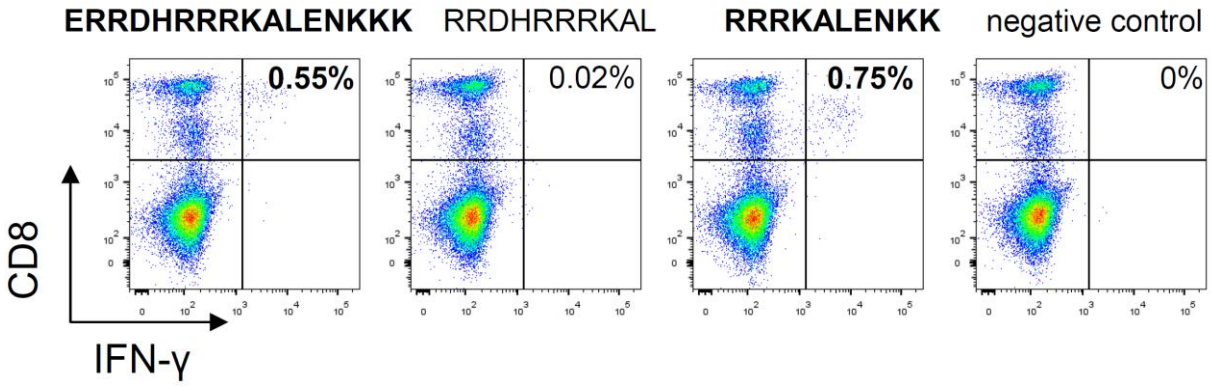


Figure 3.12 CD8 T cell assay in patient C (H36) to distinguish the exact prototype epitope.

Flow cytometry staining of the PBMCs from Patient C (HLA-B*27 positive / HDV resolved patient in the follow-up) stimulated with the 16-mer (first panel from the left) and the 10-mers. The frequency of IFN-γ+/CD8+ T cell determined by ICS is indicated as percentage in each panel.

3.3. HDV immune evasion

Several studies have demonstrated that mutations (substitutions) within epitopes are caused by immune pressure resulting in failure of T cells and promote the evolution of the virus to those variants which are not recognized by immune system and probably are responsible for establishment of a persistent viral infection. Therefore, we asked whether HDV undergoes selection pressure triggered by specific T cell response. To address this question we applied different approaches to detect molecular footprints in correlation with particular HLA alleles.
For this part of the study, we included 104 HDV chronic patients from whom we had viral sequences as well as HLA backgrounds.

3.3.1. Indication of viral immune escape in the identified HLA-B*27 restricted epitope

To answer the question whether the novel HLA-B*27-restricted epitope or any other possible regions within L-HD Ag are under selection pressure restricted by HLA-B*27, we evaluated the viral amino acid sequence variations in our cohort of chronic HDV patients in correlation with the presence of HLA-B*27 allele (Figure 3.13).

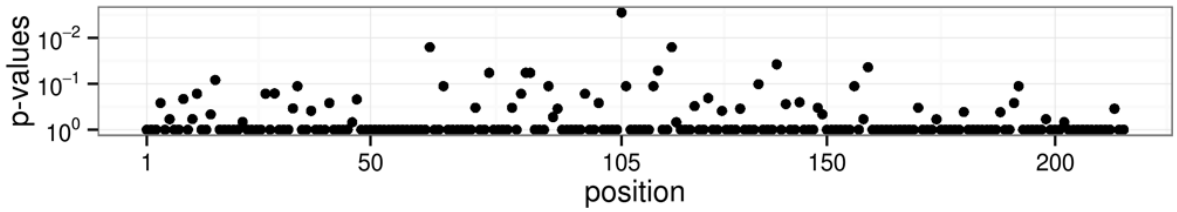


Figure 3.13 Analysis of HLA-B*27-associated viral sequence polymorphism throughout the whole ORF of the L-HD Ag.

Interestingly, the lowest *p*-value goes to the residue 105 on L-HD Ag when the correlation of variability at each position with the presence of HLA-B*27 molecule is tested (Figure 3.13). This residue is located within the newly identified HLA-B*27 epitopes and the low *p*-value indicates a possible immune escape mutation (substitution) in this region in correlation with the selected HLA allele (B*27). Figure 3.14 demonstrates a detailed alignment of the amino acid sequences of L-HD Ag at this region obtained from the studied cohort of 104 patients with chronic HDV infection. Sequences are sorted so that the sequences corresponding to the 6 HLA-B*27 positive patients are on the upper part and those of HLA-B*27 negative patients are located underneath.

	No	Patients	HLA-A and B alleles							Positions	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112
										Consensus	E	R	Q	D	H	R	R	R	K	A	L	E	N	K	K
	1	M03	HLA - A	2	29	HLA - B	**27**	45			.	.	R	.	.	.	.	**K**	.	.	.	.	.	.	.
	2	H09	HLA - A	1	29	HLA - B	**27**	44			.	.	.	.	.	.	.	.	.	.	.	.	.	.	S
HLA-B*27 +	3	B17	HLA - A	0	2	HLA - B	18	**27**			.	.	.	.	.	.	.	**K**	.	.	.	.	.	.	.
	4	H20	HLA - A	3	11	HLA - B	**27**	52			.	.	.	.	.	.	.	.	.	.	.	.	.	.	R
	5	I06	HLA - A	26	30	HLA - B	**27**	57			.	.	.	.	.	.	.	.	**M**	.	.	.	.	.	.
	6	H36	HLA - A	2	24	HLA - B	**27**	50			.	.	.	.	.	.	.	.	.	.	.	.	.	.	E
	7	B13	HLA - A	1	2	HLA - B	15	56			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	8	E11	HLA - A	2	68	HLA - B	0	15			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	9	E38	HLA - A	3	32	HLA - B	15	52			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	10	H19	HLA - A	29	68	HLA - B	15	47			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	11	I17	HLA - A	1	11	HLA - B	8	15			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	12	B02	HLA - A	26	30	HLA - B	14	41			.	.	.	.	.	.	.	.	.	.	.	.	.	.	R
	13	B03	HLA - A	3	33	HLA - B	8	14			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	14	B04	HLA - A	2	30	HLA - B	13	44			.	.	.	.	.	.	.	.	.	.	.	.	.	.	R
	15	B05	HLA - A	24	24	HLA - B	35	44			.	.	.	.	.	.	.	.	.	.	.	.	.	.	R
	16	T02	HLA - A	1	24	HLA - B	44	57			.	.	R	.	.	.	.	.	.	.	.	.	.	.	R
	17	B06	HLA - A	3	11	HLA - B	7	35			.	.	.	.	.	.	.	.	.	.	.	.	.	.	R
	18	B09	HLA - A	24	32	HLA - B	35	51			.	.	.	.	.	.	.	.	.	.	.	.	.	.	R
	19	H38	HLA - A	2	26	HLA - B	39	51			.	.	E	.	.	.	.	.	.	.	.	.	.	.	.
	20	B12	HLA - A	2	33	HLA - B	14	49			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HLA-B*27 -	21	B16	HLA - A	2	23	HLA - B	49	50			.	.	.	.	.	.	.	.	.	.	.	.	.	.	R
	22	E02	HLA - A	23	24	HLA - B	35	62			.	.	.	.	.	.	.	.	.	.	.	.	.	.	R
	23	H17	HLA - A	3	30	HLA - B	13	51			.	.	K	.	.	.	.	.	.	.	.	.	.	.	R
	24	H04	HLA - A	30	32	HLA - B	13	35			.	.	K	.	.	.	.	.	.	.	.	.	.	.	S
	25	E03	HLA - A	1	2	HLA - B	7	8			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	26	E06	HLA - A	1	3	HLA - B	18	50			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	27	E07	HLA - A	24	0	HLA - B	44	51			.	.	.	.	.	.	.	.	.	.	.	.	.	.	R
	28	E08	HLA - A	2	24	HLA - B	35	52			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	29	E10	HLA - A	2	0	HLA - B	41	51			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	30	E12	HLA - A	1	24	HLA - B	35	58			.	.	.	.	.	.	.	.	.	.	.	.	.	.	R
	31	E13	HLA - A	1	3	HLA - B	8	35			.	.	.	.	.	.	.	.	.	.	.	.	.	.	R
	32	E16	HLA - A	2	11	HLA - B	51	55			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	33	E18	HLA - A	3	11	HLA - B	35	51			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	34	E19	HLA - A	11	30	HLA - B	18	35			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	35	E20	HLA - A	2	3	HLA - B	18	41			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	36	E21	HLA - A	2	0	HLA - B	50	52			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	37	E23	HLA - A	24	31	HLA - B	35	51			.	.	.	.	.	.	.	.	.	.	.	.	.	.	R
	38	E28	HLA - A	2	24	HLA - B	38	51			.	.	.	.	.	.	.	.	.	.	.	.	.	.	R
	39	E31	HLA - A	1	11	HLA - B	35	40			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	40	E32	HLA - A	66	68	HLA - B	52	57			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	41	E33	HLA - A	24	0	HLA - B	18	51			.	.	.	.	.	.	.	.	.	.	.	.	.	.	R
	42	E36	HLA - A	2	0	HLA - B	44	0			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	43	H01	HLA - A	2	3	HLA - B	7	35			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	44	H02	HLA - A	2	24	HLA - B	38	73			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	45	H08	HLA - A	26	29	HLA - B	7	44			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	46	H14	HLA - A	3	24	HLA - B	8	51			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	47	H15	HLA - A	2	0	HLA - B	40	0			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	48	H23	HLA - A	1	2	HLA - B	41	44			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	49	H25	HLA - A	3	24	HLA - B	35	38			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	50	H26	HLA - A	11	69	HLA - B	35	52			.	.	.	.	.	.	.	.	.	.	.	.	.	.	R
	51	H27	HLA - A	1	2	HLA - B	7	8			.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

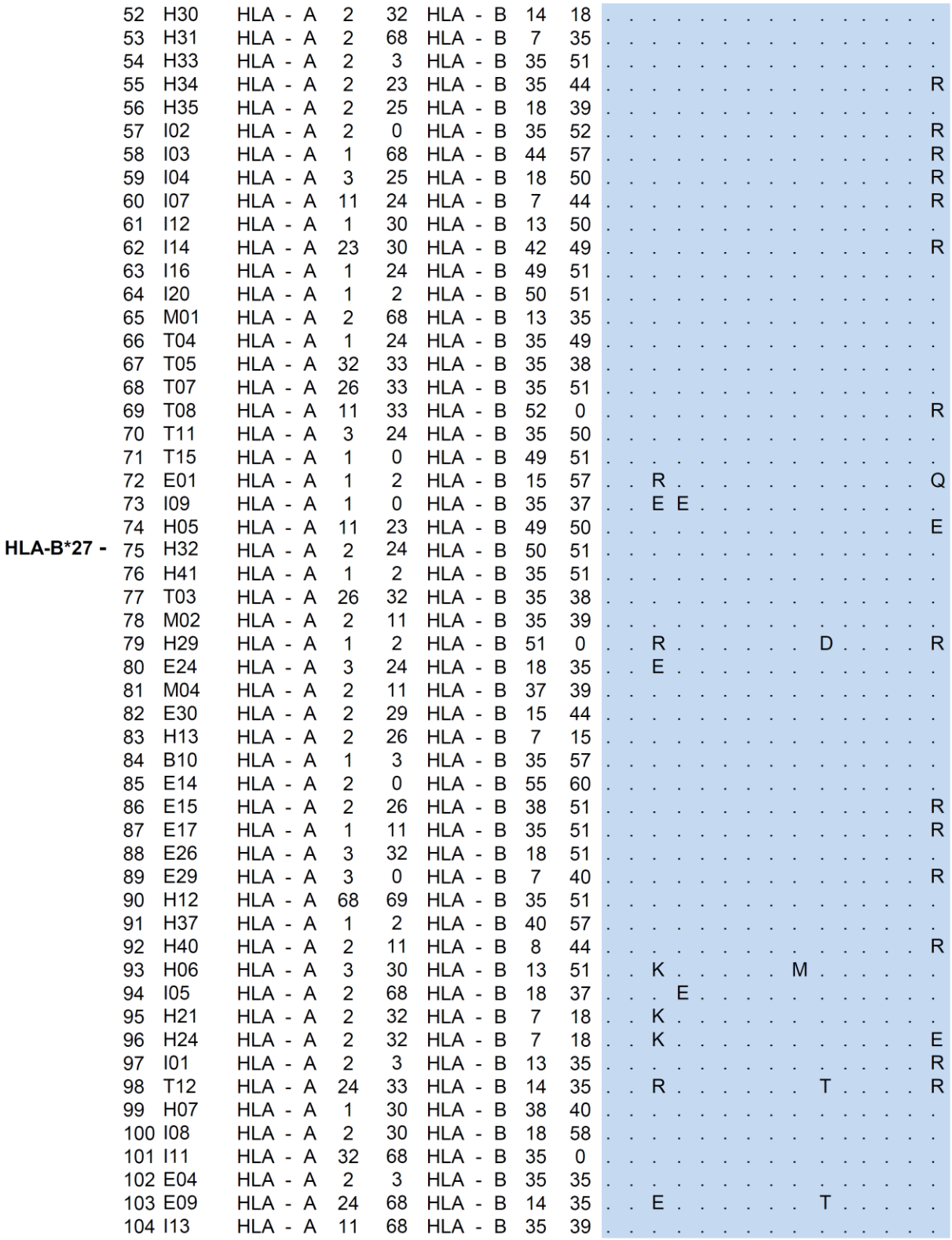


Figure 3.14 HDV amino acid sequence alignment depicting substitution pattern within the described epitope restricted to HLA-B*27.

The amino acid sequence alignment in the region restricted to HLA-B*27 is sorted by 6 HLA-B*27 positive and 98 HLA-B*27 negative patients. Substitutions regarding possible immune escape mutations inside these epitopes are indicated by a darker color.

Interestingly, we observed a higher substitution rate (p= 0.0028) in the region restricted to HLA-B*27 within the 6 HLA-B*27 positive patients when compared with HLA-B*27 negative subjects of this study (Figure 3.14). In fact, the residue 99 to 112 region, where the epitopes are located, is the least variable region of L-HD Ag in HLA-B*27 negative patients in comparison with other spots in the whole studied isolates; therefore, it is less probable for random substitutions to occur in this region. Sequence analysis also indicates some substitutions which are not significantly correlated to HLA-B*27. Q100R substitution was observed in one out of 6 HLA-B*27 positive patients (~16%). This variation at the same residue was also observed in the group of HLA-B*27 negative patients (Q100R, 4%; Q100E, 4% and Q100K, 5%). While variation at the position 100 is not associated with the HLA-B*27 allele and can be observed in both group of HLA-B*27 positive and negative subjects, there are two unique substitutions in the corresponding region, R105K and K106M, obtained from HLA-B*27 positive patients (Figure 3.13). This higher substitution rate pattern in this region in HLA-B*27 positive patients may indicate viral escape in HLA-B*27 positive individuals.

3.3.2. Amino acid substitutions and their impact on viral-specific CD8+ T cell response

Next, we asked whether the identified substitutions within HLA-B*27-restricted epitope (L-HD $Ag_{99\text{-}108}$) would have an impact on the magnitude of HDV-specific CD8+ T cell response. To address this question, we performed a T cell assay using synthetic peptides from the identified HLA-B*27 epitope with all observed substitutions. For this assay HDV-specific T cell lines (driven from PBMCs of a patient with resolved HDV infection) were stimulated and the induction of HDV-specific CD8+ T cells to produce IFN-γ was evaluated. Amino acid analysis at this region indicated that the substitutions are divided into two groups: 1) the polymorphisms at position 100 (L-HD Ag_{100}) which are apparently not associated with HLA-B*27 (R**R**DHRRRKAL, R**Q**DHRRRKAL, R**E**DHRRRKAL and R**K**DHRRRKAL) and 2) substitutions at residues 105 (R105K) and 106 (K106M) which are significantly associated with the presence of HLA-B*27 and seem to be selected under immune selection pressure restricted by HLA-B*27. Figure 3.15 demonstrates the impact of sequence polymorphism at second position of the epitope (L-HD Ag_{100}) on the HDV-specific T cell response. As shown in Figure 3.15, the 4 observed amino acid residues (R, Q, K and E) at the second position of the novel HDV epitope do not have a significant impact on the induction of CD8+ T cells to produce IFN-γ; that is, regardless of which of these 4 amino acids (R, Q, K and E) are located at this position, the magnitude of the intra cellular IFN-γ response show no differences and would be between 0.50-0.59%. To investigate whether these substitutions at the 2^{nd} residue of the epitope RRDHRRRKAL are present in previously studied genotype 1

isolates or other genotypes (2 to 8), we performed a comprehensive analysis of all available sequences from HDV genotypes 1 to 8 with respect to this region.

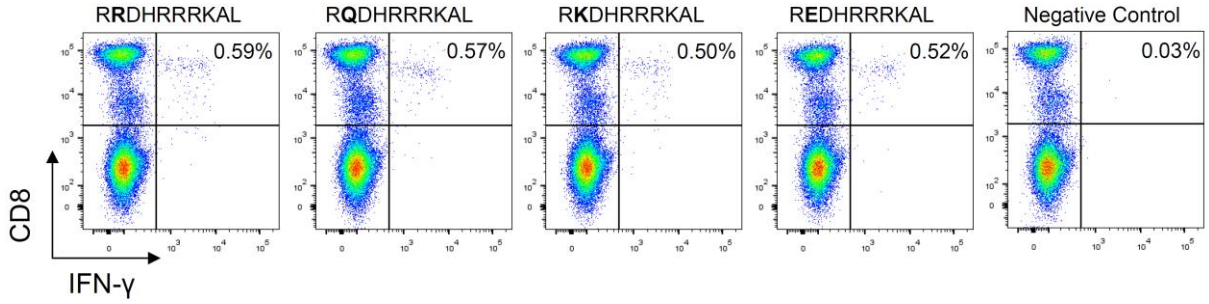


Figure 3.15 Sequence variations at 2nd position of the identified HLA-B*27-restricted epitope (L-HD Ag_{99-108}) and their impact on CD8+ T cell response.

The role of variations at the second position of HLA-B*27-restricted epitope was determined by re-stimulation of HDV-specific T cell lines with all 4 possible variants and compared with the negative control. The frequency of IFN-γ producing CD8+ T cells, determined by ICS, is indicated as percentage in each panel.

Table 3.7 demonstrates the most prevalent amino acid residues at L-HD Ag_{100} and indicates that, similar to our cohort, this position can accept any of the four amino acid residues (Q, R, K and E) regardless of the genotype. Nevertheless, the prevalence of these residues is not equal between different HDV genotypes. An analysis of 543 isolates at this region, from a local data base including all available HDV genotypes submitted in the GenBank as well as isolates from this study showed a combination of Glutamine (Q), Arginine (R), Glutamic acid (E), Lysine (K) and Glycine (G) in 57, 34, 6, 2 and 0.2% of isolates, respectively.

Table 3.7 Polymorphism of L-HD Ag in the region restricted to HLA-B*27 among all HDV genotypes (1 to 8) with respect to amino acid residue at L-HD Ag_{100}.

	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112
Consensus	**E**	**R**	**R**	**D**	**H**	**R**	**R**	**R**	**K**	**A**	**L**	**E**	**N**	**K**	**K**
HDV Genotype 1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	.	.	Q	.	.	.	.	.	.	.	.	.	.	.	.
	.	.	K	.	.	.	.	.	.	.	.	.	.	.	.
	.	.	E	.	.	.	.	.	.	.	.	.	.	.	.
HDV Genotype 2	.	.	E	.	.	.	.	.	.	.	.	.	.	.	.
	.	.	Q	.	.	.	.	.	.	.	.	Q	.	.	.
HDV Genotype 3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	.	.	Q	.	.	.	.	.	.	.	.	.	.	.	.
HDV Genotype 4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	R
	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HDV Genotype 5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	.	.	Q	.	.	.	.	.	.	.	.	.	.	.	.
HDV Genotype 6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HDV Genotype 7,8	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Next, we addressed whether R105K and K106M substitutions, observed in HLA-B*27 positive patients, have an impact on the function of CD8+ T cells. To address this issue, we used synthetic peptides having these substitutions to stimulate PBMCs of a patient with resolved HDV infection (patient B) and compared with result of the same cells when they are stimulated with the “wild type” epitope (this patient had already reacted to the “wild type” epitope, L-HD Ag_{99-108}). As for the “wild type” epitope, we used both Arginine (R) and Glutamine (Q) at L-HD Ag_{100}, RRDHRRRKAL and RQDHRRRKAL, respectively, since these two were the most prevalent residues compared to other two substitutions, Glutamic acid (E) and Lysine (K). Therefore, the corresponding mutant variants for R105K and K106M were synthetized for both “wild type” epitopes, RRDHRRRKAL and RQDHRRRKAL (Table 3.8).

Table 3.8 List of synthetic peptides corresponding to the identified HLA-B*27-restricted epitope with all observed substitutions.

#	Position	Sequence	Length	HLA	Comment
1	HDV 99-108	RQDHRRKKAL	10	B*27	R100Q, R105K
2	HDV 99-108	RQDHRRRMAL	10	B*27	R100Q, K106M
3	HDV 99-108	RRDHRRKKAL	10	B*27	Q100R, R105K
4	HDV 99-108	RRDHRRRMAL	10	B*27	Q100R, K106M
5	HDV 99-108	RKDHRRRKAL	10	B*27	Q100K Wild Type
6	HDV 99-108	REDHRRRKAL	10	B*27	Q100E Wild Type
7	HDV 99-108	RRDHRRRKAL	10	B*27	Q100R Wild Type
8	HDV 99-108	RQDHRRRKAL	10	B*27	R100Q Wild Type

To test the functional impact of amino acid substitutions within this region on T cell response, the epitope-specific cell lines were stimulated using all possible combination of mutations in a same T cell assay. The results of this assay are demonstrated in Figure 3.16. Analysis confirmed that the “wild type” epitope (aa 99-108 RRDHRRRKAL or RQDHRRRKAL) is able to reproducibly induce CD8+ T cells to produce IFN-γ; however, the mutated variants (R105K and K106M) are not able to induce IFN-γ production when PBMCs are stimulated with these mutated epitopes (Figure 3.16). Taken together, strong statistical evidence as well as the experimental findings indicate a clear viral selection in HLA-B*27 positive patients for variants which can evade an effective immune response and probably supports a persistent HDV infection in those patients.

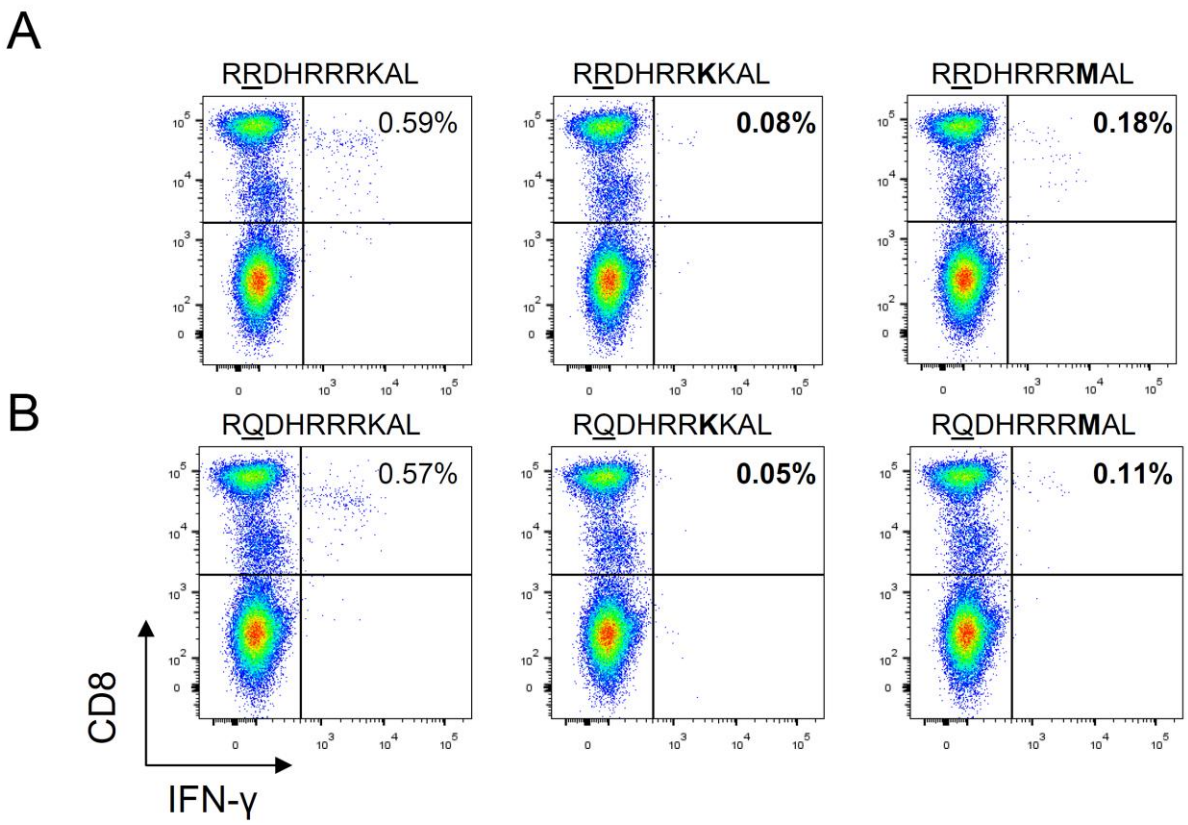


Figure 3.16 Functional impact of HLA-B*27 selected variants on the HDV-specific CD8+ T cell response.

A 10-day antigen-specific expansion of the PBMCs in the presence of the prototype epitope was performed. Then the cultures were re-stimulated with the prototype or mutated epitopes and the number of IFNγ+/CD8+ T cells was measured (A) RRDHRRRKAL (frequently seen in genotype 1 and other genotypes) was applied as the prototype along with the corresponding variants (R105K and K106M): RRDHRR**K**KAL and RRDHRRR**M**AL. (B) RQDHRRRKAL (prevalent sequence in this cohort) was used as the wild type epitope and the corresponding mutated epitopes at residues 105 and 106: RQDHRR**K**KAL and RQDHRRR**M**AL.

Next, we asked whether isolates from HLA-B*27 patients show more variations at the regions restricted by HLA-B*27 when we apply UDPS (deep sequencing) technology. Therefore, 5 isolates from HLA-B*27 positive patients were compared with 6 isolates from HLA-B*27 negative patients as control samples. The data from this study indicated that indeed, in addition to the mutations observed in conventional sequencing, there are minor variations within this region in the HLA-B*27 positive isolates which are absent in the control (HLA-B*27 negative) group at the same region. This may indicate additional HLA-B*27-restricted T cell pressure on other residues at lower level.

3.4. Detection of molecular signatures in correlation with certain HLA alleles

3.4.1. Identification of residues under CD8+ T cell selection pressure restricted by HLA class I molecules

According to the growing body of knowledge about evolution of viruses under immune selection pressure, researchers asked whether analyzing the correlation between the amino acid substitutions and HLA alleles allows identification of specific immune pressure mediated by the T cells. Virus-specific T cell response is restricted by specific epitope presentation by the HLA alleles, and, therefore, identification of amino acid mutations in the presence of particular alleles indicates immune escape from the specific T cell response at the region with substitution. Given the observations on the impact of HLA-B*27 in viral evolution and selection of specific substitutions inside the epitopes, we asked if sequence analysis allows identification of selection pressure triggered by other HLA class I molecules. Therefore, we searched for possible associations between sequence polymorphism at each residue of L-HD Ag and HLA molecules. For this purposes, we employed a recently developed tool, SeqFeatR, which identifies the correlation between the **seq**uence mutations and a **feat**ure, *e.g.* HLA alleles, of the respective sequence. These analyses determined if any amino acid residues were significantly more frequent in the presence of particular HLA class I alleles. The analysis was done based on Fisher's exact test utilizing the R package 'SeqFeatR' for all HLA class I alleles.

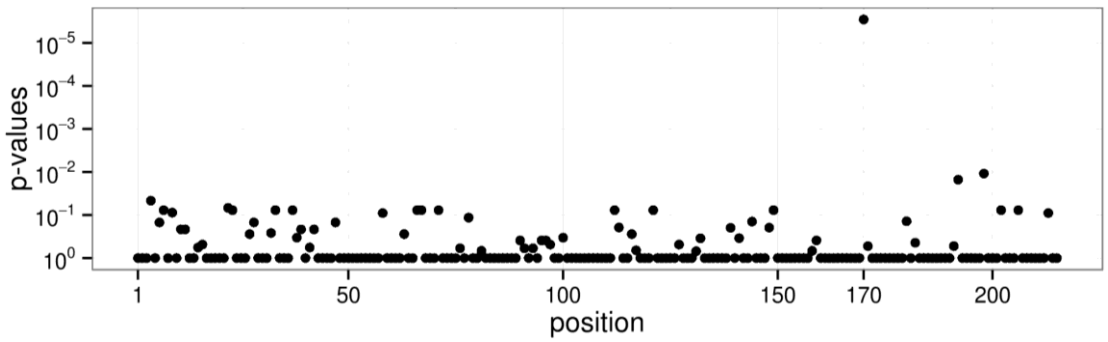
Sequence analysis was performed on sequences obtained from 104 patients whose HLA class I alleles were also determined. Correlation between each amino acid residue and HLA molecule and the corresponding *p* values were calculated. These analyses indicated several HLA footprints on L-HD Ag coding region. Comprehensive tables of HLA-sequence correlation of all HLA alleles including A and B alleles and particular amino acid substitutions are listed in Appendix IV.

3.4.2. Molecular signatures revealed a novel CD8+ T cell epitope restricted by HLA-B*15

HLA-B*15 can select for variants with L-HD Ag S170N substitution. Using SeqFeatR tool, we have found a clear footprint ($p<10^{-6}$) of HLA-B*15 on the aa position 170 of L-HD Ag (Figure 3.17A). Analyzing of the sequences obtained from this study revealed that indeed the Serine (Ser; S) at position 170 was substituted with an Asparagine (Asn; N) in all 8 (100%) HLA-B*15 positive patients; whereas, only 11 of 96 HLA-B*15 negative patients (~11%) showed this substitution (Figure 3.17B). Among the HLA-B*15 positive patients, 4 patients were from Essen, 2 patients from Hannover, 1 from Barcelona and 1 from the

collaborating center in Turin, Italy. However, these data do not necessarily reflect the origin of these patients or the source of infection. There were two additional substitutions within this region: one Methionine at position 171 in the sequence of patient E01 was substituted with Leucine (M171L) and the other one was observed in patient E38 where Valine at position 174 was substituted with Isoleucine (V174I). However, these two substitutions were also observed in the isolates from HLA-B*15 negative patients with almost same frequencies. Therefore, they were not significantly correlated to the presence of HLA-B*15 allele.

A



B

			aa Positions	169	170	171	172	173	174	175	176	177	178	179
	No	Patients	HLA-A and B alleles Cons	P	S	M	Q	G	V	P	E	S	P	F
	1	B13	HLA - A 1 2 HLA - B 15 56	.	N	.	.	.	.	.	.	.	.	.
	2	E30	HLA - A 2 29 HLA - B 15 44	.	N	.	.	.	.	.	.	.	.	.
	3	H13	HLA - A 2 26 HLA - B 7 15	.	N	.	.	.	.	.	.	.	.	.
HLA-B*15 +	4	I17	HLA - A 1 11 HLA - B 8 15	.	N	.	.	.	.	.	.	.	.	.
	5	E01	HLA - A 1 2 HLA - B 15 57	.	N	L	.	.	.	.	.	.	.	.
	6	H19	HLA - A 29 68 HLA - B 15 47	.	N	.	.	.	.	.	.	.	.	.
	7	E38	HLA - A 3 32 HLA - B 15 52	.	N	.	.	.	I	.	.	.	.	.
	8	E11	HLA - A 2 68 HLA - B 15 0	.	N	.	.	.	.	.	.	.	.	.
	9	H36	HLA - A 2 24 HLA - B 27 50	.	.	.	.	.	.	.	.	.	.	.
	10	E24	HLA - A 3 24 HLA - B 18 35	.	.	.	.	.	.	.	.	.	.	.
	11	I02	HLA - A 2 0 HLA - B 35 52	.	N	.	.	.	.	.	.	.	.	.
	12	H21	HLA - A 2 32 HLA - B 7 18	.	.	.	.	.	.	.	.	.	.	.
	13	H24	HLA - A 2 32 HLA - B 7 18	.	.	.	.	.	.	.	.	.	.	.
	14	E15	HLA - A 2 26 HLA - B 38 51	.	.	.	.	.	I	.	.	.	.	.
	15	H29	HLA - A 1 2 HLA - B 51 0	.	.	.	.	.	.	.	.	.	.	.
	16	E08	HLA - A 2 24 HLA - B 35 52	.	.	L	.	.	.	.	.	.	.	.
	17	B04	HLA - A 2 30 HLA - B 13 44	.	.	.	.	.	I	.	.	.	.	.
	18	E21	HLA - A 2 0 HLA - B 50 52	.	.	.	.	.	.	.	.	.	.	.
	19	B09	HLA - A 24 32 HLA - B 35 51	.	.	.	.	.	.	.	.	.	.	.
	20	B06	HLA - A 3 11 HLA - B 7 35	.	.	.	.	.	.	.	.	.	.	.
HLA-B*15 -	21	E13	HLA - A 1 3 HLA - B 8 35	.	.	.	.	.	.	.	.	.	.	.
	22	E31	HLA - A 1 11 HLA - B 35 40	.	N	.	.	.	I	.	.	.	.	.
	23	H04	HLA - A 30 32 HLA - B 13 35	.	.	.	.	.	.	.	.	.	.	.
	24	H26	HLA - A 11 69 HLA - B 35 52	.	.	.	.	.	.	.	.	.	.	.
	25	B17	HLA - A 0 0 HLA - B 18 27	.	.	.	.	.	.	.	.	.	.	.
	26	I04	HLA - A 3 25 HLA - B 18 50	.	.	.	.	.	.	.	.	.	.	.
	27	H06	HLA - A 3 30 HLA - B 13 51	.	.	.	.	.	.	.	.	.	.	.
	28	H14	HLA - A 3 24 HLA - B 8 51	.	.	.	.	.	I	.	.	.	.	.
	29	H17	HLA - A 3 30 HLA - B 13 51	.	.	.	.	.	.	.	.	.	.	.
	30	B02	HLA - A 26 30 HLA - B 14 41	.	.	.	L	.	.	.	.	.	.	.
	31	H09	HLA - A 1 29 HLA - B 27 44											

Figure 3.17 Analysis of HLA-B*15-associated viral sequence polymorphism throughout the whole ORF of the L-HD Ag.

(A) Manhattan plot with logarithmically scaled *p* values indicating association of amino acid substitutions with HLA-B*15 in an alignment of 104 sequences. Amino acid residue positions on the alignment are indicated by numbers in the x axis. (B) Multiple amino acid sequence alignment from the studied isolates at the region possibly restricted by HLA-B*15. Isolates from HLA-B*15 positive patients are on top. Amino acid residues similar to the consensus are indicated with dots. Substituted amino acids indicating escape mutations are marked by darker color.

Prediction tools revealed that immune escape mutation may occur within a HLA-B*15 epitope. MHC class I epitope prediction tools (available at http://tools.immuneepitope.org/mhci) indicated that the 10-mer region between 170 and 179, where we found a significant molecular signature by HLA-B*15, is indeed a possible epitope restricted to HLA-B*15 molecule. As demonstrated in Table 3.9, HLA-B*15 predicted ligands are among the best predicted epitope with strongest prediction scores (Percentile rankings<1) within the selected HLA alleles. These data indicate L-HD $Ag_{170-179,}$ SMQGVPESPF, is the best hit for HLA-B*15 throughout the whole open reading frame of L-HD Ag, which is consistent with the molecular footprint analysis.

Table 3.9 Detection of possible HDV-specific HLA class I ligands by prediction algorithms with percentile ranking lower than 1.

Allele	#	Start	End	Length	Peptide	Percentile rank
HLA-B*15:02	1	170	179	10	SMQGVPESPF	0.2
HLA-B*15:01	1	171	179	9	MQGVPESPF	0.2
HLA-B*15:01	1	198	206	9	ILFPADPPF	0.2
HLA-B*27:05	1	99	108	10	RRDHRRRKAL	0.35
HLA-B*27:05	1	103	112	10	RRRKALENKK	0.4
HLA-B*27:05	1	104	113	10	RRKALENKKK	0.45
HLA-B*15:25	1	170	179	10	SMQGVPESPF	0.5
HLA-B*15:25	1	170	179	10	SMQGVPESPF	0.5
HLA-B*35:01	1	194	203	10	FPWDILFPAD	0.6
HLA-B*07:02	1	177	186	10	SPFTRTGEGL	0.65
HLA-B*35:01	1	200	209	10	FPADPPFSPQ	0.8
HLA-B*15:01	1	170	179	10	SMQGVPESPF	0.8
HLA-B*27:05	1	103	112	10	RRRKALENKK	0.9
HLA-B*07:02	1	200	209	10	FPADPPFSPQ	0.95
HLA-B*07:02	1	147	155	9	PPVGGVNPL	1

Variation at the first position of HLA-B*15-restricted epitope alters the prediction scores for this epitope. To address whether S170N has an impact on MHC (HLA-B*15) binding prediction, we replaced the Serine residue at position 170 with Asparagine in the open reading frame of L-HD Ag and performed the prediction. Comparison of the HLA-B*15 epitope hits between L-HD Ag with S170N substitution, indicated that the L-HD $Ag_{170-179}$ is still among the best candidate ligands for HLA-B*15; however, there were dramatically

increase in the percentile rankings corresponding to this S170N variant than the “wild type” indicating the impact of this substitution on the prediction and possibly the immunogenicity and functionality of this mutated ligand in driving an HDV-specific T cell response. As demonstrated in Figure 3.18, the corresponding percentile ranks for S170N variant of the predicted HLA-B*15 epitope increase compared to the “wild type”, except for HLA-B*15:02.

Allele	#	Start	End	Length	Peptide	Method	Percentile_rank
HLA-B*15:02	1	1	10	10	SMQGVPESPF	ann	0.2
HLA-B*15:25	1	1	10	10	SMQGVPESPF	netmhcpan	0.5
HLA-B*15:01	1	1	10	10	SMQGVPESPF	Consensus (ann/smm)	0.8

Allele	#	Start	End	Length	Peptide	Method	Percentile_rank
HLA-B*15:02	1	1	10	10	NMQGVPESPF	ann	0.2
HLA-B*15:25	1	1	10	10	NMQGVPESPF	netmhcpan	1.6
HLA-B*15:01	1	1	10	10	NMQGVPESPF	Consensus (ann/smm)	1.1

Figure 3.18 Impact of HLA-B*15-assciated amino acid substitution (S170N) on prediction scores of HLA-B*15-restricted epitopes.

T cell assay confirmed that the predicted peptide with escape mutation is indeed a functional epitope. Stimulation of PBMCs isolated from a patient with resolved HDV infection demonstrated IFN-γ production upon re-stimulation with 16-mer overlapping peptide (P**SMQGVPESPF**ARTGE, from the library) as well as the 10-mer and 9-mer predicted epitopes. Interestingly, re-stimulation with both 10-mer (SMQGVPESPF) and 9-mer (MQGVPESPF) ligands resulted in elevation of CD8+/IFN-γ+ T cells indicating specific T cell responses to these epitopes. However, response to the 10-mer was significantly higher than that to the 9-mer peptide. PBMCs of the same patient responded also to the stimulation with the mutated 10-mer epitope (**N**MQGVPESPF) at slightly lower frequencies (Figure 3.19).

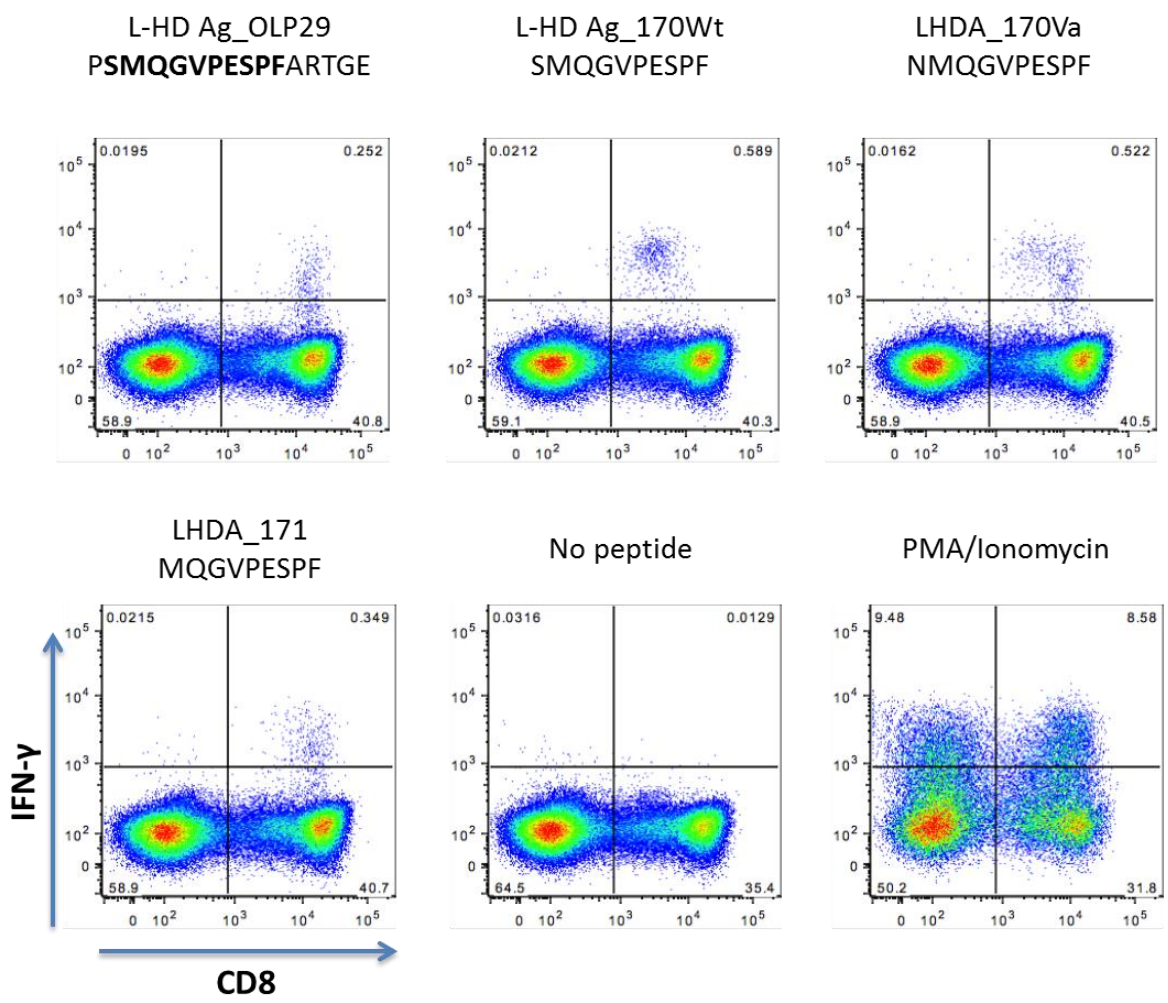


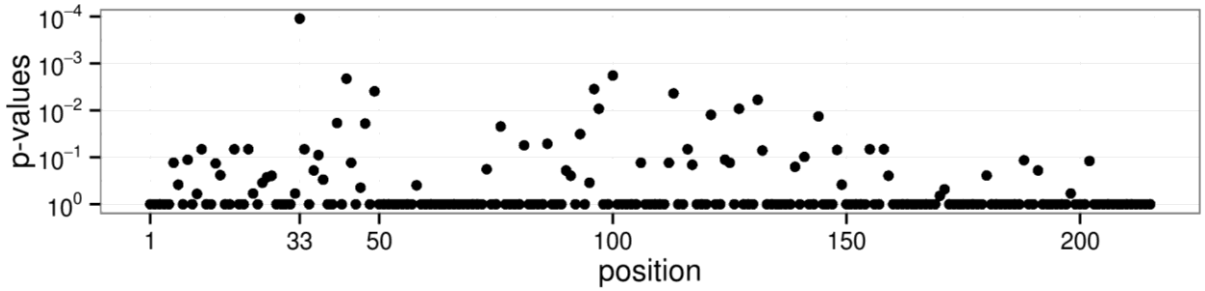
Figure 3.19 HDV-specific CD8+ T cell assay restricted by HLA-B*15.

Depiction of IFN-γ production upon peptide stimulation of PBMCs isolated from a HLA-B*15 patient with resolved HDV infection kindly provided by the collaborating center in Freiburg, Germany (AG Neumann-Haefelin)

3.4.3. Further analysis of the molecular footprints predicts novel CD8+ T cell epitopes and strong selection pressure by HLA-B alleles, *e.g.* -B*13, -B*37 and -B*41

Three alleles, HLA-B*13, -B*37 and -B*41 are given as examples in Figure 3.20, Figure 3.21 and Figure 3.22, respectively. The regions are arbitrarily marked for these three alleles and may not represent the exact amino acid sequences of the respective functional epitopes. Further analyses are needed to indicate the minimal epitope sequences restricted by these alleles.

A



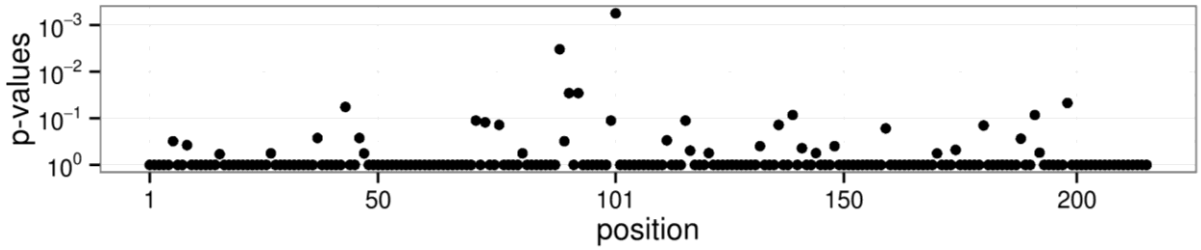
B

	No	Patients	HLA-A and B alleles						aa Positions	28	29	30	31	32	33	34	35	36	37	38	39	40
									Consensus	E	E	L	E	R	D	L	R	K	V	K	K	K
	1	I01	HLA - A	2	3	HLA - B	13	35		.	.	.	.	.	E	.	.	.	.	.	.	.
	2	M01	HLA - A	2	68	HLA - B	13	35		.	.	.	.	.	.	.	.	.	.	.	.	.
	3	B04	HLA - A	2	30	HLA - B	13	44		.	D	.	.	.	E	.	.	.	E	.	.	.
HLA-B*13 +	4	H04	HLA - A	30	32	HLA - B	13	35		.	.	.	.	.	E	.	.	.	.	.	.	.
	5	H06	HLA - A	3	30	HLA - B	13	51		.	.	.	.	.	E	.	.	R	.	R	.	.
	6	H17	HLA - A	3	30	HLA - B	13	51		.	.	.	.	.	E	S	.	.	I	.	.	.
	7	I12	HLA - A	1	30	HLA - B	13	50		.	.	.	.	.	E	.	.	.	I	.	.	.
	9	H41	HLA - A	1	2	HLA - B	35	51		.	.	.	.	.	E	.	.	.	A	.	.	.
	10	E04	HLA - A	2	3	HLA - B	35	35		.	.	.	.	.	.	.	.	.	.	.	.	.
	11	E08	HLA - A	2	24	HLA - B	35	52		.	.	.	.	.	.	.	.	.	.	.	.	.
	12	H01	HLA - A	2	3	HLA - B	7	35		.	.	.	.	.	.	V	.	.	.	.	.	.
	13	H31	HLA - A	2	68	HLA - B	7	35		.	.	.	.	.	.	.	.	.	.	.	.	.
	14	H34	HLA - A	2	23	HLA - B	35	44		.	.	.	.	.	.	.	.	.	.	.	.	.
	15	I02	HLA - A	2	0	HLA - B	35	52		.	D	.	.	.	.	.	.	.	.	.	.	.
HLA-B*13 -	16	M02	HLA - A	2	11	HLA - B	35	39		.	.	.	.	.	.	.	.	.	.	.	.	.
	17	H35	HLA - A	2	25	HLA - B	18	39		.	.	.	.	.	.	.	.	.	.	.	.	.
	18	I05	HLA - A	2	68	HLA - B	18	37		.	.	.	.	.	.	.	.	.	.	.	.	.
	19	I08	HLA - A	2	30	HLA - B	18	58		.	.	.	.	.	.	.	.	.	.	.	.	.
	20	E10	HLA - A	2	0	HLA - B	41	51		.	.	.	.	.	.	.	.	.	I	.	.	.
	21	E15	HLA - A	2	26	HLA - B	38	51		.	.	.	.	.	.	.	.	.	.	.	.	.
	22	E16	HLA - A	2	11	HLA - B	51	55		.	.	.	.	.	.	.	.	.	M	.	.	.
	23	E28	HLA - A	2	24	HLA - B	38	51		.	.	.	.	.	.	.	.	.	I	.	R	.
	24	H29	HLA - A	1	2	HLA - B	51	0		.	.	.	.	.	E	.	.	.	.	.	.	.
	25	H32	HLA - A	2	24	HLA - B	50	51		.	.	.	.	.	.	.	.	.	.	.	.	.
	26	H38	HLA - A	2	26	HLA - B	39	51		.	.	.	.	.	.	.	.	.	.	.	.	.
	27	I20	HLA - A	1	2	HLA - B	50	51			D											

Figure 3.20 Analysis of HLA-B*13-associated viral sequence polymorphism.

(A) calculated *p* values for HLA-B*13-sequence association throughout the whole ORF of the L-HD Ag. (B) Amino acid sequence alignment of some of the studied isolates at the region around residue 33. Sequences of the isolates from HLA-B*13 positive patients are sorted on top and followed by those from HLA-B*13 negative patients. Amino acid residues with statistical evidence for selection are marked with a darker color (position 33).

A



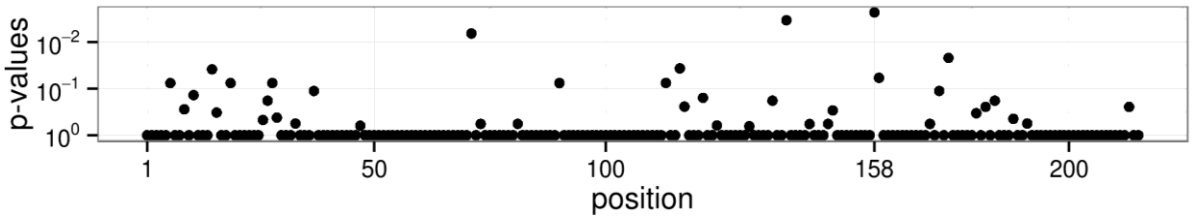
B

										aa Positions	100	101	102	103	104	105	106	107	108	109	110	111
	No	Patients	HLA-A and B alleles							Cons	Q	D	H	R	R	R	K	A	L	E	N	K
HLA-B*37 +	1	I05	HLA - A	2	68	HLA - B	18	37		.	E	.	.	.	.	.	.	.	.	.	.	
	2	M04	HLA - A	2	11	HLA - B	37	39		.	.	.	.	.	.	.	.	.	.	.	.	
	3	I09	HLA - A	1	0	HLA - B	35	37		E	E	.	.	.	.	.	.	.	.	.	.	
	4	H33	HLA - A	2	3	HLA - B	35	51		.	.	.	.	.	.	.	.	.	.	.	.	
	5	H41	HLA - A	1	2	HLA - B	35	51		.	.	.	.	.	.	.	.	.	.	.	.	
	6	E04	HLA - A	2	3	HLA - B	35	35		.	.	.	.	.	.	.	.	.	.	.	.	
	7	E08	HLA - A	2	24	HLA - B	35	52		.	.	.	.	.	.	.	.	.	.	.	.	
	8	H01	HLA - A	2	3	HLA - B	7	35		.	.	.	.	.	.	.	.	.	.	.	.	
	9	H31	HLA - A	2	68	HLA - B	7	35		.	.	.	.	.	.	.	.	.	.	.	.	
	10	H34	HLA - A	2	23	HLA - B	35	44		.	.	.	.	.	.	.	.	.	.	.	.	
	11	I01	HLA - A	2	3	HLA - B	13	35		.	.	.	.	.	.	.	.	.	.	.	.	
	12	I02	HLA - A	2	0	HLA - B	35	52		.	.	.	.	.	.	.	.	.	.	.	.	
	13	M01	HLA - A	2	68	HLA - B	13	35		.	.	.	.	.	.	.	.	.	.	.	.	
HLA-B*37 -	14	M02	HLA - A	2	11	HLA - B	35	39		.	.	.	.	.	.	.	.	.	.	.	.	
	15	E20	HLA - A	2	3	HLA - B	18	41		.	.	.	.	.	.	.	.	.	.	.	.	
	16	H30	HLA - A	2	32	HLA - B	14	18		.	.	.	.	.	.	.	.	.	.	.	.	
	17	H35	HLA - A	2	25	HLA - B	18	39		.	.	.	.	.	.	.	.	.	.	.	.	
	18	I08	HLA - A	2	30	HLA - B	18	58		.	.	.	.	.	.	.	.	.	.	.	.	
	19	E10	HLA - A	2	0	HLA - B	41	51		.	.	.	.	.	.	.	.	.	.	.	.	
	20	E15	HLA - A	2	26	HLA - B	38	51		.	.	.	.	.	.	.	.	.	.	.	.	
	21	E16	HLA - A	2	11	HLA - B	51	55		.	.	.	.	.	.	.	.	.	.	.	.	
	22	E28	HLA - A	2	24	HLA - B	38	51		.	.	.	.	.	.	.	.	.	.	.	.	
	23	H32	HLA - A	2	24	HLA - B	50	51		.	.	.	.	.	.	.	.	.	.	.	.	
	24	I20	HLA - A	1	2	HLA - B	50	51		.	.	.	.	.	.	.	.	.	.	.	.	
	25	B04	HLA - A	2	30	HLA - B	13	44		.	.	.	.	.	.	.	.	.	.	.	.	
	26	B12	HLA - A	2	33	HLA - B	14	49		.	.	.	.	.	.	.	.	.	.	.	.	
	27	B13	HLA - A	1	2	HLA - B	15	56		.	.	.	.	.	.	.	.	.	.	.	.	
	28	B16	HLA - A	2	23	HLA - B	49	50		.	.	.	.	.	.	.	.	.	.	.	.	
	29	E03	HLA - A	1	2	HLA - B	7	8		.	.	.	.	.	.	.	.	.	.	.	.	
	30	E11	HLA - A	2	68	HLA - B	15	0		.	.	.	.	.	.	.	.	.	.	.	.	

Figure 3.21 Analysis of HLA-B*37-associated viral sequence polymorphism throughout the whole ORF of the L-HD Ag.

A) calculated *p* values for HLA-B*37-sequence association throughout the whole ORF of the L-HD Ag. (B) Amino acid sequence alignment of some of the studied isolates at the region around residue 101. Sequences of the isolates from HLA-B*37 positive patients are sorted on top and followed by those from HLA-B*37 negative patients. Amino acid residues with statistical evidence for selection are marked with a darker color (position 101).

A



B

	No	Patients	HLA-A and B alleles	Con	aa Positions 132	133	134	135	136	137	138	139	140	151	152	153	154	155	156	157	158
					R	L	T	E	E	D	E	R	R	G	V	N	P	L	E	G	G
	1	H23	HLA - A 1 2 HLA - B 41 44		.	.	.	.	.	.	.	K	.	.	.	.	.	.	.	.	D
B*41 +	2	E20	HLA - A 2 3 HLA - B 18 41		.	.	.	.	.	.	.	K	.	.	.	.	.	.	.	.	A
(n=4)	3	B02	HLA - A 26 30 HLA - B 14 41		.	.	.	.	.	.	.	K	.	.	.	.	.	.	.	.	A
	4	E10	HLA - A 2 0 HLA - B 41 51		.	.	.	.	D	.	.	.	.	.	.	.	.	.	.	.	M
	5	H34	HLA - A 2 23 HLA - B 35 44		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	6	E30	HLA - A 2 29 HLA - B 15 44		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	7	H40	HLA - A 2 11 HLA - B 8 44		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	8	H08	HLA - A 26 29 HLA - B 7 44		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	9	B05	HLA - A 24 24 HLA - B 35 44		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A
	10	I03	HLA - A 1 68 HLA - B 44 57		.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	11	I07	HLA - A 11 24 HLA - B 7 44		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	12	T02	HLA - A 1 24 HLA - B 44 57		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	13	E07	HLA - A 24 0 HLA - B 44 51		.	.	.	K	.	.	.	E	.	.	.	.	.	.	.	.	.
	14	H09	HLA - A 1 29 HLA - B 27 44		.	.	.	R	.	.	.	E	.	.	.	.	.	.	.	.	.
	15	E36	HLA - A 2 0 HLA - B 44 0		K	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A
	16	B04	HLA - A 2 30 HLA - B 13 44		.	.	.	.	.	.	.	K	.	.	.	.	.	.	.	.	A
	17	H33	HLA - A 2 3 HLA - B 35 51		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	18	E04	HLA - A 2 3 HLA - B 35 35		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	19	E08	HLA - A 2 24 HLA - B 35 52		K	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	20	H01	HLA - A 2 3 HLA - B 7 35		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B*41 -	21	H31	HLA - A 2 68 HLA - B 7 35		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
(n=100)	22	I01	HLA - A 2 3 HLA - B 13 35		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	23	I02	HLA - A 2 0 HLA - B 35 52		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	24	M02	HLA - A 2 11 HLA - B 35 39		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	25	H24	HLA - A 2 32 HLA - B 7 18		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	26	E16	HLA - A 2 11 HLA - B 51 55		K	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	27	E28	HLA - A 2 24 HLA - B 38 51		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	28	H32	HLA - A 2 24 HLA - B 50 51		K	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	29	H38	HLA - A 2 26 HLA - B 39 51		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	30	I20	HLA - A 1 2 HLA - B 50 51		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	31	B16	HLA - A 2 23 HLA - B 49 50		K	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	32	E01	HLA - A 1 2 HLA - B 15 57		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	33	E11	HLA - A 2 68 HLA - B 15 0		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	34	E14	HLA - A 2 0 HLA - B 55 60																		

Figure 3.22 Analysis of HLA-B*41-associated viral sequence polymorphism throughout the whole ORF of the L-HD Ag.

(A) calculated *p* values for HLA-B*41-sequence association throughout the whole ORF of the L-HD Ag. (B) Amino acid sequence alignment of some of the studied isolates at the region around residues 139 and 158. Sequences of the isolates from HLA-B*41 positive patients are sorted on top and followed by those from HLA-B*41 negative patients. Amino acid residues with statistical evidence for selection are marked with a darker color (positions 139 and 158).

3.4.4. Molecular footprints revealed that HLA-B alleles may have stronger effect on viral evolution than locus-A alleles

Comprehensive analysis of association of amino acid polymorphisms and HLA alleles indicated several amino acid residues under selection pressure by particular HLA alleles. Table 3.10 summarizes the lowest *p* values corresponding to the associations between different HLA alleles (A and B of class I of MHC) and amino acid substitutions. Further *p* values regarding each residue/HLA allele are listed in the Appendix IV. As demonstrated below, the lowest *p*-value belonged to HLA-B*15 ($p= 10^{-6}$) for which the corresponding T cell response was detected in a HLA-B*15 positive patient with resolved HDV infection. Corresponding T cell response was also detected for HLA-B*27-restricted epitope. Further patients' samples are needed to investigate additional epitopes restricted by other HLA alleles for which we detected strong footprints.

Table 3.10 Amino acid residues within L-HD Ag with strong statistical evidence for selection by MHC class I alleles.

Locus	Allele	#	Position	Mutation	*p*-Value[a]
A-Alleles	HLA-A*29	1	63	K63R	**0.0002**
	HLA-A*30	1	49	P49L/S	**0.0022**
	HLA-A*68	1	81	V81I	**0.0036**
B-Alleles	HLA-B*13	1	33	D33E	**0.0001**
	HLA-B*13	2	113	K113R	**0.0043**
	HLA-B*14	1	107	A107T	**0.0028**
	HLA-B*15[b]	1	170	S170N	**10^-6**
	HLA-B*18	1	47	E47D	**0.0026**
	HLA-B*27[b]	1	105	R105K	**0.0028**
	HLA-B*37	1	89	P89T/I	**0.0033**
	HLA-B*37	2	101	D101E	**0.0005**
	HLA-B*38	1	131	K131G	**0.0039**
	HLA-B*41	1	71	A71T	**0.0065**
	HLA-B*41	2	139	R139K	**0.0034**
	HLA-B*41	3	158	G158A/D/M	**0.0023**
	HLA-B*49	1	37	V37A/T	**0.0015**
	HLA-B*51	1	81	V81I	**0.0018**
	HLA-B*52	1	12	G12N	**0.0052**
	HLA-B*52	2	203	D203E	**0.0052**

These data revealed that, generally, HLA-B alleles with the lowest *p* values have predominant impacts on viral evolution and viral evasion when compared with HLA-A alleles.

Chapter 4: Discussions

About 80-90% of HBV carriers superinfected with HDV are not able to clear the virus and, therefore, develop a chronic infection with both HBV and HDV. HDV-specific immune response may mediate the viral clearance in minority (10-20%) of these patients. The aim of this thesis, therefore, was to make a comprehensive analysis of the role of CD8+ T cell response in viral elimination and to evaluate a possible failure of this response due to viral genome variability or viral escape. To address this, we have studied a large cohort of patients with chronic HDV infection recruited from different medical centers, mostly located in Europe. A careful analysis of viral sequences obtained from these patients along with their HLA class I background allowed us to understand the correlation of the viral evolution with the intra familial transmission, length of the infection and particular HLA molecules.

Several aspects of this study including the role of HDV genome plasticity in determination of epitopes, the number of possible CD8+ T cell epitopes of HDV, the role of some particular HLA alleles in viral immune escape and finally application of viral escape mechanism for discovering novel epitopes are discussed in the following sections.

4.1. HDV sequence analysis and genotyping

4.1.1. Phylogenetic analysis indicates that HDV-1 is the dominant genotype in the studied isolates.

Distribution of evolutionary distances between different HDV isolates carried out in this study implies that probably the current classification of HDV into 8 clades may not completely reflect the true extent of divergence between all HDV isolates. The findings in this analysis are in some ways comparable with those found in HCV isolates and became a basis for the current classification of HCV genotypes [151]. Due to the low number of HDV isolates, especially HDV-2 to 8, when compared to those of HCV, this classification and corresponding hypothesis may need to be confirmed in a separate study with a larger dataset. Of note, access to HDV-1 isolates may not be difficult; however, it may be challenging when it comes to other HDV genotypes.

HDV-1 is the prevalent HDV genotype in most parts of the world [41, 152]. Although all studied isolates in this multi-centric project branched together with random representatives of thus far identified HDV-1 isolates, evolutionary distance analyses indicated that several isolates cluster closely together forming subpopulations of isolates in this study. These sub-clusters were mostly observed in isolates from a particular center/location (such as Spanish subjects) suggesting a close evolutionary relationship between these isolates. Moreover, there were several isolates from different geographic locations which also clustered together and showed very close evolutionary relationships. This may imply that migration has an

impact on viral evolution as well as turnover of HDV isolates in Europe and in the other parts of the world [60]. Since the origin of all studied subject was not recorded, this hypothesis remain to be investigated separately. Our analysis showed that distribution of evolutionary distances between all, so far, characterized HDV genotypes is not completely homogenous. While almost all clades showed an average percentage of similarity, the isolates from HDV-3, in our study, were the most divergent isolates from all other HDV clades which was in line with previous reports [9, 41, 42, 51].

The main message of this dissertation is that MHC class I-restricted CD8+ T cell responses are able to promote the evolution of HDV by selection of specific variants inside HDV epitopes. If this is correct, evolutionary analysis of the whole open reading frame of HD Ag (the only protein of HDV) may give us more reliable classification than using a partial ORF (generally used for genotyping so far). Some earlier studies, suggested that phylogenetic analysis of complete HDV genome sequences resulted in almost similar distribution of evolutionary distances when compared to those obtained from partial ORF [153] or S-HD Ag [51] fragments. Nevertheless, a comprehensive study on a larger dataset, comparing complete HDV genome, L-HD Ag, S-HD Ag and partial HD Ag, is needed to define which part of HDV genome must be analyzed to achieve a reliable and accurate genotyping data. Moreover, these studies may define the position of recombinant HDV isolates, already shown in several studies, in the final classification of HDV. Earlier studies suggested clinical correlations between HDV genotype and the level of severity of the symptoms [54-56]; nevertheless, whether this is the direct effect of HDV or the genotype of the helper virus (HBV) or the combination of both, needs further investigations.

4.1.2. High evolution rate at the non-synonymous sites indicated positive selection pressure

There are 61 amino acid codons (excluding stop codons), yet the total number of standard amino acids is only 20. This is due to the inherent redundancy of the genetic code, some amino acids employ more than one codon, *e.g.* Serine or Arginine can be encoded by 6 different genetic codons, whereas Methionine has only one codon, AUG. While Serine tolerates several substitutions in its codon, any changes (mutations) in Methionine's codon lead to replacement of Methionine with another amino acid. Collectively, mutations or substitutions in protein coding sequences can be either synonymous (silent) or non-synonymous (replacement). Synonymous changes do not lead to amino acid alteration; whereas the non-synonymous substitutions do. Most changes at the third position of a codon will be a silent mutation because it simply changes between the codons of the same amino acid, whereas mutations at the first or second position of a codon most likely result in replacement mutations. When the rate of non-synonymous (replacement) substitutions is

higher than that of synonymous (silent) changes, this indicates that selection acts stronger on replacement substitutions. Therefore, the ratio of the non-synonymous mutations (dN) to the synonymous mutations (dS) reflects the act of natural selection. By definition, all changes are neutral when dN/dS equals one, so any deviation from this indicates the action of selection, that is, when dN/dS <1, then non-synonymous mutations are vanishing faster by the act of negative selection, and if dN/dS >1, then non-synonymous mutations (replacement) are getting fixed faster indicating positive selection. There are multiple algorithms and methods for estimation of non-synonymous/synonymous substitution rates (dN/dS) which can lead to confusion in choosing an adequate method. It is recommended to make the calculation using as many methods as possible when the data size is limited; in contrast, it has been shown that in larger data sets, such as this multi-centric study with over 100 sequences, all methods will result in the same conclusion [154]. The calculated dN/dS ratios for the HDV isolates in this study indicate that several residues are under positive selection pressure.

It is important to note that not all silent mutations necessarily occur due to the negative selection because some of these changes may be selected due to their better action in formation of the secondary structure or replication fitness. Obviously, the unique rod-like structure of HDV RNA reflects the importance of RNA codon combination. Table 3.3 demonstrates the frequency of different codons for each amino acid calculated for the studied isolates which can probably be extended to that of all HDV-1 isolates. Of course, further investigations are needed to demystify the complexity of selection pressure due to the RNA structure. The combination and the order of the nucleotides define the secondary structures of HDV genome which then influence the replication capacity and ultimately the burden of the disease. Indeed, some earlier studies on HDV-3 isolates magnified the significance of branched versus unbranched rod like structure of HDV genomes and their functional impact on viral replication [155].

Of note, nucleotide misincorporation in HDV genome is not only caused by RNA polymerases, in fact, the impact of post-transcriptional editing action by ADAR on changes of single nucleotides is enormous and underestimated. It is believed that ADAR editing can damage the viral replication potency and lead to negative selection pressure. This was elegantly shown in three studies [156-158]. Consequently, defective substitutions leading to negative selection must be also compensated by changes rescuing viral replication [159]. Therefore, those variant with compensatory mutations will be positively selected. Taken together, evolution of HDV is extremely complex and therefore viral genome must show a high flexibility to overcome all selection pressures. Moreover, it is important to note that positive selection may not only occur due to the immune pressure but also happens to

maintain a high viral genome replication. Although a few number of studies consider positive selective pressure on HDV genome [156, 160], and even suggested that positive selection pressure is probably the main cause of HDV diversification [160], to our knowledge none of these studies addressed the question whether these selections were due to immune pressure. Therefore, the presented data of this study showed for the first time that positive selection of HDV variants also occurs under HDV-specific immune pressure. Positive selection may indeed play a crucial role in establishment of a chronic HDV infection. This hypothesis is supported in this study by experimental data. Interpretation of the data achieved exclusively from sequences, is challenging. In fact, if positive selection is detected based on sequence data, it is very difficult to pinpoint the reason behind this adaptation. Therefore, functional experiments must be done to address some of causality of the observed replacements. HDV, as mentioned throughout this dissertation, possesses only one ORF, and therefore, searching for residues under selection pressure is restricted to these 214 amino acids. By contrast, sequence-based analysis would be more complicated when it comes to the viruses with overlapping ORFs, *e.g.* HBV, where silent mutation in one ORF is a replacement substitution in another one. Although estimation of dN/dS ratio is not a very strong tool to address selection pressure, it is an easy method to look for the first evidence of natural selection and to reject the “neutrality” of nucleotide changes. Using this method, it was elegantly shown in HCV that the rate of non-synonymous mutations is significantly higher inside the CTL epitopes than that occurs outside these epitopes [161]. Of note, these findings were supported by experimental data in the same paper.

4.1.3. Long-term chronic HDV infection leads to the viral adaptation indicated by very slow viral evolution

Longitudinal study in some HDV patients of this study indicates that, rate of mutations may be related to the time of sampling after onset of infection. That is, escape mutations may not be detected in the early phase of infection where the rate of substitutions is relatively high and the substitutions are not completely fixed.

The longitudinal studies of patient samples provided novel insights and new information about evolution of HDV as an RNA virus in human. RNA viruses are known for their fast evolution and larger diversity when compared to DNA viruses. This probably relies on the lack of proof reading activity of RNA polymerases. Consequently, progeny viruses may not be all functional and replicative, so called defective viral RNA, which are not able to package or replicate without help of the functional or “wild type” viruses. It has been shown that in patients with chronic HDV infection, defective particles are indeed produced [162]. Mostly, defective viruses are resulted from insertion/deletion changes which in most cases lead to producing dysfunctional peptides or, because of frame shifted ORF, these variants are not

able to produce the HD Ag at all. Nevertheless, the focus of this study was on SNPs and their possible impact on infection outcome.

In conclusion, HDV may select for the most stable variant following the long-term infection in patients. These findings are consistent with some earlier studies of HCV infection of 2867 pregnant women known as the “single-source outbreak of HCV” or the “East-German anti-D cohort” where a large HCV outbreak occurred after administration of rhesus prophylaxis with contaminated immunoglobulin [163]. Although this cohort is known as the single-source outbreak, phylogenetic analysis proved the presence of at least 3 variants whose origins remain unclear. Clonal sequence analysis from some patients of this cohort after 30 years of chronic infection indicated the presence of only one variant in these patients [161]. These observations may suggest that after long-term infection, the fittest and the most adapted variant of a virus may take over other possible variants and guarantee an established chronic infection.

4.1.4. Deep sequencing analysis may provide a higher resolution at the region under selection pressure

In addition to data achieved by conventional sequencing, next generation sequencing such as UDPS, applied in this study, provides a high resolution data on possible or potential substitutions’ capacity at the region of interest. Obviously, it was difficult and also not necessary to perform deep sequencing for all isolates of this multi-centric cohort of patients; data analysis and interpretation of such large data sets would be a bit of bioinformatical challenge. In order to detect minor variants of the virus possibly affected by positive selection, we analyzed a section of L-HD Ag with focus on the region restricted to HLA-B*27 (L-HD Ag 99-112) with deep sequencing technology in a case/control manner. All cases of HLA-B*27 positive patients were compared with a control group of HLA-B*27 negative patients. Interestingly, we detected additional amino acid substitutions within this region in the case (HLA-B*27 positive) group indicating the impact of positive selection pressure on additional sites at lower frequencies. Although, positive selection pressure at these sites was detected at very low frequencies, this may indicate development major variants starting at very low level. This process may not be seen when the study is only focused on the major populations of the variants (detected by conventional sequencing).

Generally, mutations can be deleterious, neutral or beneficial. Variants of the virus with deleterious mutations are obviously removed by negative selection; therefore, they may, more likely, be detected at lower frequencies. Neutral and beneficial mutations, on the other hand, are those which obviously become fixed in the population and shape the genome.

4.2. Identification of HDV-specific CD8+ T cell epitopes

T cell epitope identification is probably the first and a key step in characterization of the magnitude of virus specific T cell response in patients. Therefore, we used a set of analyses and experiments to get some insight into the impact of T cell response on outcome of infection in HDV infected patients. Depending on the subset of T cells and HLA types, there are several approaches for epitope identification. Because of the main focus of this study on the HDV-specific CD8 positive T cells, we used several relevant approaches to identify HDV-specific CD8+ T cell targets in the studied cohort. In the following sections, these approaches and the corresponding results are discussed in details.

4.2.1. Using epitope prediction tools and MHC binding assay resulted in identification of two HLA-B*27-restricted HDV epitopes

MHC binding prediction tools and HLA affinity assays indicated that there are at least two HDV-specific HLA-B*27-restricted epitopes on L-HD Ag. These findings were confirmed by detection of corresponding responses in HLA-B*27 positive patients with resolved HDV infection. Indeed, all three HLA-B*27 positive patients with resolved HDV infection demonstrated HDV-specific CTL response to at least one of the characterized HLA-B*27-restricted epitopes which strongly supports the findings by prediction tools and binding assays. Interestingly, these epitopes were located in a region of L-HD Ag with a low amino acid variability which was in line with some earlier reports from HCV studies [164]. The significant roles of these HLA-B*27 epitopes and the corresponding T cell responses in viral elimination in chronic HDV-infected patients, need further investigations.

Fast growing body of literature on HLA-restricted epitopes, enriches the sources of databases used by prediction tools, resulting in more accurate and reliable epitope predictions. In addition, experimental data on binding affinities of MHC molecules leads to broader coverage of HLA alleles and more accurate predictions [165]. Since these additional data modifies the sources these tools are based on, consequently, the calculated percentile scores in this study may change upon recalculation [165]. There are slightly differences between prediction algorithms. Consensus method, mainly, used in this study was shown to be reliable in detection of T cell epitopes and, in particular, CD8+ T cell-restricted epitopes [166].

Although prediction methods are great tools in our hands to identify novel epitopes, they are not sufficient to pinpoint the exact minimal epitopes which are functional and immunogenic in individual patients. Therefore, additional experimental assays are needed to prove the results of the prediction methods. Several studies applied these epitope binding algorithms as a part of the whole process of detection of novel epitopes [106, 138, 167], yet not all of them

succeeded to prove the immunogenicity of all of the predicted epitopes indicating possible mispredictions by these tools. Nevertheless, this may not necessarily mean that those predicted epitopes are not immunogenic in other patients. In this direction, it is also important to note the subtype of the HLA allele of interest, for instance, in HCV infections it was elegantly shown that HLA-B*27 subtypes determine the specificity of HCV-related CTL responses [168].

Overall, epitope prediction and MHC binding assay are very restricted to the certain subtypes of certain HLA alleles. Therefore, overlapping these two methods provide rather narrow choices of HLA types. Interestingly, there was a very low number of HDV epitopes identified by using epitope prediction tools and MHC binding assay in this study. Significantly, these methods failed to identify HDV-specific CD8+ T cell epitopes for the most prevalent HLA alleles which are usually the center of attention when it comes to the immune therapy or vaccine design.

There are several epitope prediction tools available; however, one of the two approaches was used in this study (consensus approach), showed to be an updated approach in choosing the most reliable and accurate algorithm. We also retrieved the prediction scores based on SYFPEITHI database (http://www.syfpeithi.de) [146] and compared with the consensus method offered by IEDB. A clear consistency was observed between MHC binding affinities and percentile scores calculated by consensus method (IEDB), which was not the case for the corresponding scores calculated by SYFPEITHI. CD8+ T cell-restricted HDV epitopes have been described in only one study so far [138] where SYFPEITHI was used as the prediction tool. Using SYFPEITHI along with other approaches, Huang *et al.* introduced two HDV-specific CTL epitopes restricted to HLA-A*0201. In this study, HLA-A2.1 transgenic mice were immunized by plasmid coding for HDV Ag, then the splenocytes from these mice were stained by HLA-A*0201–peptide tetramer complexes which was constructed using six HDV peptides with high SYFPEITHI scores. These peptides include HDV 26–34, 43–51, 107–115, 114–122, 143–152 and 165–174. In this Taiwanese study, only 4 HLA-A*02 positive HDV infected patients were included: two active HDV infected (HDV RNA positive) plus two inactive HDV infected (anti-HDV positive/HDV RNA negative) patients. It is important to note that one patient from each group (with active and inactive HDV infections) was infected with HDV-2 which is also prevalent in Taiwan. In this experiment, they detected T cell specific responses for HDV 26–34 and HDV 43–51 peptides in patients with inactive HDV infection. These two peptides were tested along with the HDV library in our group and also by our collaborating group in Hannover, yet there was no detection of any specific T cell responses against these two epitopes in HLA-A*02 positive patients (data not shown). The *in silico* prediction method used by Huang *et al.,* showed low correlation with the binding and

T cell assay in our study. In fact, in the same study predicted ligands with high SYFPEITHI scores were also not able to promote a specific T cell response. Of note, these peptides were not detected in our approach of prediction methods, *e.g.*, consensus of Ann, Comblib_sidney2008 and Smm. Due to the fact that the most accurate epitope prediction methods are based on peptide data driven from experimental information about affinities to the corresponding MHC molecules, the reliability of the peptide-MHC binding prediction data driven from these methods is very dependent on the dataset used by these prediction tools [169]. The methods used in our study were also based on data from IEDB which is weekly updated [170]. Nevertheless, these data are relatively biased towards peptides' length of nine [145]. *NetMHCpan* method applies an approximation approach trained on 9-mer prediction tools, and may serve as a complementary method for prediction of ligands of lengths other than nine [145, 171].

The final proof whether a predicted and binding peptide is a functional and immunogenic CTL epitope, can be achieved by stimulating CD8+ T cells from patients, ideally, in acute phase or after recovery from infection. We could demonstrate that patients who were HDV infected, however lost HDV RNA in the follow-up, showed a significant IFN-γ production after stimulation with the peptide containing the predicted and binding epitopes restricted to HLA-B*27. These are the first and so far unique HLA epitopes described in European isolates of HDV.

The first evidence of immunogenicity of HDV-specific HLA-B*27-restricted CTL epitopes came from some earlier studies of a HLA-B*27 positive patient with resolved HDV infection. The PBMCs isolated from this patient were stimulated by the library of the HD Ag overlapping peptides. Interestingly, one of these overlapping peptides of 16-mers was able to induce IFN-γ production upon stimulation of T cells. This 16-mer included both tentative epitope, and it made difficult to distinguish exactly which fragment, as the minimal epitopes, mediated IFN-γ production. Moreover, these epitopes were not present in other overlapping peptides of this library of 16-mers overlapping by 8 amino acids. Using library of overlapping peptides for identification of virus-specific T cell response by stimulation of PBMCs may be a convenient and reliable approach for screening of patient samples; nevertheless, there are a few restrictions using this method for discovery of novel epitopes. As mentioned above, the number of amino acids by which the library peptides are overlapped, plays an important role in epitope identification; however, the higher number of amino acids by which the peptides are overlapped, the larger number of peptides within the library and consequently the larger number of cells needed for stimulation assay. It is usually a challenge and restriction when it comes to patient materials, such as PBMCs.

In conclusion, using library of peptide overlapped by, at least, 10 amino acids may ease the identification of novel epitopes using this approach. However, due to the relatively high variability of the viral protein which makes almost impossible to include all variant combinations, this may not be an attractive option. Adding this to the length and variability of the amino acid sequences, the number of overlapping peptides may increase dramatically. This is indeed one of the challenges for characterization and identification of novel epitopes using library overlapping peptides. Therefore, developing new approaches, such the one introduced in this dissertation (section 3.4.), may be of great interest.

Taken together, combination of *in silico* epitope prediction tools and MHC binding assay are able to be applied as screening approaches for identification of novel epitopes; however, due to the restricted number of HLA alleles covered by these two methods, some significant HLA molecules may be excluded. Therefore, supplementary approaches, also introduced in this dissertation, are needed for faster and more cost-effective identification of immunogenic epitopes for broader range of HLA alleles. Moreover, the identified epitopes will always need to be confirmed by detection of corresponding T cell responses in PBMCs of patients who carry the relevant HLA alleles. Finally, our study does not exclude the existence of other possible HLA class I-restricted CD8+ T cell epitopes in addition to the ones identified here. Therefore, further studies are needed for identification of possible epitopes restricted by other HLA alleles.

4.2.2. The total number of potential HDV-specific CTL epitopes within the ORF of HDV may be significantly lower than that of other hepatitis viruses such as HBV and HCV

Applying epitope prediction tools, MHC binding and T cell assays revealed a very low number of immunogenic epitopes on L-HD Ag. This may explain the low immunogenicity of HDV in comparison with other hepatitis viruses including HBV and HCV. However, the low number of predicted epitopes for HDV may be explained by the fact that most of these prediction algorithms have been trained by using the known epitope motives from the well-characterized viruses like HBV and HCV rather than from those poorly studied viruses, *e.g.* HDV. In other words, these prediction tools are developed based on the input data from the well-described epitopes. Therefore, failure in epitope identification for HDV using prediction tools may be due to the technical restrictions which may not exclude the existence of additional CTL epitopes.

Another explanation for the low number of predicted and identified HDV-specific epitopes may rely on the size of the viral protein (214aa). It is important to note that HDV has a very small genome of 1.7 kb which is the smallest genome when compared to other known viruses infecting human. For instance, HBV and HCV, two other human hepatitis viruses,

possess the genome of 3.2 and 9kb, respectively. Moreover, HDV genome encodes for only one viral protein of about 214 amino acids, whereas HCV genome, for instance, translates to a polyprotein of about 3000 amino acids. HBV, despite a relative small genome, has 4 overlapping ORFs, whereas HDV has only one. Taken together, HDV has a lower chance of having appropriate epitopes to be presented to the T cells when compared to other hepatitis viruses with larger proteins. According to our knowledge and the findings in this study, L-HD Ag may be processed and presented to the CD8+ T cells by a minority of HDV infected patients. Since to date there are not a high number of HDV-specific CD8+ T cell epitope characterized, we compared the number of predicted epitopes for HDV, HBV and HCV. L-HD Ag, as a nucleoprotein was compared with that of HBV and HCV. Interestingly, the total number of predicted MHC class I epitopes on L-HD Ag is significantly lower than that on nuclear proteins of HBV and HCV of almost similar size. In addition, it is important to note that L-HD Ag is the only protein of HDV whereas HBV and HCV have other proteins, which contain epitopes, making these two viruses more immunogenic than HDV. Hypothetically, the low number of HDV specific epitope may also rely on the origin of this virus. An early study on the origin of HDV proposed a cellular homolog for HD Ag [30], suggesting that HDV might have been evolved from a human transcriptome. This was consistent with a later report on the discovery of an HDV-like ribozyme in human genome, CPEB3 [26]. If this is correct and HDV (or HD Ag) is indeed evolved from a human cellular transcript, consequently, the low number of CTL epitopes on L-HD Ag may be due to the elimination of reactive T cells to the self-proteins.

4.2.3. HLA-B*27-restricted CTL response may support clearance of HDV infection

There is a general consensus that pathogen-specific CTL response plays a key role in fighting intracellular infectious agents including viruses and these responses may determine the final outcomes of the related diseases. In experimental HBV and HCV infections in chimpanzees, for instance, depletion of CD8+ cytotoxic T cells was associated with a prolonged infection when compared to the other animals where both main subsets of T cells (CD4+ and CD8+) were intact [172-174]. Several studies on HCV in human also suggested that an effective virus-specific CTL response may be associated with spontaneous viral clearance [164, 167, 175]. Since CTL response is dependent on presentation of viral epitopes by HLA molecules, genetic factors determine individual response to viral infections. In other words, people with different genetic background present different HLA alleles and consequently mount variable epitope-specific T cell response. Several HIV studies reported the critical influence of particular HLA alleles in progression of HIV infection to AIDS. While HLA-A*24 and HLA-B*35, for instance, were reported to be associated with a rather fast

progression to AIDS [176, 177], HLA-B*27 and -B*57 were described to be correlated to a significantly slower progression of the infection [177-181]. In HCV studies, several HLA types including HLA-A*03, HLA-B*15 and HLA-B*27 were reported to be associated with a spontaneous viral clearance [161, 164, 167, 182]. Indeed, a growing body of evidence over the past years suggests a broad antiviral activity of HLA-B*27 molecule against a variety of viral infections.

The knowledge on virus-specific T cell responses and MHC-associated viral clearance was vastly extended by the unique access to a cohort of women who were infected after inoculation with anti-D immunoglobulin contaminated with HCV genotype 1b [161, 164, 182]. Several studies have introduced this cohort as a single-source HCV outbreak; however, Timm *et al.* performed comprehensive phylogenetic analyses and indicated that there were at least three different variants of HCV possibly in the inoculum [161]. Therefore, this founder effect was considered in analysis of sequences isolated from patients after decades from the onset of infection. Despite the existence of three closely related variants, investigation of MHC-associated viral evolution is clearly easier than in usual cohorts of chronic patients, such as this cohort of chronic HDV infected patients where the sequence of infectious source is unknown.

It is not clear whether clearance of HDV is associated with multi-specific CD8+ T cell responses which, for instance, were comprehensively shown in studies of HCV infected patients [164]. While most of studies in the field of antiviral CTL response focus on HLA-A*02, our data and the growing body of literature indicate a clear dominance of HLA-B*27 over HLA-A*02, not only in hepatitis viruses but also in other viruses such as EBV or influenza virus infections.

According to several lines of evidence in our studies of HDV infected patients, we suggest that HLA-B*27-restricted CTL responses may play a protective role against HDV infection. However, this hypothesis needs to be proven in other studies of HDV infected patients with larger number of HLA-B*27 positive subjects. First, HDV-specific HLA-B*27-restricted CTL responses were detected in patients with resolved HDV infection. Second, these responses were not detectable in HLA-B*27 positive patients with ongoing chronic HDV infection, nor were they found in HLA-B*27 positive individuals without any serological markers for HDV infection or exposure. These findings were, however, consistent with those reported from HCV studies where HCV-specific CD8+ T cell responses were detected in resolvers but not in chronically infected patients [164, 183]. The next evidence comes from our study of immune escape within these two HDV-specific HLA-B*27-restricted epitopes and impairment of corresponding T cell responses due to these mutations. In fact, such T cell responses restricted to the wild type epitopes are so strong which are able to exert selection pressure

on the virus at this region with very low amino acid variability. Previous studies significantly supported the fact that immune escape mutations occur within protective epitopes [159, 161, 164, 167, 184, 185]. Viral clearance in the follow-up of a HLA-B*27 positive patient, with an ongoing HDV infection, was another significant observation implying the protective role of HDV-specific HLA-B*27-restricted T cell response in these patients. The unique access to the longitudinal samples of some of the patients within our cohort, allowed us to evaluate the kinetic of viral infection over time. Two chronically HDV-infected HLA-B*27 positive patients, without variation at the epitope restricted by HLA-B*27, resolved the infection during the follow up and demonstrated a detectable HDV-specific CD8+ T cell response (H36 and H20). These patients may not have a detectable CD8+ T cell response against HLA-B*27-restricted HDV epitope at the time of chronic and ongoing infection (patient H20 tested). This may suggest that patients who do not develop escape mutations in the critical epitope sequences may employ other mechanisms to avoid viral specific T cell responses. In HCV studies with respect to a dominant HLA-B*27-restricted CD8+ T cell responses, it was also shown that HCV-specific CD8+ T cell responses was not detectable in one chronically HCV-infected HLA-B*27 positive patient whose HCV isolate demonstrated no variation from the consensus sequence at the region restricted to HLA-B*27, implying other mechanisms of secondary failure of the virus-specific T cells [164]. Observation of HDV clearance in the follow up implies that longitudinal and prospective studies of acutely HDV-infected patients are needed in order to detect protective roles of particular HLA alleles or development of escape mutations in the process of establishment of viral persistence. Based on these observations, we hypothesized that HLA-B*27-restricted CTL responses may be correlated with control of HDV infection and possibly viral clearance. There is one rather important open question why despite the protective role of HLA-B*27 allele, some HLA-B*27 positive patients establish a chronic infection and are not able to control the viral infection. Several factors may be involved in this process: protection against HDV or other viral infections might be related to multiple HLA alleles including B*27; obviously, escape mutation may lead to the viral evasion; HLA-B*27 subtypes may play a significant role in patients with different genetic backgrounds [168]; protective HLA-B*27 responses might be (sub) genotype-specific [186].

4.3. Immune evasion of HDV from antiviral CTL response

Characterization of a successful virus-specific CD8+ T cell response is of great interest for T cell therapy to treat infectious diseases, especially for viruses such as HDV against which no direct acting antiviral agent is available. In addition, understanding the possible mechanisms causing failure of appropriate T cell responses is as important. With respect to HDV infection, here we discuss the impact of viral genome variability on failure of virus-specific immune

response. In other words, how the virus evades from the corresponding T cell response by amino acid substitutions within the key epitopes. Such amino acid substitutions are known as escape mutations. Generally, at least three main mechanisms have been described in this regard: a) one common mechanism of immune escape refers to the mutations occurring in the anchor residues of epitopes, critical for binding to the relevant MHC molecules; b) amino acid substitutions occurring beside the anchor-residues may have no direct impact on MHC binding affinity, but they may still be able to impair T cell recognition by affecting TCR binding to the peptide-MHC complex, in which case, mutated epitope presented by the corresponding HLA molecule is not recognized by the relevant T cells; c) amino acid substitutions are also able to alter post-transcriptional modification including proteasomal cleavage of viral protein inside the cytoplasm. This is a critical step in antigen presentation process by MHC molecules. Therefore, key epitopes may be disrupted if mutations within or even outside of these epitopes introduce a novel cutting site which is preferred by the proteasome compartment [103, 187, 188]. The significant role of post-transcriptional processing of viral peptides with respect to the HLA-B*27 was shown in a very recent study in influenza virus [189].

4.3.1. Evidence of HDV-specific CD8+ T cell selection pressure on HLA-B*27-restricted epitopes

HLA-B*27, as discussed in section 4.2.3. , was found to be a protective allele in several viral infections, possibly including HDV infection; therefore, we hypothesized that in HLA-B*27 positive patients with an ongoing chronic HDV infection there must be an escape mechanism involved to help the virus to persist. In order to address this issue, we studied a unique cohort of chronically HDV infected patients collected from different centers for this study. Indeed, we observed strong HLA-B*27-associated ‘footprints’ inside the novel characterized HLA-B*27 epitopes in the isolates from the group of HLA-B*27 positive patients. In a similar study in HCV, it was also shown that almost all chronically HCV-infected HLA-B*27 positive patients undergo immune escape mutations in the region restricted to the dominant HLA-B*27 epitope leading to impaired HCV-specific CD8+ T cell response. This was in contrast with the absence of variation in the same region in the isolates from HLA-B*27 negative patients indicating HLA-B*27-associated selection pressure [164].

In the first HLA-B*27-restricted epitope, L-HD Ag aa 99-108 (RRDHRRRKAL), we observed two substitutions, R105K and K106M, at the 7^{th} and the 8^{th} positions of this epitope, respectively. While R105K was exclusively observed in HLA-B*27 positive patients (in 40% of cases), K106M substitution was observed, in addition to the HLA-B*27 positive group, in only one out of 98 isolates from the HLA-B*27 negative group. It was unclear whether this mutation was already in the source of infection or evidence of a HLA-B*27 positive individual

in the family members of this patient. Although K106M mutation was also observed in an isolate from HLA-B*27 negative group, this was previously also reported to occur within a dominant HLA-B*27-resticetd HCV epitope where the HLA-B*27-associated mutation can be detected in HLA-B*27 negative group of HCV patients but in minority of the isolates [164]. Whether this mutation enforces strong fitness cost on HDV viral genome replication, is so far unclear. Longitudinal studies of the isolates from such patients may help to investigate the evolution of L-HD Ag at this position, as well as, a possible reversion to the wild type sequence which was reported in the case of HCV [190].

Two HLA-B*27-restricted HDV-specific epitopes were described in this study which are overlapping by 6 amino acids and, to our knowledge, so far unique. A protective HCV-epitope restricted by HLA-A*03 (NS3 1080-88 **TV**YHGAGTK) overlaps with a HLA-A*02-restricted HCV epitope (NS3 1073-81 CVNGVCW**TV**) in NS3 protein of HCV; these two epitopes share only two residues and no variation was reported within these two residues so far. In contrast, in our studies we observed the HLA-B*27-associated 'footprints' only within the 6 overlapping residues between both epitopes. Since we detected CTL responses to both epitopes in patients with resolved HDV infection, it is unclear whether one or both T cell responses are involved in the process of selection. In our study of HLA-B*27 positive resolvers, all three individuals were tested positive for the overlapping peptide (16-mer) including both epitopes. While there were no more PBMCs available from the first patient (Bonn Patient); therefore, it was not possible to test whether single or both HLA-B*27-restricted epitopes were able to induce HDV-specific T cell response in this patient. However, we were able to show that specific T cells can be induced by L-HD Ag 99-108 RRDHRRRKAL and 103-112 RRRKALENKK in the Freiburg patient and Hannover patient (H36), respectively. Interestingly, none of these patients demonstrated a specific response to both epitopes. This might be explained by the fact that these two epitopes are overlapping by 6 amino acids, processing towards one epitope may result in disrupting the other one. This phenomenon is perhaps dependent on the genetic background of the individuals and needs to be investigated separately. Unfortunately, the number of HLA-B*27 positive patients with resolved HDV infection was very limited; therefore, subsequent studies are needed to address these open issues.

A growing body of literature over the years demonstrated that immune escape mutations within the key epitopes of infectious agents such as HIV, HBV and HCV is a dominant mechanism of immune evasion from cytotoxic T cell response. In HCV studies, for instance, the first evidence of the impact of amino acid substitutions in the targeted epitopes on viral evasion came from experimental studies in chimpanzees [124, 125, 191]. In these studies, obviously, the genomic sequence of the virus in the inoculum was known. Subsequent

studies in human cohorts of HCV infected patients, where the source of infection was known, confirmed that HCV-specific T cell responses exert a significant selection pressure on the key epitopes which was in line with those findings in animal models [126, 192]. Amino acid substitutions within epitopes are, obviously, the result of a higher rate of non-synonymous over synonymous substitutions. In studies of HCV, it was shown that the rate of non-synonymous mutations inside the epitopes was significantly higher than that outside those epitopes which clearly supports the hypothesis of positive selection pressure by virus-specific CD8+ T cells. Because of the low number of described CTL epitopes within L-HD Ag, it was not possible to address whether this phenomenon occurs in the case of HDV infection.

Our cross-sectional study of chronically HDV infected patients indicated two HLA-B*27-associated escape mutations (R105K and K106M). It was unclear whether these substitutions were emerged under HLA-B*27-mediated pressure or they were dominantly preexisting in the sources of infection. An access to the longitudinal samples of those patients can help to address this issue. The preexisting substitutions may have profound effects on protective roles of particular HLA alleles, *e.g.* B*27. It was shown that HLA-B*27 is a protective allele in HCV genotype 1b [164, 182]; however, preexisting substitutions within the protective alleles restricted by HLA-A*03 and HLA0-B*27 in the isolates of anti-D outbreak of HCV genotype 1b, resulted in neutral outcome of these protective alleles in this study population [193].

4.3.2. Amino acid substitutions associated with HLA-B*27 impair the relevant HDV-specific T cell responses

Next we asked whether the observed amino acid substitutions within the described HLA-B*27 epitopes are recognized by the corresponding T cells. Therefore, we tested the mutated variants of those epitopes to stimulate the PBMCs isolated from patients who recovered from HDV infection. Interestingly, the mutated variants were unable to induce IFN-γ production. This strongly confirms that the mutations have occurred due to the positive selection pressure by HLA-B*27-restricted HDV-specific T cell responses. These responses were so effective and strong to eliminate the viruses carrying the ‘wild type’ sequence; therefore, the remaining variants of the virus, which contained escape mutations, were not recognized and supported the persistent HDV infection. This hypothetical process, however, needs to be confirmed in a separate study.

As presented in section 3.3.2. , we observed polymorphism at the position 100 of L-HD Ag (2^{nd} residue of the HLA-B*27-restricted epitope, L-HD Ag_{99-108}) which apparently was not related to immunosuppression by HLA-B*27, because this polymorphism was observed in HLA-B*27 negative patients as well. The amino acid residues at this position included Q, R, E and K, from high to low prevalence, respectively. Indeed, the epitope with all these four

substitutions tested positive in T cell assay; however, epitopes with R and Q, which were also the most prevalent residues, demonstrated slightly higher induction of CD8+ T cells for IFN-γ production than other substitutions, K and E. Obviously, Glutamine (Q) in our study of conventional sanger sequencing was the most prevalent residue at this position; nevertheless, a separate deep sequencing study at this region indicated frequent substitution by Arginine (R) at this position. This implies that direct sequencing may not reflect the magnitude of viral genome variability and probably its impact on T cell response in patients. It is unclear which variant of this epitope (R**R**DHRRRKAL or R**Q**DHRRRKAL) drives a stronger response in patients. Based on these observations, both of these variants were used to make the peptides including HLA-B*27-associated substitutions, R105K and K106M. Interestingly, HLA-B*27-associated variants of the epitopes impaired the T cell recognition, regardless R or Q at the position L-HD Ag 100. Of note, it seemed that T cell response was not completely impaired by K106M variant of RRDHRRRKAL when compared with R105K variants. This may indicate some levels of T cell cross-reactivity for the epitopes with K106M substitution. On the other hand, K106M substitution in the other HLA-B*27 epitope (RRRKALENKK) was able to impair the respective T cell response almost completely, whereas the R105K variant of the same epitope had rather no effect on the corresponding T cell response. However, R105K mutation in the 9-mer (104-112 RRKALENKK) was able to impair the respective response to this 9-mer. Our findings were consistent with some earlier studies of HCV where the escape mutations in the epitopes restricted by HLA-B*27 [164] and HLA-A*03 [167] were able to impair the respective immunodominant T cell responses.

Taken together, HLA-B*27-associated escape mutations at residues 105 and 106 (R105K & K106M) were strongly able to impair the T cell response in contrast to polymorphism at aa position 100, *i.e.*, Q, R, K or E. Thus far reported ligands for HLA-B*27 emphasize on R at the second residue (anchor position); we report here at least three additional possible amino acid residues, Q, K and E, supported by experimental data. Nevertheless, Glutamine (Q) at second position of HLA-B*27 epitopes has been reported in minority of natural ligands of HLA-B*27 [194], especially in the ligands with Arginine at the first position like the one identified in this study. Of note, the first Arginine residue (R) in L-HD $Ag_{99\text{-}108}$ epitope was conserved not only in genotype 1 but also throughout all genotypes implying a significant cross-genotypic biological role of this residue at this position. Further studies are needed in order to determine replication fitness or the impact on viral load for the variants of the virus with escape mutations.

4.3.3. Amino acid substitutions selected due to the immune pressure and not founder effect

While amino acid substitutions are correlated to some immune pressure, it is important to note the founder effect which may have a great impact on the final conclusion. Founder effect is defined by relatively close variants of a virus, sharing some amino acid sequences because of common lineage, and therefore, these variants should not be handled as individualistic. This may, especially, be observed when the cohort of patients is exclusively recruited from one population or one medical center. A larger data set or multi-centric data sources, as the one collected for this study, may decline the impact of this lineage effect on the final conclusions. There are pieces of evidence that the amino acid substitutions detected in this study are correlated with immune escape and not due to the founder effect: (1) the epitopes with substituted amino acids reproducibly impaired the corresponding T cell response; (2) amino acid substitutions at the escape residues were not genotype or subtype specific; (3) replaced amino acids correlated with escape were detected in patients with various origins, *e.g.*, R105K was detected in two HLA-B*27 positive patients from Spain and Germany or in the case of HLA-B*15, the amino acid substitution (S170N) was observed in 8 HLA-B*15 positive patients from 4 different collaborating centers (Essen, n=4; Hannover, n=2, Barcelona, n=1, Italy, n=1). The latter evidence is rather more important than the first two simply because it is less likely that the patients with different background and origin would share the similar lineage of the virus. Taken together, founder effect may eventually hamper the application of footprints for identification of T cell targets in a population with a relatively high frequency of certain HLA alleles. In such environment the fixed escape mutation may establish a new variant of the circulating virus and appear in different individuals with slightly different HLA backgrounds.

4.4. Sequence analysis in concert with HLA alleles can be considered as a novel approach in identification of HDV-specific CTL epitopes

The so far mentioned methods may not be able to detect all possible HDV-specific epitopes. Therefore, we applied a sequence-led approach to identify additional T cell epitopes restricted by HLA class I alleles. This approach is especially applicable when there is an access to a large dataset of viral sequences as well as the corresponding HLA alleles of the studied subjects, ideally, collected from different regions with different variants and HLA backgrounds. This method of novel epitope identification using HLA footprints was recently used for HBV [114] and HCV [161]. Obviously, this approach is biased toward those epitopes that undergo selection pressure mediated by HLA class I-restricted CD8+ T cells. That is,

sequence analysis approach may not be adequate for identification of epitopes in critical regions of the genes where no mutations are allowed or when mutations occur outside of the epitope sequence or at flanking regions [159]. This approach of epitope identification may also be not useful for identification of epitopes located in hyper-variable regions. Simply because the variations occur in all isolates often regardless of the HLA background of the subjects. Therefore, detection of HLA footprints in these regions is almost impossible. This was the case in our study of the identified HLA-A*02 epitopes [138] located in regions with a high amino acid variability, HDV 26-34 and 43-51, so that we were not able to address whether these epitopes are under selection pressure by HLA-A*02 molecules. Consequently, the major restriction for identification of epitopes through searching for HLA footprints is that this approach only works for those epitopes that continuously undergo mutations mediated by T cell pressure. Nevertheless, those virus-specific T cell responses which are that strong to promote viral evolution may imply the effectivity of T cell responses at these selected regions and indicate these as possible immune targets [159].

Strength of sequence-HLA association has a direct correlation with the number of isolates (sequences) with the relevant HLA allele. For instance, in our cohort of 104 chronic HDV infected patients with HLA types, there were only 6 HLA-B*27 positive isolates 50% of which carried immune escape mutations within the region restricted by this allele. In contrast, 100% of HLA-B*15 positive isolates (8 subjects) demonstrated immune escape substitution within the novel identified HLA-B*15 epitope. Consequently, the correlation of amino acid substitution with HLA-B*15 allele found to be significantly stronger than that of HLA-B*27. Molecular footprints were also detected in correlation with the isolates from patients with other HLA backgrounds such as HLA-B*37 (n=3) and HLA-B*41 (n=4) which are less frequent alleles. Although the *p* values calculated for the correlation of footprints with the HLA alleles with very low frequencies (*e.g.* B*37) were statistically significant, functional impact of such substitutions can only be confirmed by corresponding T cell assays.

Comprehensive analysis of molecular footprints through the whole open reading frame of L-HD Ag in correlation with all available 35 HLA alleles in this study revealed that the selection pressures are exerted predominantly by HLA-B alleles rather than HLA-A alleles. This finding was consistent with the growing body of evidence in the field [161, 195]. This may be due to the faster evolution of HLA-B alleles than A alleles [196].

In conclusion, HLA-B alleles play a significant role in viral elimination and viral evolution; this may provide a new direction of research of HDV in the future and in the field of vaccine design.

References

1. **Rizzetto, M., et al.**, *Immunofluorescence detection of new antigen-antibody system (delta/anti-delta) associated to hepatitis B virus in liver and in serum of HBsAg carriers.* Gut, **1977**. 18(12): p. 997-1003.
2. **Rizzetto, M., et al.**, *Transmission of the hepatitis B virus-associated delta antigen to chimpanzees.* J Infect Dis, **1980**. 141(5): p. 590-602.
3. **Ponzetto, A., et al.**, *Transmission of the hepatitis B virus-associated delta agent to the eastern woodchuck.* Proc Natl Acad Sci U S A, **1984**. 81(7): p. 2208-12.
4. **Rizzetto, M., et al.**, *delta Agent: association of delta antigen with hepatitis B surface antigen and RNA in serum of delta-infected chimpanzees.* Proc Natl Acad Sci U S A, **1980**. 77(10): p. 6124-8.
5. **Rizzetto, M.**, *The Delta Agent.* Hepatology, **1983**. 3(5): p. 729-737.
6. **Branch, A.D. and H.D. Robertson**, *A replication cycle for viroids and other small infectious RNA's.* Science, **1984**. 223(4635): p. 450-5.
7. **Kos, A., et al.**, *The hepatitis delta (delta) virus possesses a circular RNA.* Nature, **1986**. 323(6088): p. 558-60.
8. **Wang, K.S., et al.**, *Structure, sequence and expression of the hepatitis delta (delta) viral genome.* Nature, **1986**. 323(6088): p. 508-14.
9. **Radjef, N., et al.**, *Molecular phylogenetic analyses indicate a wide and ancient radiation of African hepatitis delta virus, suggesting a deltavirus genus of at least seven major clades.* J Virol, **2004**. 78(5): p. 2537-44.
10. **Sureau, C.**, *The role of the HBV envelope proteins in the HDV replication cycle.* Curr Top Microbiol Immunol, **2006**. 307: p. 113-31.
11. **Chen, P.J., et al.**, *Structure and replication of the genome of the hepatitis delta virus.* Proc Natl Acad Sci U S A, **1986**. 83(22): p. 8774-8.
12. **Hsieh, S.Y., et al.**, *Hepatitis delta virus genome replication: a polyadenylated mRNA for delta antigen.* J Virol, **1990**. 64(7): p. 3192-8.
13. **Lo, K., G.T. Sheu, and M.M. Lai**, *Inhibition of Cellular RNA polymerase II transcription by delta antigen of hepatitis delta virus.* Virology, **1998**. 247(2): p. 178-88.
14. **Chang, F.L., et al.**, *The large form of hepatitis delta antigen is crucial for assembly of hepatitis delta virus.* Proc Natl Acad Sci U S A, **1991**. 88(19): p. 8490-4.
15. **Glenn, J.S., et al.**, *Identification of a prenylation site in delta virus large antigen.* Science, **1992**. 256(5061): p. 1331-3.
16. **Hwang, S.B., C.Z. Lee, and M.M. Lai**, *Hepatitis delta antigen expressed by recombinant baculoviruses: comparison of biochemical properties and post-*

translational modifications between the large and small forms. Virology, **1992**. 190(1): p. 413-22.

17. **Casey, J.L. and J.L. Gerin**, *Hepatitis D virus RNA editing: specific modification of adenosine in the antigenomic RNA.* J Virol, **1995**. 69(12): p. 7593-600.
18. **Polson, A.G., B.L. Bass, and J.L. Casey**, *RNA editing of hepatitis delta virus antigenome by dsRNA-adenosine deaminase.* Nature, **1996**. 380(6573): p. 454-6.
19. **Casey, J.L. and J.L. Gerin**, *The woodchuck model of HDV infection.* Curr Top Microbiol Immunol, **2006**. 307: p. 211-25.
20. **Casey, J.L.**, *RNA editing in hepatitis delta virus.* Curr Top Microbiol Immunol, **2006**. 307: p. 67-89.
21. **Luo, G.X., et al.**, *A specific base transition occurs on replicating hepatitis delta virus RNA.* J Virol, **1990**. 64(3): p. 1021-7.
22. **Sureau, C., et al.**, *Cloned hepatitis delta virus cDNA is infectious in the chimpanzee.* J Virol, **1989**. 63(10): p. 4292-7.
23. **Kuo, M.Y., et al.**, *Characterization of self-cleaving RNA sequences on the genome and antigenome of human hepatitis delta virus.* J Virol, **1988**. 62(12): p. 4439-44.
24. **Sharmeen, L., et al.**, *Antigenomic RNA of human hepatitis delta virus can undergo self-cleavage.* J Virol, **1988**. 62(8): p. 2674-9.
25. **Ferre-D'Amare, A.R., K. Zhou, and J.A. Doudna**, *Crystal structure of a hepatitis delta virus ribozyme.* Nature, **1998**. 395(6702): p. 567-74.
26. **Salehi-Ashtiani, K., et al.**, *A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene.* Science, **2006**. 313(5794): p. 1788-92.
27. **Webb, C.H., et al.**, *Widespread occurrence of self-cleaving ribozymes.* Science, **2009**. 326(5955): p. 953.
28. **Lai, M.M.**, *RNA replication without RNA-dependent RNA polymerase: surprises from hepatitis delta virus.* J Virol, **2005**. 79(13): p. 7951-8.
29. **Reid, C.E. and D.W. Lazinski**, *A host-specific function is required for ligation of a wide variety of ribozyme-processed RNAs.* Proc Natl Acad Sci U S A, **2000**. 97(1): p. 424-9.
30. **Brazas, R. and D. Ganem**, *A cellular homolog of hepatitis delta antigen: implications for viral replication and evolution.* Science, **1996**. 274(5284): p. 90-4.
31. **Chao, M., S.Y. Hsieh, and J. Taylor**, *Role of two forms of hepatitis delta virus antigen: evidence for a mechanism of self-limiting genome replication.* J Virol, **1990**. 64(10): p. 5066-9.

32. **Gudima, S., et al.**, *Parameters of human hepatitis delta virus genome replication: the quantity, quality, and intracellular distribution of viral proteins and RNA.* J Virol, **2002**. 76(8): p. 3709-19.
33. **Ryu, W.S., et al.**, *Ribonucleoprotein complexes of hepatitis delta virus.* J Virol, **1993**. 67(6): p. 3281-7.
34. **Zeisel, M.B., et al.**, *Towards an HBV cure: state-of-the-art and unresolved questions--report of the ANRS workshop on HBV cure.* Gut, **2015**. 64(8): p. 1314-26.
35. **Gudima, S., et al.**, *Two potentially important elements of the hepatitis B virus large envelope protein are dispensable for the infectivity of hepatitis delta virus.* J Virol, **2007**. 81(8): p. 4343-7.
36. **Sureau, C., B. Guerra, and R.E. Lanford**, *Role of the large hepatitis B virus envelope protein in infectivity of the hepatitis delta virion.* J Virol, **1993**. 67(1): p. 366-72.
37. **Ryu, W.S., M. Bayer, and J. Taylor**, *Assembly of hepatitis delta virus particles.* J Virol, **1992**. 66(4): p. 2310-5.
38. **Yan, H., et al.**, *Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis B and D virus.* Elife, **2014**. 3.
39. **Netter, H.J., K. Kajino, and J.M. Taylor**, *Experimental transmission of human hepatitis delta virus to the laboratory mouse.* J Virol, **1993**. 67(6): p. 3357-62.
40. **Lutgehetmann, M., et al.**, *Humanized chimeric uPA mouse model for the study of hepatitis B and D virus interactions and preclinical drug evaluation.* Hepatology, **2012**. 55(3): p. 685-94.
41. **Casey, J.L., et al.**, *A genotype of hepatitis D virus that occurs in northern South America.* Proc Natl Acad Sci U S A, **1993**. 90(19): p. 9016-20.
42. **Le Gal, F., et al.**, *Eighth major clade for hepatitis delta virus.* Emerg Infect Dis, **2006**. 12(9): p. 1447-50.
43. **Shakil, A.O., et al.**, *Geographic distribution and genetic variability of hepatitis delta virus genotype I.* Virology, **1997**. 234(1): p. 160-7.
44. **Chao, Y.C., et al.**, *Sequence conservation and divergence of hepatitis delta virus RNA.* Virology, **1990**. 178(2): p. 384-92.
45. **Han, M., et al.**, *Molecular epidemiology of hepatitis delta virus in the Western Pacific region.* J Clin Virol, **2014**. 61(1): p. 34-9.
46. **Zhang, Y.Y., E. Tsega, and B.G. Hansson**, *Phylogenetic analysis of hepatitis D viruses indicating a new genotype I subgroup among African isolates.* J Clin Microbiol, **1996**. 34(12): p. 3023-30.

47. **Imazeki, F., M. Omata, and M. Ohto**, *Heterogeneity and evolution rates of delta virus RNA sequences.* J Virol, **1990**. 64(11): p. 5594-9.
48. **Lee, C.M., et al.**, *Characterization of a new genotype II hepatitis delta virus from Taiwan.* J Med Virol, **1996**. 49(2): p. 145-54.
49. **Ivaniushina, V., et al.**, *Hepatitis delta virus genotypes I and II cocirculate in an endemic area of Yakutia, Russia.* J Gen Virol, **2001**. 82(Pt 11): p. 2709-18.
50. **Chang, S.Y., et al.**, *Molecular epidemiology of hepatitis D virus infection among injecting drug users with and without human immunodeficiency virus infection in Taiwan.* J Clin Microbiol, **2011**. 49(3): p. 1083-9.
51. **Deny, P.**, *Hepatitis delta virus genetic variability: from genotypes I, II, III to eight major clades?* Curr Top Microbiol Immunol, **2006**. 307: p. 151-71.
52. **Barros, L.M., et al.**, *Hepatitis Delta virus genotype 8 infection in Northeast Brazil: inheritance from African slaves?* Virus Res, **2011**. 160(1-2): p. 333-9.
53. **Parana, R., et al.**, *Serial transmission of spongiocytic hepatitis to woodchucks (possible association with a specific delta strain).* J Hepatol, **1995**. 22(4): p. 468-73.
54. **Wu, J.C., et al.**, *Genotyping of hepatitis D virus by restriction-fragment length polymorphism and relation to outcome of hepatitis D.* Lancet, **1995**. 346(8980): p. 939-41.
55. **Manock, S.R., et al.**, *An outbreak of fulminant hepatitis delta in the Waorani, an indigenous people of the Amazon basin of Ecuador.* Am J Trop Med Hyg, **2000**. 63(3-4): p. 209-13.
56. **Nakano, T., et al.**, *Characterization of hepatitis D virus genotype III among Yucpa Indians in Venezuela.* J Gen Virol, **2001**. 82(Pt 9): p. 2183-9.
57. **Gaeta, G.B., et al.**, *Chronic hepatitis D: a vanishing Disease? An Italian multicenter study.* Hepatology, **2000**. 32(4 Pt 1): p. 824-7.
58. **Hadziyannis, S.J.**, *Decreasing prevalence of hepatitis D virus infection.* J Gastroenterol Hepatol, **1997**. 12(11): p. 745-6.
59. **Cross, T.J., et al.**, *The increasing prevalence of hepatitis delta virus (HDV) infection in South London.* J Med Virol, **2008**. 80(2): p. 277-82.
60. **Wedemeyer, H., B. Heidrich, and M.P. Manns**, *Hepatitis D virus infection--not a vanishing disease in Europe!* Hepatology, **2007**. 45(5): p. 1331-2; author reply 1332-3.
61. **Rizzetto, M.**, *Hepatitis D: thirty years after.* J Hepatol, **2009**. 50(5): p. 1043-50.
62. **Rizzetto, M. and S.M. Alavian**, *Hepatitis delta: the rediscovery.* Clin Liver Dis, **2013**. 17(3): p. 475-87.

63. **Smedile, A., et al.**, *Influence of delta infection on severity of hepatitis B.* Lancet, **1982**. 2(8305): p. 945-7.
64. **Buti, M., et al.**, *Serological diagnosis of acute delta hepatitis.* J Med Virol, **1986**. 18(1): p. 81-5.
65. **Grippon, P., et al.**, *Long-term delta antigenaemia without appearance of delta antibody in two immunodeficient patients.* Lancet, **1987**. 1(8540): p. 1031.
66. **Ottobrelli, A., et al.**, *Patterns of hepatitis delta virus reinfection and disease in liver transplantation.* Gastroenterology, **1991**. 101(6): p. 1649-55.
67. **Smedile, A., et al.**, *Hepatitis D viremia following orthotopic liver transplantation involves a typical HDV virion with a hepatitis B surface antigen envelope.* Hepatology, **1998**. 27(6): p. 1723-9.
68. **Giersch, K., et al.**, *Persistent hepatitis D virus mono-infection in humanized mice is efficiently converted by hepatitis B virus to a productive co-infection.* J Hepatol, **2014**. 60(3): p. 538-44.
69. **Kuo, M.Y., M. Chao, and J. Taylor**, *Initiation of replication of the human hepatitis delta virus genome from cloned DNA: role of delta antigen.* J Virol, **1989**. 63(5): p. 1945-50.
70. **Farci, P., et al.**, *Acute and chronic hepatitis delta virus infection: direct or indirect effect on hepatitis B virus replication?* J Med Virol, **1988**. 26(3): p. 279-88.
71. **Krogsgaard, K., et al.**, *Delta-infection and suppression of hepatitis B virus replication in chronic HBsAg carriers.* Hepatology, **1987**. 7(1): p. 42-5.
72. **Farci, P., et al.**, *Delta hepatitis in inapparent carriers of hepatitis B surface antigen. A disease simulating acute hepatitis B progressive to chronicity.* Gastroenterology, **1983**. 85(3): p. 669-73.
73. **Rizzetto, M., et al.**, *Chronic hepatitis in carriers of hepatitis B surface antigen, with intrahepatic expression of the delta antigen. An active and progressive disease unresponsive to immunosuppressive treatment.* Ann Intern Med, **1983**. 98(4): p. 437-41.
74. **Rosina, F., et al.**, *Changing pattern of chronic hepatitis D in Southern Europe.* Gastroenterology, **1999**. 117(1): p. 161-6.
75. **Fattovich, G., et al.**, *Influence of hepatitis delta virus infection on morbidity and mortality in compensated cirrhosis type B. The European Concerted Action on Viral Hepatitis (Eurohep).* Gut, **2000**. 46(3): p. 420-6.
76. **Romeo, R., et al.**, *A 28-year study of the course of hepatitis Delta infection: a risk factor for cirrhosis and hepatocellular carcinoma.* Gastroenterology, **2009**. 136(5): p. 1629-38.

77. **Farci, P., et al.**, *Treatment of chronic hepatitis delta virus (HDV) infection with human lymphoblastoid alpha interferon.* Q J Med, **1989**. 73(271): p. 1045-54.
78. **Rizzetto, M., et al.**, *Treatment of chronic delta hepatitis with alpha-2 recombinant interferon.* J Hepatol, **1986**. 3 Suppl 2: p. S229-33.
79. **Nucci, M.L., R. Shorr, and A. Abuchowski**, *Modified Therapeutic Proteins The therapeutic value of poly(ethylene glycol)-modified proteins.* Advanced Drug Delivery Reviews, **1991**. 6(2): p. 133-151.
80. **Hershfield, M.S., et al.**, *Treatment of adenosine deaminase deficiency with polyethylene glycol-modified adenosine deaminase.* N Engl J Med, **1987**. 316(10): p. 589-96.
81. **Garripoli, A., et al.**, *Ribavirin treatment for chronic hepatitis D: a pilot study.* Liver, **1994**. 14(3): p. 154-7.
82. **Gunsar, F., et al.**, *Two-year interferon therapy with or without ribavirin in chronic delta hepatitis.* Antivir Ther, **2005**. 10(6): p. 721-6.
83. **Yurdaydin, C., et al.**, *Treatment of chronic delta hepatitis with lamivudine vs lamivudine + interferon vs interferon.* J Viral Hepat, **2008**. 15(4): p. 314-21.
84. **Wedemeyer, H., et al.**, *Peginterferon plus adefovir versus either drug alone for hepatitis delta.* N Engl J Med, **2011**. 364(4): p. 322-31.
85. **Kabacam, G., et al.**, *Entecavir treatment of chronic hepatitis D.* Clin Infect Dis, **2012**. 55(5): p. 645-50.
86. **Farci, P., et al.**, *Treatment of chronic hepatitis D with interferon alfa-2a.* N Engl J Med, **1994**. 330(2): p. 88-94.
87. **Erhardt, A., et al.**, *Treatment of chronic hepatitis delta with pegylated interferon-alpha2b.* Liver Int, **2006**. 26(7): p. 805-10.
88. **Abbas, Z., et al.**, *Treatment of chronic hepatitis D patients with pegylated interferon: a real-world experience.* Antivir Ther, **2014**. 19(5): p. 463-8.
89. **Barrera, A., et al.**, *Mapping of the hepatitis B virus pre-S1 domain involved in receptor recognition.* J Virol, **2005**. 79(15): p. 9786-98.
90. **Gripon, P., et al.**, *Myristylation of the hepatitis B virus large surface protein is essential for viral infectivity.* Virology, **1995**. 213(2): p. 292-9.
91. **Ni, Y., et al.**, *Hepatitis B and D viruses exploit sodium taurocholate co-transporting polypeptide for species-specific entry into hepatocytes.* Gastroenterology, **2014**. 146(4): p. 1070-83.
92. **Urban, S., et al.**, *Strategies to inhibit entry of HBV and HDV into hepatocytes.* Gastroenterology, **2014**. 147(1): p. 48-64.

93. **Petersen, J., et al.**, *Prevention of hepatitis B virus infection in vivo by entry inhibitors derived from the large envelope protein.* Nat Biotechnol, **2008**. 26(3): p. 335-41.
94. **Volz, T., et al.**, *The entry inhibitor Myrcludex-B efficiently blocks intrahepatic virus spreading in humanized mice previously infected with hepatitis B virus.* J Hepatol, **2013**. 58(5): p. 861-7.
95. **Blank, A., et al.**, *First-in-human application of the novel hepatitis B and hepatitis D virus entry inhibitor myrcludex B.* J Hepatol, **2016**. 65(3): p. 483-9.
96. **Bogomolov, P., et al.**, *Treatment of chronic hepatitis D with the entry inhibitor myrcludex B: First results of a phase Ib/IIa study.* J Hepatol, **2016**. 65(3): p. 490-8.
97. **Bordier, B.B., et al.**, *A prenylation inhibitor prevents production of infectious hepatitis delta virus particles.* J Virol, **2002**. 76(20): p. 10465-72.
98. **Bordier, B.B., et al.**, *In vivo antiviral efficacy of prenylation inhibitors against hepatitis delta virus.* J Clin Invest, **2003**. 112(3): p. 407-14.
99. **Koh, C., et al.**, *Oral prenylation inhibition with lonafarnib in chronic hepatitis D infection: a proof-of-concept randomised, double-blind, placebo-controlled phase 2A trial.* Lancet Infect Dis, **2015**.
100. **Iwasaki, A. and R. Medzhitov**, *Toll-like receptor control of the adaptive immune responses.* Nat Immunol, **2004**. 5(10): p. 987-95.
101. **Theofilopoulos, A.N., et al.**, *Type I interferons (alpha/beta) in immunity and autoimmunity.* Annu Rev Immunol, **2005**. 23: p. 307-36.
102. **Perry, A.K., et al.**, *The host type I interferon response to viral and bacterial infections.* Cell Res, **2005**. 15(6): p. 407-22.
103. **Timm, J. and C.M. Walker**, *Mutational escape of CD8+ T cell epitopes: implications for prevention and therapy of persistent hepatitis virus infections.* Med Microbiol Immunol, **2015**. 204(1): p. 29-38.
104. **Feinstone, S.M., A.Z. Kapikian, and R.H. Purceli**, *Hepatitis A: detection by immune electron microscopy of a viruslike antigen associated with acute illness.* Science, **1973**. 182(4116): p. 1026-8.
105. **Daniels, D., S. Grytdal, and A. Wasley**, *Surveillance for acute viral hepatitis - United States, 2007.* MMWR Surveill Summ, **2009**. 58(3): p. 1-27.
106. **Schulte, I., et al.**, *Characterization of CD8+ T-cell response in acute and resolved hepatitis A virus infection.* J Hepatol, **2011**. 54(2): p. 201-8.
107. **Schweitzer, A., et al.**, *Estimations of worldwide prevalence of chronic hepatitis B virus infection: a systematic review of data published between 1965 and 2013.* Lancet, **2015**.

108. **Kosinska, A.D., et al.**, *Combination of DNA prime--adenovirus boost immunization with entecavir elicits sustained control of chronic hepatitis B in the woodchuck model.* PLoS Pathog, **2013**. 9(6): p. e1003391.
109. **Liu, J., et al.**, *Enhancing virus-specific immunity in vivo by combining therapeutic vaccination and PD-L1 blockade in chronic hepadnaviral infection.* PLoS Pathog, **2014**. 10(1): p. e1003856.
110. **Dunn, C., et al.**, *Temporal analysis of early immune responses in patients with acute hepatitis B virus infection.* Gastroenterology, **2009**. 137(4): p. 1289-300.
111. **Bertoletti, A., et al.**, *Cytotoxic T lymphocyte response to a wild type hepatitis B virus epitope in patients chronically infected by variant viruses carrying substitutions within the epitope.* J Exp Med, **1994**. 180(3): p. 933-43.
112. **Bertoletti, A., et al.**, *Natural variants of cytotoxic epitopes are T-cell receptor antagonists for antiviral cytotoxic T cells.* Nature, **1994**. 369(6479): p. 407-10.
113. **Desmond, C.P., et al.**, *Viral adaptation to host immune responses occurs in chronic hepatitis B virus (HBV) infection, and adaptation is greatest in HBV e antigen-negative disease.* J Virol, **2012**. 86(2): p. 1181-92.
114. **Kefalakes, H., et al.**, *Adaptation of the hepatitis B virus core protein to CD8(+) T-cell selection pressure.* Hepatology, **2015**. 62(1): p. 47-56.
115. **Hoofnagle, J.H.**, *Course and outcome of hepatitis C.* Hepatology, **2002**. 36(5 Suppl 1): p. S21-9.
116. **Gower, E., et al.**, *Global epidemiology and genotype distribution of the hepatitis C virus infection.* J Hepatol, **2014**. 61(1 Suppl): p. S45-57.
117. **Rosen, H.R., et al.**, *Frequencies of HCV-specific effector CD4+ T cells by flow cytometry: correlation with clinical disease stages.* Hepatology, **2002**. 35(1): p. 190-8.
118. **Diepolder, H.M., et al.**, *Possible mechanism involving T-lymphocyte response to non-structural protein 3 in viral clearance in acute hepatitis C virus infection.* Lancet, **1995**. 346(8981): p. 1006-7.
119. **Erickson, A.L., et al.**, *Hepatitis C virus-specific CTL responses in the liver of chimpanzees with acute and chronic hepatitis C.* J Immunol, **1993**. 151(8): p. 4189-99.
120. **Wedemeyer, H., et al.**, *Impaired effector function of hepatitis C virus-specific CD8+ T cells in chronic hepatitis C virus infection.* J Immunol, **2002**. 169(6): p. 3447-58.
121. **Golden-Mason, L., et al.**, *Upregulation of PD-1 expression on circulating and intrahepatic hepatitis C virus-specific CD8+ T cells associated with reversible immune dysfunction.* J Virol, **2007**. 81(17): p. 9249-58.

122. **Nakamoto, N., et al.**, *Functional restoration of HCV-specific CD8 T cells by PD-1 blockade is defined by PD-1 expression and compartmentalization.* Gastroenterology, **2008**. 134(7): p. 1927-37, 1937.e1-2.
123. **Raziorrouh, B., et al.**, *Inhibitory molecules that regulate expansion and restoration of HCV-specific CD4+ T cells in patients with chronic infection.* Gastroenterology, **2011**. 141(4): p. 1422-31, 1431.e1-6.
124. **Weiner, A., et al.**, *Persistent hepatitis C virus infection in a chimpanzee is associated with emergence of a cytotoxic T lymphocyte escape variant.* Proc Natl Acad Sci U S A, **1995**. 92(7): p. 2755-9.
125. **Erickson, A.L., et al.**, *The outcome of hepatitis C virus infection is predicted by escape mutations in epitopes targeted by cytotoxic T lymphocytes.* Immunity, **2001**. 15(6): p. 883-95.
126. **Cox, A.L., et al.**, *Cellular immune selection with hepatitis C virus persistence in humans.* J Exp Med, **2005**. 201(11): p. 1741-52.
127. **Guilhot, S., et al.**, *Expression of the hepatitis delta virus large and small antigens in transgenic mice.* J Virol, **1994**. 68(2): p. 1052-8.
128. **Pugnale, P., et al.**, *Hepatitis delta virus inhibits alpha interferon signaling.* Hepatology, **2009**. 49(2): p. 398-406.
129. Rizzetto, M., et al., *Hepatitis delta virus infection*. 1984, Grune & Stratton: Orlando, FL. p. 371-379.
130. **Borghesio, E., et al.**, *Serum immunoglobulin M antibody to hepatitis D as a surrogate marker of hepatitis D in interferon-treated patients and in patients who underwent liver transplantation.* Hepatology, **1998**. 27(3): p. 873-6.
131. **Farci, P., et al.**, *Diagnostic and prognostic significance of the IgM antibody to the hepatitis delta virus.* Jama, **1986**. 255(11): p. 1443-6.
132. **Negro, F., et al.**, *Reappearance of hepatitis D virus (HDV) replication in chronic hepatitis B virus carrier chimpanzees rechallenged with HDV.* J Infect Dis, **1989**. 160(4): p. 567-71.
133. **Fiedler, M., et al.**, *Immunization of woodchucks (Marmota monax) with hepatitis delta virus DNA vaccine.* Vaccine, **2001**. 19(32): p. 4618-26.
134. **Karayiannis, P., et al.**, *Immunization of woodchucks with recombinant hepatitis delta antigen does not protect against hepatitis delta virus infection.* Hepatology, **1990**. 12(5): p. 1125-8.
135. **Karayiannis, P., et al.**, *Partial control of hepatitis delta virus superinfection by immunisation of woodchucks (Marmota monax) with hepatitis delta antigen expressed by a recombinant vaccinia or baculovirus.* J Med Virol, **1993**. 41(3): p. 210-4.

136. **Nisini, R., et al.**, *Human CD4+ T-cell response to hepatitis delta virus: identification of multiple epitopes and characterization of T-helper cytokine profiles.* J Virol, **1997**. 71(3): p. 2241-51.
137. **Mauch, C., et al.**, *Induction of cytotoxic T lymphocyte responses against hepatitis delta virus antigens which protect against tumor formation in mice.* Vaccine, **2001**. 20(1-2): p. 170-80.
138. **Huang, Y.H., et al.**, *Identification of novel HLA-A*0201-restricted CD8+ T-cell epitopes on hepatitis delta virus.* J Gen Virol, **2004**. 85(Pt 10): p. 3089-98.
139. **D'Ugo, E., et al.**, *Immunization of woodchucks with adjuvanted sHDAg (p24): immune response and outcome following challenge.* Vaccine, **2004**. 22(3-4): p. 457-66.
140. **Fiedler, M. and M. Roggendorf**, *Immunology of HDV infection.* Curr Top Microbiol Immunol, **2006**. 307: p. 187-209.
141. **Heinemann, F.M.**, *HLA Genotyping and Antibody Characterization Using the Luminex Multiplex Technology.* Transfus Med Hemother, **2009**. 36(4): p. 273-278.
142. **Larkin, M.A., et al.**, *Clustal W and Clustal X version 2.0.* Bioinformatics, **2007**. 23(21): p. 2947-8.
143. **Tamura, K., et al.**, *MEGA6: Molecular Evolutionary Genetics Analysis version 6.0.* Mol Biol Evol, **2013**. 30(12): p. 2725-9.
144. **Gregori, J., et al.**, *Inference with viral quasispecies diversity indices: clonal and NGS approaches.* Bioinformatics, **2014**. 21: p. 21.
145. **Nielsen, M. and M. Andreatta**, *NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets.* Genome Med, **2016**. 8(1): p. 33.
146. **Rammensee, H., et al.**, *SYFPEITHI: database for MHC ligands and peptide motifs.* Immunogenetics, **1999**. 50(3-4): p. 213-9.
147. **Rodenko, B., et al.**, *Generation of peptide-MHC class I complexes through UV-mediated ligand exchange.* Nat Protoc, **2006**. 1(3): p. 1120-32.
148. **Kern, F., et al.**, *T-cell epitope mapping by flow cytometry.* Nat Med, **1998**. 4(8): p. 975-8.
149. **Budeus, B., J. Timm, and D. Hoffmann**, *SeqFeatR for the Discovery of Feature-Sequence Associations.* PLoS One, **2016**. 11(1): p. e0146409.
150. **Fisher, R.A.**, *On the Interpretation of χ2 from from Contingency Tables, and the Calculation of P.* Journal of the Royal Statistical Society, **1922**. 85(1): p. 87-94.

151. **Simmonds, P., et al.**, *Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region.* J Gen Virol, **1993**. 74 (Pt 11): p. 2391-9.
152. **Niro, G.A., et al.**, *The predominance of hepatitis delta virus genotype I among chronically infected Italian patients.* Hepatology, **1997**. 25(3): p. 728-34.
153. **Celik, I., et al.**, *Complete genome sequences and phylogenetic analysis of hepatitis delta viruses isolated from nine Turkish patients.* Arch Virol, **2011**. 156(12): p. 2215-20.
154. **Pond, S.L. and S.D. Frost**, *A genetic algorithm approach to detecting lineage-specific variation in selection pressure.* Mol Biol Evol, **2005**. 22(3): p. 478-85.
155. **Linnstaedt, S.D., et al.**, *The role of a metastable RNA secondary structure in hepatitis delta virus genotype III RNA editing.* Rna, **2006**. 12(8): p. 1521-33.
156. **Chang, J., S.O. Gudima, and J.M. Taylor**, *Evolution of hepatitis delta virus RNA genome following long-term replication in cell culture.* J Virol, **2005**. 79(21): p. 13310-6.
157. **Netter, H.J., et al.**, *Nucleotide sequence stability of the genome of hepatitis delta virus.* J Virol, **1995**. 69(3): p. 1687-92.
158. **Polson, A.G., et al.**, *Hepatitis delta virus RNA editing is highly specific for the amber/W site and is suppressed by hepatitis delta antigen.* Mol Cell Biol, **1998**. 18(4): p. 1919-26.
159. **Ruhl, M., et al.**, *Escape from a dominant HLA-B*15-restricted CD8+ T cell response against hepatitis C virus requires compensatory mutations outside the epitope.* J Virol, **2012**. 86(2): p. 991-1000.
160. **Anisimova, M. and Z. Yang**, *Molecular evolution of the hepatitis delta virus antigen gene: recombination or positive selection?* J Mol Evol, **2004**. 59(6): p. 815-26.
161. **Ruhl, M., et al.**, *CD8+ T-cell response promotes evolution of hepatitis C virus nonstructural proteins.* Gastroenterology, **2011**. 140(7): p. 2064-73.
162. **Wu, J.C., et al.**, *"Defective" mutations of hepatitis D viruses in chronic hepatitis D patients.* World J Gastroenterol, **2005**. 11(11): p. 1658-62.
163. **Wiese, M., et al.**, *Outcome in a hepatitis C (genotype 1b) single source outbreak in Germany--a 25-year multicenter study.* J Hepatol, **2005**. 43(4): p. 590-8.
164. **Neumann-Haefelin, C., et al.**, *Dominant influence of an HLA-B27 restricted CD8+ T cell response in mediating HCV clearance and evolution.* Hepatology, **2006**. 43(3): p. 563-572.
165. **Kim, Y., et al.**, *Immune epitope database analysis resource.* Nucleic Acids Res, **2012**. 40(Web Server issue): p. W525-30.

166. **Moutaftsi, M., et al.**, *A consensus epitope prediction approach identifies the breadth of murine T(CD8+)-cell responses to vaccinia virus.* Nat Biotechnol, **2006**. 24(7): p. 817-9.
167. **Fitzmaurice, K., et al.**, *Molecular footprints reveal the impact of the protective HLA-A*03 allele in hepatitis C virus infection.* Gut, **2011**. 60(11): p. 1563-71.
168. **Nitschke, K., et al.**, *HLA-B*27 subtype specificity determines targeting and viral evolution of a hepatitis C virus-specific CD8+ T cell epitope.* J Hepatol, **2014**. 60(1): p. 22-9.
169. **Kim, Y., et al.**, *Dataset size and composition impact the reliability of performance benchmarks for peptide-MHC binding predictions.* BMC Bioinformatics, **2014**. 15: p. 241.
170. **Trolle, T., et al.**, *Automated benchmarking of peptide-MHC class I binding predictions.* Bioinformatics, **2015**. 31(13): p. 2174-81.
171. **Lundegaard, C., O. Lund, and M. Nielsen**, *Accurate approximation method for prediction of class I MHC affinities for peptides of length 8, 10 and 11 using prediction tools trained on 9mers.* Bioinformatics, **2008**. 24(11): p. 1397-8.
172. **Shoukry, N.H., A.G. Cawthon, and C.M. Walker**, *Cell-mediated immunity and the outcome of hepatitis C virus infection.* Annu Rev Microbiol, **2004**. 58: p. 391-424.
173. **Shoukry, N.H., et al.**, *Memory CD8+ T cells are required for protection from persistent hepatitis C virus infection.* J Exp Med, **2003**. 197(12): p. 1645-55.
174. **Thimme, R., et al.**, *CD8(+) T cells mediate viral clearance and disease pathogenesis during acute hepatitis B virus infection.* J Virol, **2003**. 77(1): p. 68-76.
175. **Kuniholm, M.H., et al.**, *Specific human leukocyte antigen class I and II alleles associated with hepatitis C virus viremia.* Hepatology, **2010**. 51(5): p. 1514-22.
176. **Carrington, M., et al.**, *HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage.* Science, **1999**. 283(5408): p. 1748-52.
177. **Mann, D.L., et al.**, *Major histocompatibility complex genotype is associated with disease progression and virus load levels in a cohort of human immunodeficiency virus type 1-infected Caucasians and African Americans.* J Infect Dis, **1998**. 178(6): p. 1799-802.
178. **den Uyl, D., I.E. van der Horst-Bruinsma, and M. van Agtmael**, *Progression of HIV to AIDS: a protective role for HLA-B27?* AIDS Rev, **2004**. 6(2): p. 89-96.
179. **O'Brien, S.J., X. Gao, and M. Carrington**, *HLA and AIDS: a cautionary tale.* Trends Mol Med, **2001**. 7(9): p. 379-81.
180. **Gao, X., et al.**, *AIDS restriction HLA allotypes target distinct intervals of HIV-1 pathogenesis.* Nat Med, **2005**. 11(12): p. 1290-2.

181. **Kaslow, R.A., et al.**, *Polymorphisms in HLA class I genes associated with both favorable prognosis of human immunodeficiency virus (HIV) type 1 infection and positive cytotoxic T-lymphocyte responses to ALVAC-HIV recombinant canarypox vaccines.* J Virol, **2001**. 75(18): p. 8681-9.
182. **McKiernan, S.M., et al.**, *Distinct MHC class I and II alleles are associated with hepatitis C viral clearance, originating from a single source.* Hepatology, **2004**. 40(1): p. 108-14.
183. **Takaki, A., et al.**, *Cellular immune responses persist and humoral responses decrease two decades after recovery from a single-source outbreak of hepatitis C.* Nat Med, **2000**. 6(5): p. 578-82.
184. **Iglesias, M.C., et al.**, *Escape from highly effective public CD8+ T-cell clonotypes by HIV.* Blood, **2011**. 118(8): p. 2138-49.
185. **Shahid, A., et al.**, *Consequences of HLA-B*13-Associated Escape Mutations on HIV-1 Replication and Nef Function.* J Virol, **2015**. 89(22): p. 11557-71.
186. **Neumann-Haefelin, C., et al.**, *Protective effect of human leukocyte antigen B27 in hepatitis C virus infection requires the presence of a genotype-specific immunodominant CD8+ T-cell epitope.* Hepatology, **2010**. 51(1): p. 54-62.
187. **Kimura, Y., et al.**, *Escape mutations alter proteasome processing of major histocompatibility complex class I-restricted epitopes in persistent hepatitis C virus infection.* J Virol, **2005**. 79(8): p. 4870-6.
188. **Seifert, U., et al.**, *Hepatitis C virus mutation affects proteasomal epitope processing.* J Clin Invest, **2004**. 114(2): p. 250-9.
189. **Akram, A., et al.**, *HLA-B27, but not HLA-B7, immunodominance to influenza is ERAP dependent.* J Immunol, **2014**. 192(12): p. 5520-8.
190. **Timm, J., et al.**, *CD8 epitope escape and reversion in acute HCV infection.* J Exp Med, **2004**. 200(12): p. 1593-604.
191. **Cooper, S., et al.**, *Analysis of a successful immune response against hepatitis C virus.* Immunity, **1999**. 10(4): p. 439-49.
192. **Tester, I., et al.**, *Immune evasion versus recovery after acute hepatitis C virus infection from a shared source.* J Exp Med, **2005**. 201(11): p. 1725-31.
193. **Ziegler, S., et al.**, *Susceptibility to chronic hepatitis C virus infection is influenced by sequence differences in immunodominant CD8+ T cell epitopes.* J Hepatol, **2013**. 58(1): p. 24-30.
194. **Lopez de Castro, J.A., et al.**, *HLA-B27: a registry of constitutive peptide ligands.* Tissue Antigens, **2004**. 63(5): p. 424-45.

195. **Kiepiela, P., et al.**, *Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA.* Nature, **2004**. 432(7018): p. 769-75.

196. **McAdam, S.N., et al.**, *A uniquely high level of recombination at the HLA-B locus.* Proc Natl Acad Sci U S A, **1994**. 91(13): p. 5893-7.

Appendix

Appendix I: Genomic DNA purification protocol

Blood or Body Fluid Spin Protocol

Protocol: DNA Purification from Blood or Body Fluids (Spin Protocol)

This protocol is for purification of total (genomic, mitochondrial, and viral) DNA from whole blood, plasma, serum, buffy coat, lymphocytes, and body fluids using a microcentrifuge. For total DNA purification using a vacuum manifold, see "Protocol: DNA Purification from Blood or Body Fluids (Vacuum Protocol)" on page 29.

Important points before starting

- All centrifugation steps are carried out at room temperature (15–25°C).
- Use carrier DNA if the sample contains <10,000 genome equivalents (see page 17).
- 200 µl of whole blood yields 3–12 µg of DNA. Preparation of buffy coat (see page 18) is recommended if a higher yield is required.

Things to do before starting

- Equilibrate samples to room temperature (15–25°C).
- Heat a water bath or heating block to 56°C for use in step 4.
- Equilibrate Buffer AE or distilled water to room temperature for elution in step 11.
- Ensure that Buffer AW1, Buffer AW2, and QIAGEN Protease have been prepared according to the instructions on page 16.
- If a precipitate has formed in Buffer AL, dissolve by incubating at 56°C.

Procedure

1. **Pipet 20 µl QIAGEN Protease (or proteinase K) into the bottom of a 1.5 ml microcentrifuge tube.**
2. **Add 200 µl sample to the microcentrifuge tube. Use up to 200 µl whole blood, plasma, serum, buffy coat, or body fluids, or up to 5 x 10^6 lymphocytes in 200 µl PBS.**

   If the sample volume is less than 200 µl, add the appropriate volume of PBS.

   QIAamp Mini spin columns copurify RNA and DNA when both are present in the sample. RNA may inhibit some downstream enzymatic reactions, but not PCR. If RNA-free genomic DNA is required, 4 µl of an RNase A stock solution (100 mg/ml) should be added to the sample before addition of Buffer AL.

   Note: It is possible to add QIAGEN Protease (or proteinase K) to samples that have already been dispensed into microcentrifuge tubes. In this case, it is important to ensure proper mixing after adding the enzyme.

3. **Add 200 µl Buffer AL to the sample. Mix by pulse-vortexing for 15 s.**

   To ensure efficient lysis, it is essential that the sample and Buffer AL are mixed thoroughly to yield a homogeneous solution.

   If the sample volume is larger than 200 µl, increase the amount of QIAGEN Protease (or proteinase K) and Buffer AL proportionally; for example, a 400 µl sample will require 40 µl QIAGEN Protease (or proteinase K) and 400 µl Buffer AL. If sample volumes larger than 400 µl are required, use of QIAamp DNA Blood Midi or Maxi Kits is recommended; these can process up to 2 ml or up to 10 ml of sample, respectively.

   Note: Do not add QIAGEN Protease or proteinase K directly to Buffer AL.

4. **Incubate at 56°C for 10 min.**

   DNA yield reaches a maximum after lysis for 10 min at 56°C. Longer incubation times have no effect on yield or quality of the purified DNA.

5. **Briefly centrifuge the 1.5 ml microcentrifuge tube to remove drops from the inside of the lid.**

6. **Add 200 µl ethanol (96–100%) to the sample, and mix again by pulse-vortexing for 15 s. After mixing, briefly centrifuge the 1.5 ml microcentrifuge tube to remove drops from the inside of the lid.**

   If the sample volume is greater than 200 µl, increase the amount of ethanol proportionally; for example, a 400 µl sample will require 400 µl of ethanol.

7. **Carefully apply the mixture from step 6 to the QIAamp Mini spin column (in a 2 ml collection tube) without wetting the rim. Close the cap, and centrifuge at 6000 x *g* (8000 rpm) for 1 min. Place the QIAamp Mini spin column in a clean 2 ml collection tube (provided), and discard the tube containing the filtrate.***

   Close each spin column to avoid aerosol formation during centrifugation.

   Centrifugation is performed at 6000 x *g* (8000 rpm) to reduce noise. Centrifugation at full speed will not affect the yield or purity of the DNA. If the lysate has not completely passed through the column after centrifugation, centrifuge again at higher speed until the QIAamp Mini spin column is empty.

   Note: When preparing DNA from buffy coat or lymphocytes, centrifugation at full speed is recommended to avoid clogging.

8. **Carefully open the QIAamp Mini spin column and add 500 µl Buffer AW1 without wetting the rim. Close the cap and centrifuge at 6000 x *g* (8000 rpm) for 1 min. Place the QIAamp Mini spin column in a clean 2 ml collection tube (provided), and discard the collection tube containing the filtrate.***

   It is not necessary to increase the volume of Buffer AW1 if the original sample volume is larger than 200 µl.

* Flow-through contains Buffer AL or Buffer AW1 and is therefore not compatible with bleach. See page 6 for safety information.

9. **Carefully open the QIAamp Mini spin column and add 500 µl Buffer AW2 without wetting the rim. Close the cap and centrifuge at full speed (20,000 x *g*; 14,000 rpm) for 3 min.**

10. **Recommended: Place the QIAamp Mini spin column in a new 2 ml collection tube (not provided) and discard the old collection tube with the filtrate. Centrifuge at full speed for 1 min.**

    This step helps to eliminate the chance of possible Buffer AW2 carryover.

11. **Place the QIAamp Mini spin column in a clean 1.5 ml microcentrifuge tube (not provided), and discard the collection tube containing the filtrate. Carefully open the QIAamp Mini spin column and add 200 µl Buffer AE or distilled water. Incubate at room temperature (15–25°C) for 1 min, and then centrifuge at 6000 x *g* (8000 rpm) for 1 min.**

    Incubating the QIAamp Mini spin column loaded with Buffer AE or water for 5 min at room temperature before centrifugation generally increases DNA yield.

    A second elution step with a further 200 µl Buffer AE will increase yields by up to 15%.

    Volumes of more than 200 µl should not be eluted into a 1.5 ml microcentrifuge tube because the spin column will come into contact with the eluate, leading to possible aerosol formation during centrifugation.

    Elution with volumes of less than 200 µl increases the final DNA concentration in the eluate significantly, but slightly reduces the overall DNA yield (see Table 5, page 25). For samples containing less than 1 µg of DNA, elution in 50 µl Buffer AE or water is recommended. Eluting with 2 x 100 µl instead of 1 x 200 µl does not increase elution efficiency.

    For long-term storage of DNA, eluting in Buffer AE and storing at –20°C is recommended, since DNA stored in water is subject to acid hydrolysis.

    A 200 µl sample of whole human blood (approximately 5 x 10^6 leukocytes/ml) typically yields 6 µg of DNA in 200 µl water (30 ng/µl) with an A_{260}/A_{280} ratio of 1.7–1.9.

    For more information about elution and how to determine DNA yield, purity, and length, refer to pages 24–25 and Appendix A, page 50.

Appendix II: Purification of PCR products by Gel extraction

QIAquick Gel Extraction Kit Protocol

using a microcentrifuge

This protocol is designed to extract and purify DNA of 70 bp to 10 kb from standard or low-melt agarose gels in TAE or TBE buffer. Up to 400 mg agarose can be processed per spin column. This kit can also be used for DNA cleanup from enzymatic reactions (see page 8). For DNA cleanup from enzymatic reactions using this protocol, add 3 volumes of Buffer QG and 1 volume of isopropanol to the reaction, mix, and proceed with step 6 of the protocol. Alternatively, use the new MinElute Reaction Cleanup Kit.

Notes:
- The yellow color of Buffer QG indicates a pH ≤7.5.
- Add ethanol (96–100%) to Buffer PE before use (see bottle label for volume).
- Isopropanol (100%) and a heating block or water bath at 50°C are required.
- All centrifugation steps are carried out at ≥10,000 x *g* (~13,000 rpm) in a conventional table-top microcentrifuge.
- 3 M sodium acetate, pH 5.0, may be necessary.

1. **Excise the DNA fragment from the agarose gel with a clean, sharp scalpel.**

   Minimize the size of the gel slice by removing extra agarose.

2. **Weigh the gel slice in a colorless tube. Add 3 volumes of Buffer QG to 1 volume of gel (100 mg ~ 100 µl).**

   For example, add 300 µl of Buffer QG to each 100 mg of gel. For >2% agarose gels, add 6 volumes of Buffer QG. The maximum amount of gel slice per QIAquick column is 400 mg; for gel slices >400 mg use more than one QIAquick column.

3. **Incubate at 50°C for 10 min (or until the gel slice has completely dissolved). To help dissolve gel, mix by vortexing the tube every 2–3 min during the incubation.**

   IMPORTANT: Solubilize agarose completely. For >2% gels, increase incubation time.

4. **After the gel slice has dissolved completely, check that the color of the mixture is yellow (similar to Buffer QG without dissolved agarose).**

   If the color of the mixture is orange or violet, add 10 µl of 3 M sodium acetate, pH 5.0, and mix. The color of the mixture will turn to yellow.

   The adsorption of DNA to the QIAquick membrane is efficient only at pH ≤7.5. Buffer QG contains a pH indicator which is yellow at pH ≤7.5 and orange or violet at higher pH, allowing easy determination of the optimal pH for DNA binding.

5. **Add 1 gel volume of isopropanol to the sample and mix.**

   For example, if the agarose gel slice is 100 mg, add 100 µl isopropanol. This step increases the yield of DNA fragments <500 bp and >4 kb. For DNA fragments between 500 bp and 4 kb, addition of isopropanol has no effect on yield. Do not centrifuge the sample at this stage.

6. **Place a QIAquick spin column in a provided 2 ml collection tube.**
7. **To bind DNA, apply the sample to the QIAquick column, and centrifuge for 1 min.**

   The maximum volume of the column reservoir is 800 µl. For sample volumes of more than 800 µl, simply load and spin again.
8. **Discard flow-through and place QIAquick column back in the same collection tube.**

   Collection tubes are re-used to reduce plastic waste.
9. **(Optional): Add 0.5 ml of Buffer QG to QIAquick column and centrifuge for 1 min.**

   This step will remove all traces of agarose. It is only required when the DNA will subsequently be used for direct sequencing, in vitro transcription or microinjection.
10. **To wash, add 0.75 ml of Buffer PE to QIAquick column and centrifuge for 1 min.**

    Note: If the DNA will be used for salt sensitive applications, such as blunt-end ligation and direct sequencing, let the column stand 2–5 min after addition of Buffer PE, before centrifuging.
11. **Discard the flow-through and centrifuge the QIAquick column for an additional 1 min at ≥10,000 x *g* (~13,000 rpm).**

    IMPORTANT: Residual ethanol from Buffer PE will not be completely removed unless the flow-through is discarded before this additional centrifugation.
12. **Place QIAquick column into a clean 1.5 ml microcentrifuge tube.**
13. **To elute DNA, add 50 µl of Buffer EB (10 mM Tris·Cl, pH 8.5) or H_2O to the center of the QIAquick membrane and centrifuge the column for 1 min at maximum speed. Alternatively, for increased DNA concentration, add 30 µl elution buffer to the center of the QIAquick membrane, let the column stand for 1 min, and then centrifuge for 1 min.**

    IMPORTANT: Ensure that the elution buffer is dispensed directly onto the QIAquick membrane for complete elution of bound DNA. The average eluate volume is 48 µl from 50 µl elution buffer volume, and 28 µl from 30 µl.

    Elution efficiency is dependent on pH. The maximum elution efficiency is achieved between pH 7.0 and 8.5. When using water, make sure that the pH value is within this range, and store DNA at –20°C as DNA may degrade in the absence of a buffering agent. The purified DNA can also be eluted in TE (10 mM Tris·Cl, 1 mM EDTA, pH 8.0), but the EDTA may inhibit subsequent enzymatic reactions.

Appendix III: IUPAC nucleotide codes

Single-letter Code	Nucleotide/s	Explanation
A	A	Adenine
C	C	Cytosine
G	G	Guanine
T	T	Thymine
I	I	Inosine
R	A or G	pu**R**ine
Y	C or T	p**Y**rimidine
M	A or C	a**M**ino
K	G or T	**K**eto
S	C or G	**S**trong interaction
W	A or T	**W**eak interaction
H	A or C or T	not G, H follows G in alphabet
B	C or G or T	not A, B follows A in alphabet
V	A or C or G	not T/U, V follows U in alphabet
D	A or G or T	not C, Dfollows C in alphabet
N	A or C or G or T	aNy

Appendix IV: List of HLA- HDV sequence associations

HLA-A*01

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	0.5317401	L	87	1	K	130	1	L	173	1	S
2	1	S	45	1	E	88	0.5317401	R	131	0.1503590	K	174	0.3050730	I
3	1	Q	46	0.6804567	D	89	0.3653846	I	132	0.2786916	R	175	1	P
4	0.4809178	S	47	0.5317401	G	90	0.1878768	H	133	1	L	176	1	E
5	1	E	48	1	?	91	0.6218285	K	134	0.5523986	A	177	1	S
6	0.1312546	A	49	0.1312546	T	92	0.3653846	R	135	0.1556140	K	178	1	P
7	0.2525894	R	50	1	W	93	0.5285952	G	136	0.6501914	D	179	1	Y
8	0.3653846	D	51	1	L	94	1	F	137	1	D	180	0.3498904	A
9	0.1378855	K	52	1	G	95	0.0236518	S	138	1	E	181	1	R
10	1	R	53	1	N	96	0.4116855	E	139	0.2976101	I	182	0.7430464	T
11	0.2816477	E	54	0.2555179	V	97	0.3653846	R	140	1	R	183	1	G
12	0.2527799	G	55	1	K	98	1	E	141	0.4357733	K	184	1	G
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	0.1556140	K	143	1	R	186	1	L
15	0.0125324	D	58	0.7042527	I	101	1	E	144	0.3653846	A	187	1	D
16	0.1312546	A	59	1	G	102	1	H	145	0.3653846	A	188	0.2139290	I
17	1	L	60	1	K	103	1	R	146	1	G	189	0.1371408	R
18	1	E	61	1	?	104	1	R	147	1	P	190	1	G
19	0.3523813	Q	62	1	D	105	0.5317401	R	148	0.0090541	S	191	0.3523813	G
20	1	W	63	0.6218285	K	106	0.5317401	K	149	0.3194729	P	192	0.1312546	G
21	1	V	64	1	D	107	0.3653846	D	150	1	G	193	1	G
22	0.3653846	D	65	1	G	108	1	L	151	1	G	194	1	F
23	1	A	66	0.2976101	E	109	1	E	152	1	V	195	1	P
24	1	R	67	1	E	110	1	N	153	1	N	196	1	*
25	0.1878768	R	68	1	A	111	1	K	154	1	P	197	1	D
26	0.2644260	K	69	1	P	112	0.3653846	Q	155	0.3653846	S	198	0.3653846	?
27	0.1371408	Q	70	1	P	113	0.1312546	?	156	1	E	199	1	L
28	0.5317401	E	71	0.6218285	T	114	1	Q	157	0.3653846	G	200	1	F
29	0.2197463	D	72	1	K	115	1	L	158	0.1312546	S	201	1	P
30	1	L	73	0.3653846	?	116	0.1371408	N	159	0.0388780	S	202	0.2202327	A
31	1	E	74	1	S	117	0.6855633	A	160	1	R	203	0.5317401	D
32	0.7643337	R	75	1	R	118	1	G	161	1	G	204	1	P
33	0.8042677	D	76	0.2197463	T	119	1	G	162	1	A	205	1	P
34	0.2976101	L	77	0.3653846	E	120	1	K	163	1	P	206	0.3653846	F
35	1	R	78	0.4593211	R	121	0.1878768	H	164	1	G	207	1	S
36	0.2976101	K	79	1	M	122	1	L	165	1	G	208	1	P
37	0.2811874	T	80	1	E	123	1	S	166	1	G	209	1	Q
38	0.6501914	R	81	0.0170624	I	124	0.5415598	K	167	0.3653846	?	210	1	S
39	1	K	82	1	D	125	0.5317401	E	168	0.3653846	V	211	1	C
40	1	K	83	0.1556140	S	126	1	E	169	1	P	212	1	R
41	0.0885464	V	84	0.5317401	G	127	1	K	170	0.3653846	?	213	0.7042527	T
42	0.2976101	K	85	1	S	128	0.5317401	E	171	1	L	214	1	Q
43	0.1312546	?	86	0.0835304	R	129	1	E	172	1	L	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-A*02

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	1	I	87	1	K	130	1	L	173	0.3215377	G
2	1	S	45	1	E	88	0.4957057	K	131	0.1039095	E	174	0.5845124	I
3	1	Q	46	0.2784021	D	89	0.2884615	I	132	0.7733926	K	175	1	P
4	1	A	47	0.0506254	E	90	0.1990620	F	133	1	L	176	1	E
5	1	G	48	0.3180132	N	91	1	G	134	0.0712996	T	177	1	S
6	0.2884615	P	49	0.0202832	P	92	0.2884615	G	135	0.0594033	E	178	1	P
7	0.0860541	R	50	1	W	93	0.4085282	E	136	0.3180132	D	179	1	Y
8	0.2884615	D	51	1	L	94	1	F	137	1	D	180	0.1473538	H
9	0.2884615	?	52	1	G	95	0.3516560	S	138	1	E	181	1	R
10	1	R	53	1	N	96	0.3516560	E	139	0.1110419	R	182	0.4957057	S
11	0.4413282	E	54	0.6701407	I	97	0.5550344	E	140	1	R	183	1	G
12	0.2884615	E	55	1	K	98	1	E	141	0.5791425	E	184	0.6246429	D
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	0.1428466	K	143	1	R	186	1	L
15	0.0812173	N	58	1	L	101	0.4957057	E	144	0.1160820	I	187	1	D
16	0.1916489	I	59	1	G	102	1	H	145	0.2884615	A	188	0.3642637	V
17	1	L	60	1	K	103	1	R	146	1	G	189	0.4957057	T
18	1	E	61	1	K	104	1	R	147	1	P	190	1	G
19	0.2884615	E	62	0.4957057	E	105	0.4957057	K	148	0.4085282	Q	191	0.2884615	V
20	1	W	63	0.3215377	K	106	0.4957057	M	149	0.0318792	P	192	0.0812173	G
21	1	I	64	1	D	107	0.5550344	A	150	1	G	193	1	G
22	0.2884615	E	65	1	G	108	1	L	151	0.4957057	D	194	1	F
23	1	A	66	0.5550344	E	109	1	E	152	1	V	195	1	P
24	1	R	67	1	E	110	1	N	153	1	N	196	0.0222949	?
25	0.0812173	N	68	1	A	111	1	K	154	1	P	197	1	D
26	0.0326142	R	69	1	P	112	0.4957057	S	155	0.0812173	L	198	0.2884615	M
27	0.4085282	A	70	1	P	113	0.4085282	N	156	0.4957057	G	199	1	L
28	0.4957057	D	71	0.3180132	A	114	1	Q	157	0.2884615	G	200	1	F
29	0.3384023	E	72	1	K	115	1	L	158	0.1990620	E	201	1	P
30	1	L	73	0.6024698	K	116	0.4629646	A	159	0.1782089	Q	202	0.4957057	V
31	1	E	74	0.3516560	A	117	0.0810818	A	160	1	R	203	0.2884615	?
32	0.3384023	R	75	1	R	118	1	G	161	1	G	204	1	P
33	0.4106806	E	76	0.0059600	A	119	1	G	162	1	A	205	1	P
34	0.2884615	S	77	0.2884615	D	120	1	K	163	1	P	206	1	F
35	1	R	78	0.7200760	Q	121	0.0484450	N	164	1	G	207	1	S
36	1	R	79	1	M	122	1	L	165	1	G	208	1	P
37	0.4332315	A	80	0.2884615	E	123	1	S	166	1	G	209	1	Q
38	0.6246429	K	81	0.8006742	I	124	0.5165426	R	167	0.2884615	F	210	1	S
39	1	R	82	1	D	125	1	G	168	1	V	211	1	C
40	1	K	83	0.1428466	S	126	1	E	169	1	P	212	1	R
41	0.3180132	L	84	0.2884615	?	127	0.6871438	K	170	0.0249370	S	213	1	P
42	0.5550344	K	85	1	S	128	0.4957057	G	171	1	M	214	1	Q
43	0.0812173	?	86	0.6701407	R	129	1	E	172	0.5550344	Q	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-A*03

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	0.1685959	I	87	1	K	130	1	L	173	0.6406166	G
2	1	S	45	1	E	88	0.5102688	K	131	0.1685959	G	174	0.4519801	V
3	0.5678842	Q	46	0.1074378	D	89	0.1403321	P	132	0.7884705	R	175	1	P
4	0.4134615	P	47	0.0100459	D	90	0.2654032	P	133	1	L	176	1	E
5	1	G	48	0.2654032	H	91	0.4439532	G	134	0.4134615	I	177	1	S
6	0.4134615	P	49	0.1685959	L	92	1	R	135	0.4134615	I	178	1	P
7	0.3695566	R	50	1	W	93	0.4134615	V	136	1	D	179	1	Y
8	0.4134615	E	51	1	L	94	1	F	137	1	D	180	0.0880158	H
9	0.2649541	D	52	1	G	95	1	S	138	1	E	181	1	R
10	1	R	53	1	N	96	0.0793078	D	139	0.3560445	K	182	0.3117868	T
11	0.4134615	?	54	0.6968361	V	97	0.0629408	Q	140	1	R	183	1	G
12	0.4134615	S	55	1	K	98	1	E	141	0.7970908	E	184	1	G
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	0.0104672	K	143	1	R	186	1	L
15	0.0372973	E	58	0.6968361	I	101	0.5102688	E	144	0.1571481	I	187	1	D
16	0.5102688	L	59	1	G	102	1	H	145	1	T	188	0.0591009	V
17	1	L	60	1	K	103	1	R	146	1	G	189	0.6406166	R
18	1	E	61	1	?	104	1	R	147	1	P	190	1	G
19	0.4134615	E	62	0.5102688	E	105	0.5102688	K	148	0.4439532	Q	191	0.5678842	E
20	1	W	63	0.1403321	R	106	0.1685959	M	149	0.0150131	P	192	1	L
21	1	I	64	1	D	107	1	D	150	1	G	193	1	G
22	0.1685959	?	65	1	G	108	1	L	151	1	D	194	1	F
23	0.7742199	G	66	0.2654032	E	109	1	E	152	1	V	195	1	P
24	1	R	67	1	E	110	1	N	153	1	N	196	0.5678842	?
25	0.6894419	R	68	1	A	111	1	K	154	1	P	197	1	D
26	0.1568054	R	69	1	P	112	0.6661167	R	155	0.1685959	L	198	0.3874827	L
27	0.1224822	A	70	1	P	113	0.3759139	R	156	0.5102688	G	199	1	L
28	0.1685959	D	71	0.4010277	A	114	1	Q	157	0.4134615	G	200	1	F
29	0.0681592	D	72	1	K	115	1	L	158	0.4134615	V	201	1	P
30	1	L	73	0.8152607	K	116	0.4134615	E	159	0.4134615	?	202	0.5102688	V
31	1	E	74	0.0268391	P	117	0.0755037	A	160	1	R	203	0.5102688	D
32	0.7682783	K	75	1	R	118	1	G	161	1	G	204	1	P
33	0.0410112	E	76	0.3040512	A	119	1	G	162	1	A	205	1	P
34	0.4134615	S	77	1	E	120	1	K	163	1	P	206	1	L
35	1	R	78	0.4134615	H	121	0.0400507	S	164	1	G	207	1	S
36	0.0677689	R	79	1	M	122	1	L	165	1	G	208	1	P
37	0.1685959	?	80	0.4134615	E	123	1	S	166	1	G	209	1	Q
38	0.6466063	K	81	0.2486119	V	124	1	R	167	1	?	210	1	S
39	1	K	82	1	D	125	0.5102688	G	168	1	A	211	1	C
40	1	K	83	0.4134615	A	126	1	E	169	1	P	212	1	R
41	1	L	84	0.4134615	?	127	0.2702724	K	170	0.0233878	S	213	0.6968361	T
42	1	R	85	1	S	128	0.5102688	G	171	0.7327782	L	214	1	Q
43	1	R	86	0.6894419	R	129	1	E	172	1	L	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-A*11

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	1	I	87	1	K	130	1	L	173	1	S
2	1	S	45	1	E	88	1	K	131	0.4213609	K	174	0.7311734	V
3	0.3973992	Q	46	0.7894276	E	89	1	P	132	0.4558077	K	175	1	P
4	0.3484079	S	47	0.5822699	E	90	0.0610640	F	133	1	L	176	1	E
5	1	G	48	0.3973992	H	91	0.2929908	G	134	0.3973992	A	177	1	S
6	0.0710163	S	49	0.2852875	T	92	0.1538461	G	135	0.1538461	D	178	1	P
7	0.3600310	R	50	1	W	93	0.2852875	D	136	1	D	179	1	F
8	0.3973992	R	51	1	L	94	1	F	137	1	D	180	0.4558077	S
9	0.2412458	D	52	1	G	95	1	S	138	1	E	181	1	R
10	1	R	53	1	N	96	0.5869120	D	139	0.1538461	?	182	1	S
11	0.6256811	E	54	1	I	97	0.2053136	K	140	1	R	183	1	G
12	0.2852875	N	55	1	K	98	1	E	141	0.7323793	E	184	0.1682916	D
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	0.1197557	Q	143	1	R	186	1	L
15	0.2108062	E	58	0.5918297	I	101	1	D	144	0.1538461	L	187	1	D
16	0.2852875	A	59	1	G	102	1	H	145	0.1538461	A	188	0.0942010	I
17	1	L	60	1	K	103	1	R	146	1	G	189	0.1110147	R
18	1	E	61	1	K	104	1	R	147	1	P	190	1	G
19	1	?	62	1	E	105	1	R	148	0.0028037	S	191	0.0250994	G
20	1	W	63	0.4928607	K	106	1	K	149	0.1813143	A	192	0.0224047	G
21	1	V	64	1	D	107	1	D	150	1	G	193	1	G
22	0.1538461	D	65	1	G	108	1	L	151	0.2852875	D	194	1	F
23	1	G	66	1	G	109	1	E	152	1	V	195	1	P
24	1	R	67	1	E	110	1	N	153	1	N	196	0.3973992	?
25	1	K	68	1	A	111	1	K	154	1	P	197	1	D
26	0.1538461	L	69	1	P	112	0.2356754	R	155	0.1538461	S	198	0.1538461	M
27	0.2288343	L	70	1	P	113	0.2852875	?	156	0.2852875	G	199	1	L
28	1	E	71	0.1110147	T	114	1	Q	157	0.1538461	G	200	1	F
29	0.6860398	D	72	1	K	115	1	L	158	0.0610640	E	201	1	P
30	1	L	73	0.1538461	?	116	0.4928607	N	159	0.1110147	R	202	1	S
31	1	E	74	0.2852875	S	117	0.4162792	A	160	1	R	203	1	E
32	1	R	75	1	R	118	1	G	161	1	G	204	1	P
33	0.1834294	D	76	0.0250994	V	119	1	G	162	1	A	205	1	P
34	1	L	77	0.1538461	D	120	1	K	163	1	P	206	1	L
35	1	R	78	0.1036335	R	121	0.1538461	?	164	1	G	207	1	S
36	1	R	79	1	M	122	1	L	165	1	G	208	1	P
37	0.2852875	M	80	1	D	123	1	S	166	1	G	209	1	Q
38	0.5740030	K	81	0.1094447	V	124	0.4156666	K	167	1	F	210	1	S
39	0.3973992	K	82	1	D	125	1	E	168	1	V	211	1	C
40	1	K	83	1	S	126	1	E	169	1	P	212	1	R
41	0.1191502	I	84	1	G	127	1	E	170	0.1538461	?	213	1	T
42	1	R	85	1	P	128	1	E	171	0.3484079	L	214	1	Q
43	0.2074425	R	86	0.5869120	R	129	1	E	172	1	L	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-A*23

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	1	I	87	1	K	130	1	L	173	0.0480769	D
2	1	S	45	1	E	88	1	K	131	0.0480769	Q	174	0.5803723	I
3	1	Q	46	0.1570178	D	89	1	I	132	0.5740030	K	175	1	P
4	1	A	47	0.0942867	G	90	0.4652481	S	133	1	L	176	1	E
5	0.0480769	E	48	0.0054913	H	91	0.1813238	K	134	1	A	177	1	S
6	0.5816674	S	49	1	S	92	1	R	135	0.0480769	D	178	1	P
7	1	K	50	1	W	93	0.0942867	D	136	1	D	179	1	F
8	1	R	51	1	L	94	1	F	137	1	D	180	0.1786445	T
9	0.1503926	D	52	1	G	95	1	S	138	1	E	181	1	R
10	1	R	53	1	N	96	1	E	139	0.4348645	K	182	0.0480769	?
11	0.0480769	S	54	0.2992142	V	97	1	E	140	1	R	183	1	G
12	1	G	55	1	K	98	1	E	141	1	E	184	0.0480769	G
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	1	E	143	1	R	186	1	L
15	1	D	58	1	I	101	1	E	144	0.1867331	T	187	1	D
16	0.0622106	I	59	1	G	102	1	H	145	1	A	188	0.4652481	L
17	1	L	60	1	K	103	1	R	146	1	G	189	0.0480769	V
18	1	E	61	1	?	104	1	R	147	1	P	190	1	G
19	1	?	62	1	E	105	1	K	148	0.2992142	Q	191	0.0705607	D
20	1	W	63	1	R	106	1	M	149	0.0480769	G	192	0.0206614	R
21	1	V	64	1	D	107	1	A	150	1	G	193	1	G
22	0.0565221	N	65	1	G	108	1	L	151	1	G	194	1	F
23	0.4943107	A	66	0.0942867	G	109	1	E	152	1	V	195	1	P
24	1	R	67	1	E	110	1	N	153	1	N	196	0.2222576	W
25	0.2615375	R	68	1	A	111	1	K	154	1	P	197	1	D
26	1	L	69	1	P	112	0.0054581	K	155	1	H	198	0.1331041	L
27	0.1386844	M	70	1	P	113	0.5838954	K	156	1	G	199	1	L
28	1	E	71	1	T	114	1	Q	157	1	G	200	1	F
29	1	D	72	1	K	115	1	L	158	0.0480769	L	201	1	P
30	1	L	73	0.5838954	K	116	0.0583943	T	159	0.2615375	P	202	0.0480769	?
31	1	E	74	1	P	117	0.0622106	A	160	1	R	203	1	?
32	0.4943107	R	75	1	R	118	1	G	161	1	G	204	1	P
33	0.2249839	E	76	1	A	119	1	G	162	1	A	205	1	P
34	1	L	77	1	D	120	1	K	163	1	P	206	1	L
35	1	R	78	1	H	121	0.0707567	N	164	1	G	207	1	S
36	1	R	79	1	M	122	1	L	165	1	G	208	1	P
37	0.0463534	A	80	1	E	123	1	S	166	1	G	209	1	Q
38	1	R	81	0.3335738	V	124	1	K	167	1	F	210	1	S
39	1	R	82	1	D	125	1	G	168	1	A	211	1	C
40	1	K	83	1	P	126	1	E	169	1	P	212	1	R
41	0.4031153	V	84	1	G	127	0.3353372	E	170	0.3258347	N	213	1	T
42	0.1386844	K	85	1	S	128	1	G	171	0.3699550	L	214	1	Q
43	0.4348645	R	86	1	G	129	1	E	172	0.1386844	Q	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-A*24

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	1	I	87	1	K	130	1	L	173	0.1958941	G
2	1	S	45	1	E	88	0.3799477	K	131	0.2115384	Q	174	0.5554678	I
3	0.5136844	Q	46	1	D	89	0.5762795	P	132	0.2115384	?	175	1	P
4	0.2115384	P	47	0.4701961	D	90	0.5136844	P	133	1	L	176	1	E
5	0.2115384	G	48	0.5815760	N	91	0.1617030	G	134	0.5762795	T	177	1	S
6	0.3508711	S	49	0.5136844	S	92	1	R	135	0.2115384	I	178	1	P
7	0.1134158	K	50	1	W	93	0.4530397	G	136	0.2848754	E	179	1	Y
8	0.5136844	R	51	1	L	94	1	F	137	1	D	180	0.0396602	S
9	0.1124741	H	52	1	G	95	0.3381294	S	138	1	E	181	1	R
10	1	R	53	1	N	96	0.3381294	E	139	0.1073622	E	182	0.6845644	T
11	0.2115384	?	54	0.6373659	I	97	0.1124741	E	140	1	R	183	1	G
12	0.1124741	D	55	1	K	98	1	E	141	1	E	184	0.2848754	D
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	0.1958941	E	143	1	R	186	1	L
15	0.7306914	E	58	0.6373659	L	101	1	E	144	0.2115384	A	187	1	D
16	0.3612513	T	59	1	G	102	1	H	145	1	T	188	0.4530397	L
17	1	L	60	1	K	103	1	R	146	1	G	189	0.2115384	G
18	1	E	61	1	?	104	1	R	147	1	P	190	1	G
19	0.2115384	?	62	0.3799477	E	105	1	K	148	0.3406653	Q	191	0.2263217	S
20	1	W	63	0.5762795	R	106	1	M	149	0.2115384	G	192	0.3107694	Q
21	0.3799477	I	64	1	D	107	0.0431292	T	150	1	G	193	1	G
22	0.3339861	S	65	1	G	108	1	L	151	1	D	194	1	F
23	0.4662794	A	66	0.3799477	G	109	1	E	152	1	V	195	1	P
24	1	R	67	1	E	110	1	N	153	1	N	196	0.1981169	*
25	1	R	68	1	A	111	1	K	154	1	P	197	1	D
26	0.5127636	K	69	1	P	112	0.0336787	R	155	1	S	198	0.2115384	K
27	0.1617030	A	70	1	P	113	0.2115384	V	156	1	G	199	1	L
28	1	D	71	0.5762795	T	114	1	Q	157	1	A	200	1	F
29	1	E	72	1	K	115	1	L	158	0.5136844	E	201	1	P
30	1	L	73	0.0217332	R	116	0.1073622	Q	159	0.1958941	T	202	0.2115384	?
31	1	E	74	0.3799477	S	117	0.4701961	A	160	1	R	203	1	E
32	0.0288587	K	75	1	R	118	1	G	161	1	G	204	1	P
33	0.0685551	E	76	0.2115384	?	119	1	G	162	1	A	205	1	P
34	1	V	77	1	E	120	1	K	163	1	P	206	1	L
35	1	R	78	0.2115384	P	121	0.1958941	K	164	1	G	207	1	S
36	0.5136844	R	79	1	M	122	1	L	165	1	G	208	1	P
37	0.2044575	T	80	1	D	123	1	S	166	1	G	209	1	Q
38	1	K	81	0.7797022	I	124	0.8107111	R	167	1	?	210	1	S
39	0.5136844	R	82	1	D	125	0.3799477	G	168	1	A	211	1	C
40	1	K	83	0.5815760	S	126	1	E	169	1	P	212	1	R
41	0.2848754	L	84	1	A	127	0.1981169	E	170	0.2704285	N	213	0.6373659	P
42	0.5136844	R	85	1	S	128	1	G	171	0.0920310	M	214	1	Q
43	0.0653258	K	86	1	G	129	1	E	172	0.5136844	L	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-A*25

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	1	L	87	1	K	130	1	L	173	1	N
2	1	S	45	1	E	88	1	R	131	0.5093353	R	174	1	I
3	1	Q	46	1	E	89	0.0192307	L	132	1	K	175	1	P
4	0.0571321	A	47	0.4998132	D	90	0.0571321	P	133	1	L	176	1	E
5	1	G	48	1	H	91	1	A	134	1	A	177	1	S
6	1	A	49	1	S	92	1	R	135	1	K	178	1	P
7	1	K	50	1	W	93	1	R	136	1	D	179	1	F
8	1	R	51	1	L	94	1	F	137	1	D	180	0.4247572	A
9	0.0758028	G	52	1	G	95	1	T	138	1	E	181	1	R
10	1	R	53	1	N	96	1	E	139	0.0192307	G	182	1	S
11	1	R	54	1	I	97	1	E	140	1	R	183	1	G
12	1	D	55	1	K	98	1	E	141	0.3175877	K	184	1	E
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	1	K	143	1	R	186	1	L
15	1	E	58	1	L	101	1	D	144	1	?	187	1	D
16	0.1486183	T	59	1	G	102	1	H	145	1	A	188	0.0123226	L
17	1	L	60	1	K	103	1	R	146	1	G	189	1	T
18	1	E	61	1	K	104	1	R	147	1	P	190	1	G
19	1	?	62	1	E	105	1	R	148	0.0606796	P	191	0.1685959	N
20	1	W	63	1	R	106	1	K	149	0.4958924	V	192	0.5491038	Q
21	1	V	64	1	D	107	1	D	150	1	G	193	1	G
22	1	S	65	1	G	108	1	L	151	1	D	194	1	F
23	0.2354368	A	66	1	D	109	1	E	152	1	V	195	1	P
24	1	R	67	1	G	110	1	N	153	1	N	196	1	?
25	1	R	68	1	A	111	1	K	154	1	P	197	1	D
26	1	K	69	1	P	112	0.5093353	R	155	1	S	198	1	K
27	1	M	70	1	P	113	1	R	156	1	E	199	1	L
28	1	D	71	1	T	114	1	Q	157	1	G	200	1	F
29	0.2354368	D	72	1	K	115	1	L	158	1	E	201	1	P
30	1	L	73	1	K	116	0.1663554	T	159	1	P	202	1	S
31	1	E	74	1	A	117	0.4998132	A	160	1	R	203	1	E
32	0.2354368	K	75	1	R	118	1	G	161	1	G	204	1	P
33	1	N	76	1	T	119	1	G	162	1	A	205	1	P
34	1	S	77	1	E	120	1	K	163	1	P	206	1	L
35	1	R	78	1	Q	121	1	D	164	1	G	207	1	S
36	1	R	79	1	M	122	1	L	165	1	G	208	1	P
37	0.5102688	V	80	1	D	123	1	S	166	1	G	209	1	Q
38	1	R	81	1	V	124	1	K	167	1	F	210	1	S
39	1	R	82	1	D	125	1	G	168	1	A	211	1	C
40	1	K	83	1	P	126	1	E	169	1	P	212	1	R
41	1	V	84	1	G	127	1	K	170	1	?	213	1	T
42	1	R	85	1	P	128	0.0382748	G	171	1	M	214	1	Q
43	1	R	86	0.1125840	R	129	1	E	172	1	L	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-A*26

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	1	I	87	1	K	130	1	L	173	1	G
2	1	S	45	1	E	88	1	R	131	0.1486183	G	174	0.0425796	I
3	1	Q	46	0.7149635	D	89	1	I	132	0.3477077	R	175	1	P
4	0.0205219	S	47	0.0081944	E	90	0.6759134	L	133	1	L	176	1	E
5	1	G	48	0.1486183	?	91	0.4389210	G	134	1	I	177	1	S
6	0.1486183	G	49	1	P	92	1	R	135	0.3353372	K	178	1	P
7	0.4415181	R	50	1	W	93	0.0769230	V	136	1	D	179	1	Y
8	0.0769230	E	51	1	L	94	1	F	137	1	D	180	0.3531794	S
9	0.4389210	R	52	1	G	95	1	S	138	1	E	181	1	R
10	1	R	53	1	N	96	1	E	139	0.2153934	I	182	1	T
11	0.0150683	R	54	1	I	97	0.5918608	K	140	1	R	183	1	G
12	1	G	55	1	K	98	1	E	141	0.6245077	K	184	1	D
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	0.2775404	E	143	1	R	186	1	L
15	0.5914911	E	58	1	L	101	1	E	144	0.5280959	I	187	1	D
16	0.4628764	I	59	1	G	102	1	H	145	1	T	188	0.4433326	V
17	1	L	60	1	K	103	1	R	146	1	G	189	1	G
18	1	E	61	0.0769230	K	104	1	R	147	1	P	190	1	G
19	0.1486183	K	62	1	D	105	1	R	148	0.1486183	R	191	0.2153934	E
20	1	W	63	1	R	106	0.1486183	M	149	0.4415181	T	192	0.7144291	Q
21	1	V	64	1	D	107	1	T	150	1	G	193	1	G
22	0.2783102	S	65	1	G	108	1	L	151	0.1486183	D	194	1	F
23	0.5914911	A	66	1	E	109	1	E	152	1	V	195	1	P
24	1	R	67	0.0769230	G	110	1	N	153	1	N	196	1	*
25	0.3890473	R	68	1	A	111	1	K	154	1	P	197	1	D
26	1	R	69	1	P	112	0.7073016	K	155	1	S	198	1	L
27	0.2775404	Q	70	1	P	113	0.1945643	K	156	1	G	199	1	L
28	1	E	71	0.2775404	T	114	1	Q	157	1	A	200	1	F
29	0.5914911	E	72	1	K	115	1	L	158	0.1486183	D	201	1	P
30	1	L	73	0.3707295	K	116	0.2775404	R	159	0.0313534	S	202	0.2783102	S
31	1	E	74	1	S	117	0.2886244	A	160	1	R	203	1	E
32	1	R	75	1	R	118	1	G	161	1	G	204	1	P
33	1	D	76	0.5914911	T	119	1	G	162	1	A	205	1	P
34	1	V	77	1	E	120	1	K	163	1	P	206	1	L
35	1	R	78	0.4851955	R	121	0.4805011	S	164	1	G	207	1	S
36	1	K	79	1	M	122	1	L	165	1	G	208	1	P
37	0.2153934	E	80	1	D	123	1	S	166	1	G	209	1	Q
38	1	R	81	0.6762549	I	124	0.7154542	K	167	1	?	210	1	S
39	0.2153934	K	82	1	D	125	1	E	168	1	A	211	1	C
40	1	K	83	0.0769230	A	126	1	E	169	1	P	212	1	R
41	0.5987428	I	84	1	G	127	1	E	170	1	S	213	0.4389210	P
42	1	R	85	1	S	128	1	E	171	1	L	214	1	Q
43	1	K	86	1	R	129	1	E	172	0.2153934	Q	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-A*29

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	1	I	87	1	K	130	1	L	173	1	S
2	1	S	45	1	E	88	1	R	131	0.1552775	R	174	1	I
3	1	Q	46	1	E	89	0.0942867	T	132	0.5740030	K	175	1	P
4	0.2222576	T	47	0.0751582	E	90	0.2615375	H	133	1	L	176	1	E
5	1	E	48	1	H	91	1	G	134	1	A	177	1	S
6	0.0942867	K	49	1	S	92	1	G	135	0.0942867	R	178	1	P
7	0.0480769	E	50	1	W	93	1	E	136	1	D	179	1	F
8	1	R	51	1	L	94	1	F	137	1	D	180	0.1786445	T
9	0.1813238	G	52	1	G	95	1	T	138	1	E	181	1	R
10	1	R	53	1	N	96	1	D	139	0.2615375	E	182	1	H
11	0.1813238	D	54	1	V	97	0.1386844	E	140	1	R	183	1	G
12	0.1386844	D	55	1	K	98	1	E	141	1	E	184	1	E
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	0.2222576	R	143	1	R	186	1	L
15	0.0715127	D	58	0.0354776	L	101	1	E	144	0.0206614	V	187	1	D
16	0.1633149	V	59	1	G	102	1	H	145	1	A	188	0.0476556	V
17	1	L	60	1	K	103	1	R	146	1	G	189	1	T
18	1	E	61	1	?	104	1	R	147	1	P	190	1	G
19	1	R	62	1	E	105	0.0942867	K	148	0.3278593	P	191	0.3052863	D
20	1	W	63	0.0002163	R	106	1	M	149	0.4031153	A	192	0.1386844	H
21	1	I	64	1	D	107	1	T	150	1	G	193	1	G
22	0.1597670	S	65	1	G	108	1	L	151	1	D	194	1	F
23	0.0480769	E	66	0.0480769	D	109	1	E	152	1	V	195	1	P
24	1	R	67	1	E	110	1	N	153	1	N	196	1	*
25	1	K	68	1	A	111	1	K	154	1	P	197	1	D
26	1	L	69	1	P	112	0.0942867	S	155	1	H	198	0.5220958	L
27	0.2992142	V	70	1	P	113	0.2992142	N	156	0.0942867	G	199	1	L
28	1	E	71	0.0480769	P	114	1	Q	157	1	G	200	1	F
29	0.4943107	E	72	1	K	115	1	L	158	0.5872457	G	201	1	P
30	1	L	73	1	K	116	0.0480769	I	159	0.2992142	A	202	1	P
31	1	E	74	1	P	117	1	G	160	1	R	203	1	E
32	1	K	75	1	R	118	1	G	161	1	G	204	1	P
33	0.0480769	N	76	0.0480769	S	119	1	G	162	1	A	205	1	P
34	0.0942867	V	77	1	E	120	1	K	163	1	P	206	1	F
35	1	R	78	0.3353372	R	121	0.3702547	N	164	1	G	207	1	S
36	1	R	79	1	M	122	1	L	165	1	G	208	1	P
37	0.0480769	S	80	1	D	123	1	S	166	1	G	209	1	Q
38	0.2222576	K	81	0.3335738	I	124	1	K	167	1	F	210	1	S
39	1	R	82	1	D	125	1	E	168	1	A	211	1	C
40	1	K	83	0.1386844	P	126	1	E	169	1	P	212	1	R
41	0.2222576	L	84	0.0480769	A	127	0.3353372	E	170	0.3258347	N	213	1	T
42	1	R	85	0.0480769	P	128	1	E	171	1	L	214	1	Q
43	1	K	86	1	G	129	1	E	172	1	L	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-A*30

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	1	L	87	1	K	130	1	L	173	1	G
2	1	S	45	1	E	88	1	R	131	0.0247888	E	174	1	I
3	1	R	46	0.0478345	E	89	1	P	132	0.0639467	K	175	1	P
4	0.5922594	S	47	0.0022285	D	90	0.2012696	A	133	1	L	176	1	E
5	1	E	48	1	?	91	0.5540427	G	134	0.2874071	A	177	1	S
6	0.2087109	S	49	0.0021108	P	92	1	G	135	0.5946438	E	178	1	P
7	1	K	50	1	W	93	0.1593449	E	136	1	D	179	1	F
8	1	R	51	1	L	94	1	F	137	1	D	180	0.3382618	T
9	0.0080500	N	52	1	G	95	0.4976573	S	138	1	E	181	1	R
10	1	R	53	1	N	96	0.0150697	E	139	0.0910457	K	182	0.0704989	T
11	0.2421791	E	54	0.5540427	I	97	0.0372734	Q	140	1	R	183	1	G
12	0.1057692	E	55	1	K	98	1	E	141	0.0949990	K	184	1	D
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	0.0080143	K	143	1	R	186	1	L
15	0.1391463	E	58	0.5540427	L	101	1	D	144	0.0067555	I	187	1	D
16	0.5253019	V	59	1	G	102	1	H	145	1	A	188	0.3549913	L
17	1	L	60	1	K	103	1	R	146	1	G	189	0.2012696	T
18	1	E	61	1	K	104	1	R	147	1	P	190	1	G
19	0.1057692	R	62	1	E	105	1	R	148	0.1593449	Q	191	0.2087109	S
20	1	W	63	1	K	106	0.0102688	M	149	0.4816096	T	192	0.1757138	Q
21	1	V	64	1	D	107	1	A	150	1	G	193	1	G
22	0.1057692	E	65	1	G	108	1	L	151	1	D	194	1	F
23	1	E	66	1	E	109	1	E	152	1	V	195	1	P
24	1	R	67	1	G	110	1	N	153	1	N	196	0.4348645	W
25	0.1209003	R	68	1	A	111	1	K	154	1	P	197	1	D
26	1	L	69	1	P	112	0.2012696	S	155	0.0102688	L	198	0.2041765	I
27	0.0545798	Q	70	1	P	113	0.0301715	R	156	1	E	199	1	L
28	1	D	71	0.3650163	T	114	1	Q	157	0.1057692	G	200	1	F
29	1	E	72	1	K	115	1	L	158	0.1057692	V	201	1	P
30	1	L	73	1	?	116	0.1057692	E	159	0.1610956	S	202	0.5211825	S
31	1	E	74	0.0545798	P	117	0.6905301	S	160	1	R	203	1	?
32	0.3513249	K	75	1	R	118	1	G	161	1	G	204	1	P
33	0.0044825	E	76	0.0545798	A	119	1	G	162	1	A	205	1	P
34	0.1057692	S	77	1	D	120	1	K	163	1	P	206	1	F
35	1	R	78	1	R	121	0.0204244	S	164	1	G	207	1	S
36	0.2874071	K	79	1	M	122	1	L	165	1	G	208	1	P
37	0.2000354	I	80	1	E	123	1	S	166	1	G	209	1	Q
38	0.0856234	K	81	0.4541560	V	124	0.0079180	K	167	1	F	210	1	S
39	1	K	82	1	D	125	1	E	168	1	V	211	1	C
40	1	K	83	1	H	126	1	E	169	1	P	212	1	R
41	0.2850992	V	84	1	A	127	0.0372734	K	170	0.0623584	S	213	0.5540427	P
42	1	K	85	1	P	128	1	G	171	1	M	214	1	Q
43	0.0157073	R	86	1	G	129	1	E	172	0.2874071	Q	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-A*32

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	1	I	87	1	K	130	1	L	173	1	G
2	1	S	45	1	E	88	1	K	131	0.0880158	K	174	0.0969568	I
3	1	Q	46	1	D	89	1	P	132	0.6497601	K	175	1	P
4	0.0715127	T	47	0.5059451	D	90	0.4448799	L	133	1	L	176	1	E
5	1	E	48	0.1839058	?	91	0.5181611	G	134	0.0961538	I	177	1	S
6	0.2017529	S	49	0.1839058	L	92	1	G	135	0.0059210	K	178	1	P
7	0.2764753	K	50	1	W	93	0.0151182	G	136	1	E	179	1	F
8	1	R	51	1	L	94	1	F	137	1	D	180	0.2460809	H
9	0.0929870	N	52	1	G	95	1	T	138	1	E	181	1	R
10	1	R	53	1	N	96	0.0112213	E	139	0.0238874	I	182	0.5181611	H
11	0.2038545	G	54	1	V	97	0.0281937	Q	140	1	R	183	1	G
12	0.0961538	S	55	1	K	98	1	E	141	0.0681931	K	184	0.4031153	D
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	0.0059210	K	143	1	R	186	1	L
15	0.0071060	D	58	1	I	101	1	E	144	0.0961538	?	187	1	D
16	0.5678352	T	59	1	G	102	1	H	145	1	A	188	0.1588073	V
17	1	L	60	1	K	103	1	R	146	1	G	189	1	R
18	1	E	61	1	K	104	1	R	147	1	P	190	1	G
19	0.1839058	K	62	1	D	105	1	R	148	0.4448799	P	191	0.3849964	S
20	1	W	63	1	K	106	1	M	149	0.0541657	P	192	0.4889784	Q
21	1	V	64	1	D	107	1	D	150	1	G	193	1	G
22	0.3188872	N	65	1	G	108	1	L	151	0.1839058	D	194	1	F
23	0.6096321	A	66	1	E	109	1	E	152	1	V	195	1	P
24	1	R	67	1	G	110	1	N	153	1	N	196	0.1704423	*
25	0.1704423	K	68	1	A	111	1	K	154	1	P	197	1	D
26	0.6347107	R	69	1	P	112	0.1839058	S	155	1	L	198	0.1317533	L
27	0.4031153	I	70	1	P	113	0.6096321	R	156	1	E	199	1	L
28	0.1839058	D	71	1	T	114	1	Q	157	1	G	200	1	F
29	0.3556579	E	72	1	K	115	1	L	158	0.1115678	G	201	1	P
30	1	L	73	0.6882280	K	116	0.0158715	S	159	0.0961538	?	202	0.3346765	A
31	1	E	74	0.3367948	P	117	0.0402567	A	160	1	R	203	1	D
32	0.6096321	R	75	1	R	118	1	G	161	1	G	204	1	P
33	1	E	76	0.3556579	T	119	1	G	162	1	A	205	1	P
34	1	V	77	1	D	120	1	K	163	1	P	206	1	F
35	1	R	78	0.5924762	Q	121	0.1778501	S	164	1	G	207	1	S
36	1	R	79	1	M	122	1	L	165	1	G	208	1	P
37	0.2639151	E	80	0.0961538	E	123	1	S	166	1	G	209	1	Q
38	1	R	81	0.0117851	V	124	1	R	167	1	F	210	1	S
39	1	R	82	1	D	125	1	E	168	1	V	211	1	C
40	1	K	83	0.0715127	S	126	1	E	169	1	P	212	1	R
41	1	I	84	0.0961538	?	127	0.0281937	K	170	0.1839058	G	213	1	P
42	1	R	85	1	P	128	1	E	171	0.5948455	M	214	1	Q
43	1	R	86	0.1016584	G	129	1	E	172	1	L	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-A*33

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	1	L	87	1	K	130	1	L	173	1	G
2	1	S	45	1	E	88	1	R	131	0.3595873	R	174	0.0836009	V
3	1	Q	46	1	D	89	1	P	132	1	K	175	1	P
4	0.0080411	S	47	0.2281434	E	90	0.3327476	L	133	1	L	176	1	E
5	1	G	48	1	H	91	0.1411561	R	134	1	T	177	1	S
6	0.0576923	P	49	1	L	92	1	G	135	0.0576923	I	178	1	P
7	0.6590867	R	50	1	W	93	1	A	136	0.2615375	E	179	1	F
8	0.0576923	E	51	1	L	94	1	F	137	1	D	180	0.0294849	T
9	0.0054870	K	52	1	G	95	1	S	138	1	E	181	1	R
10	1	R	53	1	N	96	1	E	139	0.1209003	K	182	1	S
11	0.0629038	G	54	1	I	97	0.1647849	E	140	1	R	183	1	G
12	1	N	55	1	K	98	1	E	141	0.5868715	K	184	0.0258580	D
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	0.2615375	R	143	1	R	186	1	L
15	0.4634067	D	58	1	L	101	1	D	144	0.0576923	?	187	1	D
16	0.0070763	I	59	1	G	102	1	H	145	1	T	188	0.4145631	V
17	1	L	60	1	K	103	1	R	146	1	G	189	1	R
18	1	E	61	1	?	104	1	R	147	1	P	190	1	G
19	1	E	62	1	D	105	1	R	148	0.3327476	P	191	0.0258580	G
20	1	W	63	1	K	106	1	K	149	0.1016584	P	192	0.6610093	Q
21	1	V	64	1	D	107	0.1125840	T	150	1	G	193	1	G
22	0.0576923	D	65	1	G	108	1	L	151	1	G	194	1	F
23	0.5607143	A	66	1	E	109	1	E	152	1	V	195	1	P
24	1	R	67	1	E	110	1	N	153	1	N	196	0.0081821	?
25	1	K	68	1	A	111	1	K	154	1	P	197	1	D
26	0.2067113	R	69	1	P	112	1	K	155	1	L	198	0.5896782	L
27	0.0576923	?	70	1	P	113	0.3487647	N	156	1	E	199	1	L
28	1	E	71	1	P	114	1	Q	157	1	A	200	1	F
29	1	D	72	1	K	115	1	L	158	0.1340499	G	201	1	P
30	1	L	73	0.3320885	R	116	0.3062928	Q	159	0.2144016	R	202	0.6959876	S
31	1	E	74	1	S	117	0.1484338	G	160	1	R	203	1	D
32	1	R	75	1	R	118	1	G	161	1	G	204	1	P
33	0.3438548	D	76	1	L	119	1	G	162	1	A	205	1	P
34	1	S	77	1	E	120	1	K	163	1	P	206	1	L
35	1	R	78	1	Q	121	0.6733891	N	164	1	G	207	1	S
36	0.1647849	R	79	1	M	122	1	L	165	1	G	208	1	P
37	0.0028005	?	80	0.0576923	D	123	1	S	166	1	G	209	1	Q
38	0.2615375	K	81	1	V	124	0.6781231	K	167	1	?	210	1	S
39	1	K	82	1	D	125	1	E	168	1	A	211	1	C
40	1	K	83	0.0258580	S	126	1	E	169	1	P	212	1	R
41	0.4634067	V	84	0.0576923	?	127	1	K	170	1	S	213	1	T
42	1	K	85	1	P	128	1	E	171	0.4272318	M	214	1	Q
43	1	R	86	0.3062928	G	129	1	E	172	1	L	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-A*68

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	1	L	87	1	K	130	1	L	173	1	G
2	1	S	45	1	E	88	0.0102688	R	131	0.1593449	E	174	0.4388279	I
3	1	R	46	0.7586127	D	89	0.2012696	T	132	0.6872525	R	175	1	P
4	0.5922594	S	47	0.2012696	S	90	0.2874071	P	133	1	L	176	1	E
5	1	E	48	1	N	91	0.1057692	A	134	0.0034089	T	177	1	S
6	0.0102688	K	49	1	L	92	0.1057692	G	135	0.2012696	R	178	1	P
7	0.1057692	E	50	1	W	93	0.1057692	A	136	0.4348645	D	179	0.1057692	F
8	0.2874071	R	51	1	L	94	1	F	137	1	D	180	0.0545798	V
9	0.1593449	R	52	1	G	95	0.1209003	S	138	1	E	181	1	R
10	1	R	53	1	N	96	1	D	139	0.2874071	I	182	0.5946438	T
11	0.3650163	D	54	1	I	97	0.3549913	K	140	1	R	183	1	G
12	0.2874071	D	55	1	K	98	1	E	141	1	K	184	1	D
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	0.3650163	E	143	1	R	186	1	L
15	0.0704989	D	58	0.1593449	I	101	0.2012696	E	144	1	I	187	1	D
16	0.0623304	I	59	1	G	102	1	H	145	1	A	188	0.0265007	I
17	1	L	60	1	K	103	1	R	146	1	G	189	1	R
18	1	E	61	1	K	104	1	R	147	1	P	190	1	G
19	1	E	62	1	D	105	1	R	148	0.2012696	R	191	0.1057692	V
20	1	W	63	0.3650163	K	106	1	M	149	0.2850992	A	192	0.7467277	Q
21	1	V	64	1	D	107	0.2012696	T	150	1	G	193	1	G
22	0.2012696	T	65	1	G	108	1	L	151	1	D	194	1	F
23	0.1057692	E	66	0.1057692	D	109	1	E	152	1	V	195	1	P
24	1	R	67	1	G	110	1	N	153	1	N	196	0.2874071	?
25	0.2012696	N	68	1	A	111	1	K	154	1	P	197	1	D
26	0.6580271	R	69	1	P	112	0.0525061	K	155	1	H	198	0.1057692	M
27	0.1593449	V	70	1	P	113	0.6254130	R	156	1	E	199	1	L
28	1	E	71	0.1057692	P	114	1	Q	157	1	G	200	1	F
29	1	D	72	1	K	115	1	L	158	0.0642269	G	201	1	P
30	1	L	73	0.7040745	K	116	0.1371999	A	159	0.0335603	S	202	0.0102688	V
31	1	E	74	0.0102688	S	117	0.5390112	A	160	1	R	203	0.1057692	?
32	1	K	75	1	R	118	1	G	161	1	G	204	1	P
33	0.1057692	N	76	0.3650163	A	119	1	G	162	1	A	205	1	P
34	1	L	77	1	D	120	1	K	163	1	P	206	1	F
35	1	R	78	1	Q	121	0.1209003	H	164	1	G	207	1	S
36	1	K	79	1	M	122	1	L	165	1	G	208	1	P
37	0.1057692	S	80	1	E	123	1	S	166	1	G	209	1	Q
38	0.4348645	R	81	0.0036510	I	124	1	R	167	0.1057692	F	210	1	S
39	1	K	82	1	D	125	0.2012696	E	168	1	V	211	1	C
40	1	K	83	0.0856234	S	126	1	E	169	1	P	212	1	R
41	0.2850992	V	84	1	G	127	0.5954202	E	170	0.2728318	N	213	1	P
42	0.2874071	K	85	1	P	128	0.2012696	G	171	0.2421791	M	214	1	Q
43	1	K	86	0.4976573	G	129	1	E	172	1	Q	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-A*69

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	1	I	87	1	K	130	1	L	173	1	N
2	1	S	45	1	E	88	1	R	131	0.5093353	R	174	1	I
3	1	Q	46	0.1685959	E	89	1	L	132	1	R	175	1	P
4	1	A	47	0.4998132	D	90	1	P	133	1	L	176	1	E
5	1	G	48	1	H	91	1	R	134	1	A	177	1	S
6	1	R	49	1	S	92	1	R	135	1	V	178	1	P
7	0.4678864	R	50	1	W	93	1	V	136	1	D	179	1	F
8	1	R	51	1	L	94	1	F	137	1	D	180	0.4247572	A
9	0.1306945	R	52	1	G	95	1	S	138	1	E	181	1	R
10	1	R	53	1	N	96	1	D	139	1	I	182	1	T
11	1	R	54	1	V	97	1	E	140	1	R	183	1	G
12	0.0382748	N	55	1	K	98	1	E	141	1	K	184	1	E
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	1	R	143	1	R	186	1	L
15	1	E	58	1	I	101	1	E	144	0.3015309	T	187	1	D
16	1	L	59	1	G	102	1	H	145	1	A	188	0.1176250	V
17	1	L	60	1	K	103	1	R	146	1	G	189	1	G
18	1	E	61	1	K	104	1	R	147	1	P	190	1	G
19	1	Q	62	1	E	105	1	R	148	0.2852875	S	191	0.1685959	N
20	1	W	63	1	R	106	1	M	149	0.1839058	P	192	0.5491038	Q
21	1	I	64	1	D	107	1	T	150	1	G	193	1	G
22	1	S	65	1	G	108	1	L	151	1	D	194	1	F
23	1	A	66	1	E	109	1	E	152	1	V	195	1	P
24	1	R	67	1	E	110	1	N	153	1	N	196	0.0571321	?
25	0.0382748	N	68	1	A	111	1	K	154	1	P	197	1	D
26	0.2688573	R	69	1	P	112	0.5093353	R	155	1	L	198	1	L
27	0.0942867	I	70	1	P	113	1	R	156	1	G	199	1	L
28	1	E	71	1	T	114	1	Q	157	1	G	200	1	F
29	1	D	72	1	K	115	1	L	158	0.2522404	A	201	1	P
30	1	L	73	0.3950709	K	116	0.1663554	T	159	1	Q	202	0.4981329	A
31	1	E	74	1	P	117	0.4247572	G	160	1	R	203	1	D
32	1	R	75	1	R	118	1	G	161	1	G	204	1	P
33	1	E	76	1	T	119	1	G	162	1	A	205	1	P
34	1	V	77	0.0192307	D	120	1	K	163	1	P	206	1	L
35	1	R	78	1	R	121	0.1125840	R	164	1	G	207	1	S
36	1	R	79	1	M	122	1	L	165	1	G	208	1	P
37	0.0382748	M	80	1	E	123	1	S	166	1	G	209	1	Q
38	1	R	81	0.4247572	I	124	1	K	167	1	F	210	1	S
39	1	R	82	1	D	125	1	E	168	1	A	211	1	C
40	1	K	83	1	P	126	1	E	169	1	P	212	1	R
41	1	I	84	1	A	127	1	K	170	1	?	213	1	P
42	1	R	85	1	S	128	1	G	171	1	L	214	1	Q
43	1	K	86	1	R	129	1	E	172	1	L	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-B*07

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	0.2184466	L	87	1	K	130	1	L	173	0.1153846	N
2	1	S	45	1	E	88	1	R	131	0.1153846	?	174	1	I
3	0.0345516	R	46	0.3509867	D	89	1	P	132	0.2053136	K	175	1	P
4	0.4652481	T	47	0.2184466	S	90	0.6268495	S	133	1	L	176	1	E
5	1	E	48	0.2184466	?	91	0.3515372	R	134	1	A	177	1	S
6	0.6904502	S	49	1	T	92	1	G	135	0.4652481	K	178	1	P
7	1	K	50	1	W	93	0.1848093	E	136	1	E	179	0.1153846	Y
8	0.3103940	R	51	1	L	94	1	F	137	1	D	180	0.2171691	T
9	0.0668068	S	52	1	G	95	1	S	138	1	E	181	1	R
10	1	R	53	1	N	96	0.1411561	D	139	0.3103940	I	182	0.6005875	T
11	0.3103940	R	54	1	V	97	0.2304876	Q	140	1	R	183	1	G
12	0.5918608	G	55	1	K	98	1	E	141	1	E	184	0.4652481	D
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	0.1006446	K	143	1	R	186	1	L
15	1	E	58	1	I	101	1	D	144	1	L	187	1	D
16	0.2184466	A	59	1	G	102	1	H	145	1	T	188	0.3618845	T
17	1	L	60	1	K	103	1	R	146	1	G	189	1	V
18	1	E	61	1	K	104	1	R	147	1	P	190	1	G
19	1	Q	62	1	E	105	1	R	148	0.7257035	P	191	0.2271784	N
20	1	W	63	1	R	106	1	M	149	0.0808585	T	192	0.7467277	Q
21	0.2184466	V	64	1	D	107	1	D	150	1	G	193	1	G
22	0.2184466	T	65	1	G	108	1	L	151	1	G	194	1	F
23	1	A	66	1	G	109	1	E	152	1	V	195	1	P
24	1	R	67	0.1153846	E	110	1	N	153	1	N	196	0.4652481	W
25	0.5918608	K	68	1	A	111	1	K	154	1	P	197	1	D
26	0.6866591	K	69	1	P	112	0.3103940	E	155	1	L	198	0.2080428	I
27	0.1783269	L	70	1	P	113	0.0418893	R	156	1	G	199	1	L
28	1	E	71	1	A	114	1	Q	157	1	A	200	1	F
29	1	D	72	1	K	115	1	L	158	0.2873876	G	201	1	P
30	1	L	73	0.4577882	K	116	0.1153846	D	159	0.1736276	S	202	0.7695027	A
31	1	E	74	1	P	117	0.2852043	G	160	1	R	203	1	E
32	0.1715653	R	75	1	R	118	1	G	161	1	G	204	1	P
33	0.4516060	D	76	0.4652481	V	119	1	G	162	1	A	205	1	P
34	0.2184466	V	77	1	E	120	1	K	163	1	P	206	1	F
35	1	R	78	0.0889591	Q	121	0.1153846	?	164	1	G	207	1	S
36	1	R	79	1	M	122	1	L	165	1	G	208	1	P
37	0.2304876	I	80	1	D	123	1	S	166	1	G	209	1	Q
38	1	K	81	0.4761239	V	124	0.3579355	R	167	1	?	210	1	S
39	0.0345516	K	82	1	D	125	1	G	168	0.1153846	A	211	1	C
40	1	K	83	1	S	126	1	E	169	1	P	212	1	R
41	0.4652481	L	84	1	A	127	0.2304876	K	170	0.2184466	G	213	0.5876094	T
42	1	R	85	1	P	128	1	E	171	1	L	214	1	Q
43	0.2184466	?	86	0.5300665	R	129	1	E	172	1	L	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-B*08

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	1	L	87	1	K	130	1	L	173	1	G
2	1	S	45	1	E	88	1	K	131	0.1924956	R	174	1	I
3	1	Q	46	0.2348747	E	89	1	P	132	1	K	175	1	P
4	1	A	47	1	E	90	0.5876094	S	133	1	L	176	1	E
5	1	G	48	0.1306945	?	91	1	R	134	1	T	177	1	S
6	0.5876094	R	49	1	S	92	1	R	135	1	K	178	1	P
7	1	R	50	1	W	93	1	G	136	1	E	179	1	Y
8	1	K	51	1	L	94	1	F	137	1	D	180	0.0716937	S
9	0.1306945	Y	52	1	G	95	0.3487647	S	138	1	E	181	1	R
10	1	R	53	1	N	96	1	E	139	0.5540427	K	182	1	?
11	0.4798330	E	54	1	I	97	1	E	140	1	R	183	1	G
12	1	D	55	1	K	98	1	E	141	0.6016261	E	184	1	D
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	0.5896782	Q	143	1	R	186	1	L
15	0.5920919	E	58	1	I	101	1	E	144	0.6701407	V	187	1	D
16	0.1306945	A	59	1	G	102	1	H	145	1	T	188	0.5876094	L
17	1	L	60	1	K	103	1	R	146	1	G	189	1	R
18	1	E	61	1	?	104	1	R	147	1	P	190	1	G
19	1	Q	62	1	D	105	1	R	148	0.3952815	Q	191	0.3447934	S
20	1	W	63	1	K	106	1	M	149	0.5181611	P	192	0.3952580	R
21	0.1306945	I	64	1	D	107	1	D	150	1	G	193	1	G
22	0.0399217	S	65	1	G	108	1	L	151	1	G	194	1	F
23	0.2110788	A	66	1	E	109	1	E	152	1	V	195	1	P
24	1	R	67	1	E	110	1	N	153	1	N	196	1	?
25	1	K	68	1	A	111	1	K	154	1	P	197	1	D
26	0.0673076	L	69	1	P	112	1	E	155	1	H	198	0.5896782	L
27	0.1903527	M	70	1	P	113	0.0059561	N	156	1	G	199	1	L
28	1	E	71	1	T	114	1	Q	157	1	A	200	1	F
29	0.2110788	E	72	1	K	115	1	L	158	0.1974818	G	201	1	P
30	1	L	73	1	?	116	0.0220102	N	159	0.1852578	S	202	0.1200957	A
31	1	E	74	1	P	117	0.1245033	A	160	1	R	203	1	E
32	1	K	75	1	R	118	1	G	161	1	G	204	1	P
33	0.3398584	D	76	0.5920919	T	119	1	G	162	1	A	205	1	P
34	1	V	77	1	E	120	1	K	163	1	P	206	1	L
35	1	R	78	0.4389210	R	121	0.3487647	H	164	1	G	207	1	S
36	0.1903527	K	79	1	M	122	1	L	165	1	G	208	1	P
37	0.3486904	T	80	1	D	123	1	S	166	1	G	209	1	Q
38	1	K	81	1	V	124	1	R	167	1	?	210	1	S
39	1	R	82	1	D	125	1	E	168	0.0673076	V	211	1	C
40	1	K	83	1	P	126	1	E	169	1	P	212	1	R
41	0.5896782	I	84	1	A	127	1	E	170	0.0673076	?	213	1	T
42	1	R	85	1	S	128	1	G	171	0.4798330	M	214	1	Q
43	0.5540427	R	86	1	G	129	1	E	172	1	L	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-B*13

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	0.1306945	I	87	1	K	130	1	L	173	1	G
2	1	S	45	1	E	88	1	R	131	0.0059561	E	174	1	I
3	1	Q	46	0.4439532	E	89	1	P	132	0.0716937	K	175	1	P
4	1	A	47	0.0190227	E	90	0.1903527	F	133	1	L	176	1	E
5	1	G	48	1	H	91	0.2464669	K	134	1	A	177	1	S
6	0.1306945	G	49	0.0039208	L	92	1	R	135	1	R	178	1	P
7	0.3831093	R	50	1	W	93	0.0320230	G	136	1	E	179	1	Y
8	1	R	51	1	L	94	1	F	137	1	D	180	0.2442263	T
9	0.1127530	N	52	1	G	95	0.3487647	S	138	1	E	181	1	R
10	1	R	53	1	N	96	0.0035096	D	139	0.1593449	K	182	1	?
11	0.6016261	G	54	1	I	97	0.0092400	Q	140	1	R	183	1	G
12	0.0673076	E	55	1	K	98	1	E	141	0.0978624	K	184	1	G
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	0.0018091	K	143	1	R	186	1	L
15	0.1348802	D	58	0.3952815	I	101	1	D	144	0.0134361	I	187	1	D
16	0.2397472	I	59	1	G	102	1	H	145	1	T	188	0.1154538	I
17	1	L	60	1	K	103	1	R	146	1	G	189	1	R
18	1	E	61	1	?	104	1	R	147	1	P	190	1	G
19	0.0673076	R	62	1	E	105	1	K	148	0.0696639	Q	191	0.1903527	E
20	1	W	63	1	R	106	0.1306945	M	149	0.3831093	T	192	1	H
21	1	V	64	1	D	107	1	T	150	1	G	193	1	G
22	0.0673076	G	65	1	G	108	1	L	151	1	G	194	1	F
23	0.5896782	G	66	1	E	109	1	E	152	1	V	195	1	P
24	1	R	67	1	E	110	1	N	153	1	N	196	1	?
25	0.3487647	R	68	1	A	111	1	K	154	1	P	197	1	D
26	0.2653287	R	69	1	P	112	0.1306945	S	155	0.0673076	H	198	0.5896782	L
27	0.2464669	Q	70	1	P	113	0.0043444	R	156	1	G	199	1	L
28	1	E	71	1	A	114	1	Q	157	1	A	200	1	F
29	1	D	72	1	K	115	1	L	158	0.0673076	V	201	1	P
30	1	L	73	0.1797945	K	116	0.0673076	E	159	0.2464669	T	202	0.1200957	A
31	1	E	74	1	P	117	0.1444127	S	160	1	R	203	1	?
32	0.5920919	K	75	1	R	118	1	G	161	1	G	204	1	P
33	0.0001109	E	76	0.0220102	A	119	1	G	162	1	A	205	1	P
34	0.0673076	S	77	1	E	120	1	K	163	1	P	206	1	L
35	1	R	78	1	P	121	0.0124415	S	164	1	G	207	1	S
36	0.1903527	K	79	1	M	122	1	L	165	1	G	208	1	P
37	0.0898052	I	80	1	D	123	1	S	166	1	G	209	1	Q
38	0.2992142	R	81	0.0554768	I	124	0.1122636	K	167	1	?	210	1	S
39	1	R	82	1	D	125	0.1306945	G	168	1	A	211	1	C
40	1	K	83	1	S	126	1	E	169	1	P	212	1	R
41	0.0186039	V	84	1	G	127	0.0092400	K	170	0.6605051	N	213	1	P
42	1	R	85	1	S	128	1	E	171	0.4798330	L	214	1	Q
43	0.0021108	R	86	0.0514617	G	129	1	E	172	1	L	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-B*14

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	1	L	87	1	K	130	1	L	173	1	G
2	1	S	45	1	E	88	0.1125840	R	131	0.3487647	E	174	1	I
3	1	Q	46	0.3967320	D	89	1	P	132	0.2297631	K	175	1	P
4	0.0576923	P	47	0.6894419	E	90	0.1647849	P	133	1	L	176	1	E
5	1	E	48	1	H	91	0.3487647	G	134	1	T	177	1	S
6	0.1125840	K	49	1	P	92	1	R	135	0.0112213	E	178	1	P
7	0.1823207	K	50	1	W	93	1	G	136	0.2615375	E	179	1	F
8	1	K	51	1	L	94	1	F	137	1	D	180	0.0453488	S
9	0.0576923	A	52	1	G	95	1	T	138	1	E	181	1	R
10	1	R	53	1	N	96	1	D	139	0.0174821	R	182	1	?
11	0.4272318	E	54	1	V	97	0.1647849	E	140	1	R	183	1	G
12	0.0576923	S	55	1	K	98	1	E	141	1	K	184	0.2615375	D
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	0.1841838	Q	143	1	R	186	1	L
15	0.4634067	D	58	0.3487647	L	101	1	D	144	1	A	187	1	D
16	0.0886134	I	59	1	G	102	1	H	145	1	A	188	0.0382153	I
17	1	L	60	1	K	103	1	R	146	1	G	189	1	R
18	1	E	61	1	K	104	1	R	147	1	P	190	1	G
19	1	K	62	1	D	105	1	R	148	0.1125840	R	191	0.2281434	N
20	1	W	63	1	K	106	1	K	149	0.2071044	V	192	0.0948345	Q
21	1	V	64	1	D	107	0.0028005	T	150	1	G	193	1	G
22	0.1125840	?	65	1	G	108	1	L	151	1	G	194	1	F
23	1	G	66	1	E	109	1	E	152	1	V	195	1	P
24	1	R	67	1	E	110	1	N	153	1	N	196	1	W
25	1	N	68	1	A	111	1	K	154	1	P	197	1	D
26	0.5869120	K	69	1	P	112	1	E	155	1	L	198	0.5860319	I
27	0.0576923	?	70	1	P	113	0.3487647	N	156	1	E	199	1	L
28	1	E	71	0.2144016	T	114	1	Q	157	1	G	200	1	F
29	1	E	72	1	K	115	1	L	158	0.1841838	A	201	1	P
30	1	L	73	0.3320885	R	116	0.3062928	Q	159	0.3062928	Q	202	0.2135126	A
31	1	E	74	0.1125840	S	117	0.0576923	T	160	1	R	203	1	D
32	0.1622925	R	75	1	R	118	1	G	161	1	G	204	1	P
33	0.3438548	D	76	1	V	119	1	G	162	1	A	205	1	P
34	1	S	77	1	D	120	1	K	163	1	P	206	1	F
35	1	R	78	1	P	121	0.3062928	R	164	1	G	207	1	S
36	0.1647849	R	79	1	M	122	1	L	165	1	G	208	1	P
37	0.0028005	?	80	1	E	123	1	S	166	1	G	209	1	Q
38	0.2615375	K	81	0.3316195	V	124	0.6781231	R	167	1	?	210	1	S
39	1	R	82	1	D	125	1	E	168	1	A	211	1	C
40	1	K	83	1	H	126	1	E	169	1	P	212	1	R
41	0.4634067	V	84	1	G	127	1	E	170	0.1125840	G	213	0.3487647	P
42	1	R	85	1	S	128	1	E	171	0.0835710	L	214	1	Q
43	1	R	86	1	R	129	1	E	172	0.1647849	L	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-B*15

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	1	I	87	1	K	130	1	L	173	1	N
2	1	S	45	1	E	88	1	K	131	0.6926836	R	174	1	V
3	1	R	46	1	D	89	1	L	132	0.3477077	R	175	1	P
4	0.0463534	T	47	0.1486183	S	90	0.3890473	H	133	1	L	176	1	E
5	1	G	48	1	?	91	0.5918608	R	134	1	I	177	1	S
6	0.1486183	K	49	1	T	92	1	R	135	1	K	178	1	P
7	0.0769230	E	50	1	W	93	0.5918608	G	136	1	D	179	1	Y
8	1	D	51	1	L	94	1	F	137	1	D	180	0.1395755	T
9	0.0876636	D	52	1	G	95	0.3890473	T	138	1	E	181	1	R
10	1	R	53	1	N	96	0.3890473	D	139	0.1981169	R	182	0.4389210	H
11	0.2153934	R	54	1	I	97	0.4851955	Q	140	1	R	183	1	G
12	0.2153934	D	55	1	K	98	1	E	141	0.3452353	K	184	1	D
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	0.3353372	R	143	1	R	186	1	L
15	0.5678352	D	58	0.0898052	L	101	1	D	144	0.1419926	I	187	1	D
16	0.4851955	T	59	1	G	102	1	H	145	1	T	188	1	T
17	1	L	60	1	K	103	1	R	146	1	G	189	1	R
18	1	E	61	1	?	104	1	R	147	1	P	190	1	G
19	1	R	62	1	E	105	1	R	148	0.1961384	P	191	0.5280959	T
20	1	W	63	0.2775404	K	106	1	K	149	0.0769230	I	192	0.0150683	H
21	1	V	64	1	D	107	1	A	150	1	G	193	1	G
22	0.0680561	S	65	1	G	108	1	L	151	1	G	194	1	F
23	0.0769230	E	66	0.0769230	D	109	1	E	152	1	V	195	1	P
24	1	R	67	0.0769230	G	110	1	N	153	1	N	196	1	W
25	1	K	68	1	A	111	1	K	154	1	P	197	1	D
26	1	R	69	1	P	112	0.0769230	Q	155	1	S	198	0.0108910	L
27	0.2775404	Q	70	1	P	113	0.1945643	K	156	1	E	199	1	L
28	0.1486183	D	71	0.0769230	P	114	1	Q	157	1	A	200	1	F
29	1	E	72	1	K	115	1	L	158	0.6780049	G	201	1	P
30	1	L	73	1	K	116	0.2775404	N	159	0.3890473	P	202	0.0769230	P
31	1	E	74	1	S	117	0.6614962	S	160	1	R	203	1	E
32	0.2615417	R	75	1	R	118	1	G	161	1	G	204	1	P
33	0.0769230	N	76	0.5914911	T	119	1	G	162	1	A	205	1	P
34	1	V	77	1	E	120	1	K	163	1	P	206	0.0769230	F
35	1	R	78	0.1149991	R	121	0.0769230	D	164	1	G	207	1	S
36	1	R	79	1	M	122	1	L	165	1	G	208	1	P
37	0.0769230	S	80	1	D	123	1	S	166	1	G	209	1	Q
38	0.3353372	R	81	0.6762549	I	124	1	R	167	1	?	210	1	S
39	0.2153934	K	82	1	D	125	1	E	168	1	A	211	1	C
40	1	K	83	1	A	126	1	E	169	1	P	212	1	R
41	0.5678352	V	84	1	?	127	0.4851955	K	170	2,86E+08	N	213	0.0898052	P
42	0.2153934	K	85	1	S	128	1	E	171	0.5280959	M	214	1	Q
43	1	K	86	1	G	129	1	E	172	1	L	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-B*18

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	1	I	87	1	K	130	1	L	173	1	G
2	1	S	45	1	E	88	1	R	131	0.1346153	?	174	0.7300066	V
3	1	Q	46	0.1462666	E	89	0.0866039	P	132	0.1346153	?	175	1	P
4	0.3548741	A	47	0.0026978	D	90	0.1841838	H	133	1	L	176	1	E
5	1	E	48	1	N	91	0.1346153	A	134	0.3548741	A	177	1	S
6	0.0677886	S	49	0.3548741	S	92	1	G	135	0.1331041	K	178	1	P
7	0.3397362	K	50	1	W	93	0.1346153	A	136	0.5220958	D	179	1	F
8	1	E	51	1	L	94	1	F	137	1	D	180	0.1317533	H
9	0.0073423	G	52	1	G	95	0.5896782	T	138	1	E	181	1	R
10	1	R	53	1	N	96	0.1841838	E	139	0.0469731	I	182	0.2379739	H
11	0.0866039	D	54	0.5896782	I	97	0.2929655	Q	140	1	R	183	1	G
12	0.1346153	S	55	1	K	98	1	E	141	0.0643701	E	184	1	D
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	0.1331041	K	143	1	R	186	1	L
15	0.1346153	S	58	0.5896782	L	101	0.2522404	E	144	0.1055631	V	187	1	D
16	0.3850895	V	59	1	G	102	1	H	145	1	A	188	0.0930720	T
17	1	L	60	1	K	103	1	R	146	1	G	189	0.2522404	T
18	1	E	61	1	K	104	1	R	147	1	P	190	1	G
19	1	E	62	0.2522404	E	105	0.2522404	K	148	1	?	191	0.2805302	S
20	1	W	63	1	K	106	1	M	149	0.1946963	T	192	0.3548741	H
21	1	V	64	1	D	107	1	T	150	1	G	193	1	G
22	0.2522404	T	65	1	G	108	1	L	151	1	D	194	1	F
23	0.3770538	A	66	1	G	109	1	E	152	1	V	195	1	P
24	1	R	67	1	G	110	1	N	153	1	N	196	0.1331041	W
25	0.2522404	N	68	1	A	111	1	K	154	1	P	197	1	D
26	0.4213101	R	69	1	P	112	0.2210582	R	155	0.1346153	S	198	0.1346153	K
27	0.0497086	A	70	1	P	113	0.6865430	R	156	1	E	199	1	L
28	0.2522404	D	71	1	P	114	1	Q	157	0.1346153	G	200	1	F
29	0.6865430	E	72	1	K	115	1	L	158	0.0469731	E	201	1	P
30	1	L	73	0.2957556	K	116	0.0059790	A	159	0.0866039	R	202	1	?
31	1	E	74	0.4442975	P	117	0.0444455	A	160	1	R	203	1	E
32	0.3770538	R	75	1	R	118	1	G	161	1	G	204	1	P
33	0.7170617	E	76	0.2068347	T	119	1	G	162	1	A	205	1	P
34	1	V	77	1	D	120	1	K	163	1	P	206	1	F
35	1	R	78	0.2929655	R	121	0.0169902	T	164	1	G	207	1	S
36	1	R	79	1	M	122	1	L	165	1	G	208	1	P
37	0.2929655	A	80	1	E	123	1	S	166	1	G	209	1	Q
38	0.5220958	R	81	0.0375979	I	124	0.0413346	K	167	1	F	210	1	S
39	1	K	82	1	D	125	1	E	168	1	V	211	1	C
40	1	K	83	1	P	126	1	E	169	1	P	212	1	R
41	0.2123388	I	84	1	A	127	0.2929655	E	170	0.1801946	N	213	1	P
42	1	K	85	1	P	128	0.2522404	G	171	0.6050124	M	214	1	Q
43	0.2068347	K	86	0.1841838	G	129	1	E	172	1	Q	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-B*27

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	1	L	87	1	K	130	1	L	173	1	G
2	1	S	45	1	E	88	1	R	131	0.3487647	E	174	0.5930882	I
3	1	R	46	0.6894419	E	89	0.1125840	T	132	1	R	175	1	P
4	0.2615375	T	47	0.2183065	D	90	0.5300665	S	133	1	L	176	1	E
5	1	G	48	1	?	91	0.3487647	G	134	1	T	177	1	S
6	0.5892735	S	49	1	S	92	1	G	135	0.1016584	E	178	1	P
7	1	E	50	1	W	93	1	G	136	1	E	179	1	Y
8	1	K	51	1	L	94	1	F	137	1	D	180	0.4109930	T
9	0.2144016	G	52	1	G	95	1	T	138	1	E	181	1	R
10	1	R	53	1	N	96	1	E	139	0.0377611	E	182	1	?
11	0.5868715	G	54	1	I	97	0.1647849	E	140	1	R	183	1	G
12	0.1647849	D	55	1	K	98	1	E	141	0.2770252	E	184	1	E
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	0.2615375	R	143	1	R	186	1	L
15	0.4634067	D	58	1	L	101	1	D	144	0.2532340	T	187	1	D
16	0.0825674	V	59	1	G	102	1	H	145	1	A	188	0.4145631	V
17	1	L	60	1	K	103	1	R	146	1	G	189	1	R
18	1	E	61	1	?	104	1	R	147	1	P	190	1	G
19	1	K	62	1	D	105	0.0028005	K	148	0.3327476	P	191	0.2615375	G
20	1	W	63	0.0159347	K	106	0.1125840	K	149	0.4634067	P	192	0.1125840	G
21	1	V	64	1	D	107	1	A	150	1	G	193	1	G
22	0.6781231	N	65	1	G	108	1	L	151	1	G	194	1	F
23	1	G	66	0.1125840	G	109	1	E	152	1	V	195	1	P
24	1	R	67	1	E	110	1	N	153	1	N	196	1	*
25	1	R	68	1	A	111	1	K	154	1	P	197	1	D
26	1	R	69	1	P	112	0.1125840	S	155	1	L	198	0.5896782	L
27	0.1647849	M	70	1	P	113	0.0514617	N	156	0.1125840	E	199	1	L
28	1	E	71	1	P	114	1	Q	157	1	A	200	1	F
29	0.1622925	E	72	1	K	115	1	L	158	0.5896782	A	201	1	P
30	1	L	73	0.3320885	R	116	0.0159347	R	159	0.0436276	S	202	0.6840312	A
31	1	E	74	1	A	117	0.6873226	A	160	1	R	203	1	D
32	1	R	75	1	R	118	1	G	161	1	G	204	1	P
33	0.3438548	D	76	0.0576923	P	119	1	G	162	1	A	205	1	P
34	0.1125840	V	77	1	D	120	1	K	163	1	P	206	1	F
35	1	R	78	1	R	121	0.3062928	R	164	1	G	207	1	S
36	1	K	79	1	M	122	1	L	165	1	G	208	1	P
37	0.3890473	I	80	1	D	123	1	S	166	1	G	209	1	Q
38	1	K	81	0.3316195	V	124	0.2049641	R	167	1	F	210	1	S
39	1	K	82	1	D	125	1	E	168	1	A	211	1	C
40	1	K	83	0.1647849	P	126	1	E	169	1	P	212	1	R
41	0.2615375	L	84	0.0576923	A	127	0.3890473	E	170	0.3320885	N	213	0.3487647	P
42	1	R	85	0.0576923	S	128	1	E	171	1	M	214	1	Q
43	1	R	86	1	R	129	1	E	172	1	L	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-B*35

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	0.1312546	L	87	1	K	130	1	L	173	1	S
2	1	S	45	1	E	88	1	K	131	0.1312546	G	174	0.3050730	I
3	0.5523986	R	46	0.4099066	D	89	0.3653846	I	132	0.0263216	R	175	1	P
4	0.3653846	P	47	0.0655867	D	90	0.0463251	F	133	1	L	176	1	E
5	0.3653846	E	48	0.2976101	H	91	0.2555179	G	134	0.3653846	I	177	1	S
6	0.3013697	S	49	0.2555179	P	92	0.3653846	R	135	0.3653846	I	178	1	P
7	0.4928560	R	50	1	W	93	0.3653846	R	136	1	D	179	0.3653846	Y
8	0.3523813	K	51	1	L	94	1	F	137	1	D	180	0.3653846	?
9	0.1586911	D	52	1	G	95	0.0835304	T	138	1	E	181	1	R
10	1	R	53	1	N	96	0.4116855	E	139	0.1446444	R	182	0.7430464	T
11	0.1499777	E	54	1	V	97	0.5285952	K	140	1	R	183	1	G
12	0.1312546	N	55	1	K	98	1	E	141	0.1903148	E	184	0.3523813	D
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	0.1371408	E	143	1	R	186	1	L
15	0.1647764	D	58	0.7042527	I	101	1	D	144	0.3639941	V	187	1	D
16	0.4593211	T	59	1	G	102	1	H	145	0.3653846	A	188	0.0895201	V
17	1	L	60	1	K	103	1	R	146	1	G	189	0.3653846	G
18	1	E	61	1	?	104	1	R	147	1	P	190	1	G
19	0.3653846	E	62	0.5317401	D	105	0.5317401	R	148	0.0252391	S	191	0.2816477	T
20	1	W	63	0.2938829	K	106	0.5317401	K	149	0.3653846	G	192	0.2976101	H
21	1	I	64	1	D	107	0.1312546	T	150	1	G	193	1	G
22	0.3653846	G	65	1	G	108	1	L	151	0.1312546	G	194	1	F
23	0.5409899	A	66	0.2976101	E	109	1	E	152	1	V	195	1	P
24	1	R	67	1	G	110	1	N	153	1	N	196	0.4593211	*
25	0.4116855	R	68	1	A	111	1	K	154	1	P	197	1	D
26	0.7780306	R	69	1	P	112	0.2976101	E	155	0.3653846	S	198	0.0531213	I
27	0.2555179	A	70	1	P	113	0.1312546	?	156	1	G	199	1	L
28	0.5317401	E	71	0.6501914	A	114	1	Q	157	0.3653846	G	200	1	F
29	0.0285618	D	72	1	K	115	1	L	158	0.5317401	D	201	1	P
30	1	L	73	0.3653846	?	116	0.2816477	T	159	0.0835304	P	202	0.3653846	?
31	1	E	74	0.1312546	S	117	0.6130455	S	160	1	R	203	0.5317401	D
32	0.2197463	K	75	1	R	118	1	G	161	1	G	204	1	P
33	0.3653846	?	76	0.3653846	L	119	1	G	162	1	A	205	1	P
34	1	V	77	0.3653846	D	120	1	K	163	1	P	206	1	F
35	1	R	78	0.3653846	H	121	0.3653846	?	164	1	G	207	1	S
36	0.2976101	K	79	1	M	122	1	L	165	1	G	208	1	P
37	0.0257222	I	80	0.3653846	D	123	1	S	166	1	G	209	1	Q
38	1	R	81	0.4754029	I	124	0.1536074	K	167	1	?	210	1	S
39	1	K	82	1	D	125	0.1312546	E	168	1	A	211	1	C
40	1	K	83	0.3523813	S	126	1	E	169	1	P	212	1	R
41	0.3523813	L	84	0.3653846	?	127	0.2527799	K	170	0.3363256	N	213	1	P
42	0.5523986	K	85	1	S	128	0.5317401	E	171	0.7210219	L	214	1	Q
43	0.1312546	?	86	0.1878768	G	129	1	E	172	1	L	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-B*37

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	1	I	87	1	K	130	1	L	173	1	S
2	1	S	45	1	E	88	1	R	131	1	G	174	0.4767605	I
3	1	Q	46	0.2654032	E	89	0.0033167	P	132	0.3973992	K	175	1	P
4	1	A	47	0.5678842	E	90	0.3103940	S	133	1	L	176	1	E
5	1	G	48	1	?	91	0.0288461	A	134	1	T	177	1	S
6	0.3103940	R	49	1	T	92	1	G	135	1	E	178	1	P
7	1	K	50	1	W	93	0.0288461	A	136	0.1386844	D	179	1	F
8	1	E	51	1	L	94	1	F	137	1	D	180	0.1427755	A
9	0.3763783	D	52	1	G	95	1	T	138	1	E	181	1	R
10	1	R	53	1	N	96	1	D	139	0.0848635	I	182	1	T
11	1	E	54	1	I	97	1	K	140	1	R	183	1	G
12	1	N	55	1	K	98	1	E	141	0.4380134	K	184	1	E
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	0.1120458	E	143	1	R	186	1	L
15	1	N	58	1	I	101	0.0005601	E	144	0.5580821	V	187	1	D
16	0.5880815	I	59	1	G	102	1	H	145	1	T	188	0.2744585	V
17	1	L	60	1	K	103	1	R	146	1	G	189	1	T
18	1	E	61	1	K	104	1	R	147	1	P	190	1	G
19	1	K	62	1	D	105	1	R	148	0.3973992	S	191	0.0851381	S
20	1	W	63	1	R	106	1	M	149	1	P	192	0.5491038	Q
21	1	V	64	1	D	107	1	T	150	1	G	193	1	G
22	1	?	65	1	G	108	1	L	151	1	D	194	1	F
23	1	G	66	1	G	109	1	E	152	1	V	195	1	P
24	1	R	67	1	G	110	1	N	153	1	N	196	1	W
25	1	N	68	1	A	111	1	K	154	1	P	197	1	D
26	1	R	69	1	P	112	0.2976101	K	155	1	L	198	0.0469731	L
27	0.5661709	L	70	1	P	113	1	?	156	1	G	199	1	L
28	1	D	71	0.1120458	T	114	1	Q	157	1	G	200	1	F
29	1	E	72	1	K	115	1	L	158	1	S	201	1	P
30	1	L	73	0.1222598	K	116	0.1120458	R	159	0.1647849	Q	202	1	V
31	1	E	74	1	S	117	0.4954476	S	160	1	R	203	1	D
32	1	R	75	1	R	118	1	G	161	1	G	204	1	P
33	1	D	76	0.1386844	V	119	1	G	162	1	A	205	1	P
34	1	V	77	1	E	120	1	K	163	1	P	206	1	F
35	1	R	78	1	R	121	0.5571761	N	164	1	G	207	1	S
36	1	K	79	1	M	122	1	L	165	1	G	208	1	P
37	0.2654032	V	80	1	E	123	1	S	166	1	G	209	1	Q
38	1	R	81	0.5657481	I	124	1	K	167	1	F	210	1	S
39	1	K	82	1	D	125	1	E	168	1	A	211	1	C
40	1	K	83	1	A	126	1	E	169	1	P	212	1	R
41	1	V	84	1	G	127	1	K	170	0.5661709	S	213	1	T
42	1	K	85	1	S	128	1	G	171	1	L	214	1	Q
43	0.0571321	?	86	1	G	129	1	E	172	1	Q	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-B*38

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	1	L	87	1	K	130	1	L	173	0.0673076	N
2	1	S	45	1	E	88	1	K	131	0.0039208	G	174	0.0244326	I
3	1	Q	46	1	D	89	1	P	132	1	R	175	1	P
4	0.0354776	T	47	0.1224822	E	90	0.1848093	S	133	1	L	176	1	E
5	1	G	48	0.1306945	?	91	0.3952815	G	134	1	T	177	1	S
6	0.1306945	G	49	1	S	92	1	R	135	0.2992142	K	178	1	P
7	1	R	50	1	W	93	0.0673076	V	136	1	E	179	1	Y
8	1	R	51	1	L	94	1	F	137	1	D	180	0.1348802	H
9	0.0673076	T	52	1	G	95	1	S	138	1	E	181	1	R
10	1	R	53	1	N	96	1	E	139	0.1903527	I	182	0.3952815	H
11	0.6016261	G	54	0.0696639	V	97	0.4389210	Q	140	1	R	183	1	G
12	0.0673076	R	55	1	K	98	1	E	141	1	K	184	0.2992142	D
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	0.5896782	Q	143	1	R	186	1	L
15	0.1306945	N	58	1	I	101	1	E	144	0.0673076	?	187	1	D
16	0.4626758	V	59	1	G	102	1	H	145	1	T	188	0.0497086	T
17	1	L	60	1	K	103	1	R	146	1	G	189	1	R
18	1	E	61	0.0673076	K	104	1	R	147	1	P	190	1	G
19	0.1306945	K	62	1	D	105	1	K	148	0.1306945	R	191	0.0410930	D
20	1	W	63	1	R	106	1	M	149	0.1348802	P	192	0.0172634	R
21	0.1306945	I	64	1	D	107	1	D	150	1	G	193	1	G
22	0.0673076	E	65	1	G	108	1	L	151	0.1306945	D	194	1	F
23	1	A	66	1	E	109	1	E	152	1	V	195	1	P
24	1	R	67	1	E	110	1	N	153	1	N	196	0.1903527	?
25	0.0514617	R	68	1	A	111	1	K	154	1	P	197	1	D
26	0.5896782	R	69	1	P	112	1	E	155	1	L	198	0.2379739	L
27	0.0696639	V	70	1	P	113	1	R	156	1	G	199	1	L
28	1	E	71	1	A	114	1	Q	157	1	A	200	1	F
29	0.5920919	D	72	1	K	115	1	L	158	0.1306945	D	201	1	P
30	1	L	73	1	R	116	0.3487647	Q	159	0.0673076	?	202	0.1304081	S
31	1	E	74	0.2464669	P	117	0.3398584	S	160	1	R	203	1	D
32	0.5920919	K	75	1	R	118	1	G	161	1	G	204	1	P
33	0.3398584	D	76	0.5920919	T	119	1	G	162	1	A	205	1	P
34	1	S	77	1	E	120	1	K	163	1	P	206	1	L
35	1	R	78	0.4389210	R	121	0.1306945	I	164	1	G	207	1	S
36	1	R	79	1	M	122	1	L	165	1	G	208	1	P
37	0.0898052	I	80	0.0673076	E	123	1	S	166	1	G	209	1	Q
38	1	K	81	1	V	124	1	K	167	1	?	210	1	S
39	0.1903527	K	82	1	D	125	1	G	168	1	A	211	1	C
40	1	K	83	0.1903527	P	126	1	E	169	1	P	212	1	R
41	0.1348802	V	84	0.0673076	?	127	1	E	170	0.6735451	S	213	1	P
42	1	R	85	1	S	128	1	G	171	0.4798330	M	214	1	Q
43	0.5920919	K	86	0.3487647	G	129	1	E	172	1	L	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-B*39

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	1	I	87	1	K	130	1	L	173	1	S
2	1	S	45	1	E	88	1	K	131	0.0751582	K	174	1	I
3	1	Q	46	0.0751582	D	89	0.0480769	L	132	0.5885256	R	175	1	P
4	1	A	47	0.1570178	E	90	0.1386844	P	133	1	L	176	1	E
5	1	E	48	1	H	91	0.2992142	G	134	0.1386844	A	177	1	S
6	0.0413257	S	49	1	S	92	0.0480769	G	135	1	D	178	1	P
7	0.6093702	R	50	1	W	93	0.0942867	D	136	1	E	179	1	Y
8	1	R	51	1	L	94	1	F	137	1	D	180	0.0114446	A
9	0.1503926	D	52	1	G	95	1	T	138	1	E	181	1	R
10	1	R	53	1	N	96	1	E	139	0.0480769	G	182	1	H
11	0.5843525	G	54	1	V	97	1	E	140	1	R	183	1	G
12	1	D	55	1	K	98	1	E	141	0.2056514	E	184	1	E
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	0.1813238	E	143	1	R	186	1	L
15	1	D	58	1	I	101	1	E	144	0.3186006	V	187	1	D
16	0.6560548	I	59	1	G	102	1	H	145	1	A	188	0.4652481	L
17	1	L	60	1	K	103	1	R	146	1	G	189	1	V
18	1	E	61	1	?	104	1	R	147	1	P	190	1	G
19	1	R	62	1	D	105	1	K	148	0.0133954	P	191	0.0175886	G
20	1	W	63	1	R	106	1	M	149	0.4031153	A	192	0.6167128	R
21	1	V	64	1	D	107	1	T	150	1	G	193	1	G
22	0.6479857	S	65	1	G	108	1	L	151	1	G	194	1	F
23	1	G	66	1	G	109	1	E	152	1	V	195	1	P
24	1	R	67	1	G	110	1	N	153	1	N	196	1	*
25	1	K	68	1	A	111	1	K	154	1	P	197	1	D
26	1	R	69	1	P	112	0.1556140	K	155	1	S	198	0.0299809	I
27	0.3234844	L	70	1	P	113	0.4943107	R	156	1	E	199	1	L
28	1	E	71	0.0107663	T	114	1	Q	157	1	G	200	1	F
29	1	E	72	1	K	115	1	L	158	0.5872457	G	201	1	P
30	1	L	73	0.0097059	R	116	0.5967155	A	159	0.0258580	Q	202	0.6479857	S
31	1	E	74	0.0942867	S	117	0.2646524	S	160	1	R	203	1	E
32	1	K	75	1	R	118	1	G	161	1	G	204	1	P
33	0.5803723	D	76	0.0175886	V	119	1	G	162	1	A	205	1	P
34	1	V	77	1	D	120	1	K	163	1	P	206	1	F
35	1	R	78	1	R	121	0.2615375	H	164	1	G	207	1	S
36	1	R	79	1	M	122	1	L	165	1	G	208	1	P
37	0.0751582	V	80	1	D	123	1	S	166	1	G	209	1	Q
38	1	R	81	1	I	124	0.3626838	R	167	1	F	210	1	S
39	1	R	82	1	D	125	1	G	168	1	V	211	1	C
40	1	K	83	1	P	126	1	E	169	1	P	212	1	R
41	1	V	84	1	?	127	1	E	170	1	G	213	1	T
42	1	R	85	1	S	128	0.0942867	G	171	1	M	214	1	Q
43	0.4348645	R	86	0.2615375	G	129	1	E	172	1	L	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-B*40

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	1	I	87	1	K	130	1	L	173	1	G
2	1	S	45	1	E	88	1	R	131	0.6330969	R	174	0.2446701	V
3	1	Q	46	0.6466063	E	89	1	L	132	0.1682916	K	175	1	P
4	1	A	47	1	D	90	0.4652481	S	133	1	L	176	1	E
5	1	E	48	1	H	91	1	K	134	1	A	177	1	S
6	0.1006446	R	49	1	S	92	1	G	135	1	K	178	1	P
7	0.1195158	R	50	1	W	93	1	E	136	1	E	179	1	Y
8	1	R	51	1	L	94	1	F	137	1	D	180	0.3703521	T
9	0.1503926	D	52	1	G	95	1	T	138	1	E	181	1	R
10	1	R	53	1	N	96	1	D	139	0.4348645	K	182	0.2992142	H
11	0.5843525	G	54	0.0354776	I	97	1	E	140	1	R	183	1	G
12	1	D	55	1	K	98	1	E	141	0.5843525	E	184	1	G
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	1	E	143	1	R	186	1	L
15	0.0942867	N	58	1	I	101	1	E	144	0.0480769	L	187	1	D
16	0.1633149	V	59	1	G	102	1	H	145	0.0480769	A	188	0.1331041	T
17	1	L	60	1	K	103	1	R	146	1	G	189	1	T
18	1	E	61	1	?	104	1	R	147	1	P	190	1	G
19	1	E	62	1	D	105	1	R	148	0.2992142	Q	191	0.2249839	S
20	1	W	63	1	R	106	1	M	149	0.4031153	A	192	0.6167128	R
21	1	V	64	1	D	107	1	T	150	1	G	193	1	G
22	0.0480769	E	65	1	G	108	1	L	151	0.0942867	D	194	1	F
23	1	G	66	1	D	109	1	E	152	1	V	195	1	P
24	1	R	67	1	G	110	1	N	153	1	N	196	1	*
25	0.0258580	R	68	1	A	111	1	K	154	1	P	197	1	D
26	1	K	69	1	P	112	0.6501914	K	155	1	H	198	0.5885256	I
27	0.0107663	Q	70	1	P	113	0.5838954	K	156	0.0942867	G	199	1	L
28	1	D	71	1	T	114	1	Q	157	1	G	200	1	F
29	0.4943107	D	72	1	K	115	1	L	158	0.5872457	G	201	1	P
30	1	L	73	0.3052863	K	116	0.1813238	R	159	0.2615375	Q	202	0.6479857	S
31	1	E	74	0.1813238	P	117	1	G	160	1	R	203	1	E
32	0.4943107	R	75	1	R	118	1	G	161	1	G	204	1	P
33	1	E	76	1	A	119	1	G	162	1	A	205	1	P
34	1	S	77	1	D	120	1	K	163	1	P	206	1	L
35	1	R	78	0.3353372	R	121	0.2615375	H	164	1	G	207	1	S
36	1	R	79	1	M	122	1	L	165	1	G	208	1	P
37	0.0480769	L	80	1	D	123	1	S	166	1	G	209	1	Q
38	1	K	81	0.3335738	V	124	1	K	167	1	F	210	1	S
39	1	R	82	1	D	125	1	G	168	1	V	211	1	C
40	1	K	83	1	P	126	1	E	169	1	P	212	1	R
41	1	I	84	1	G	127	1	E	170	0.3258347	N	213	1	T
42	1	R	85	1	P	128	1	G	171	1	L	214	1	Q
43	0.4348645	R	86	1	G	129	1	E	172	1	L	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-B*41

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	1	I	87	1	K	130	1	L	173	1	D
2	1	S	45	1	E	88	1	R	131	0.6406166	K	174	0.0218795	V
3	1	R	46	1	E	89	1	I	132	1	?	175	1	P
4	1	T	47	0.6247871	D	90	0.0758028	A	133	1	L	176	1	E
5	1	E	48	1	N	91	1	A	134	1	I	177	1	S
6	0.0758028	A	49	1	T	92	1	G	135	1	I	178	1	P
7	1	E	50	1	W	93	1	R	136	0.1813238	E	179	1	F
8	1	E	51	1	L	94	1	F	137	1	D	180	0.3367948	H
9	0.2775404	K	52	1	G	95	1	T	138	1	E	181	1	R
10	1	R	53	1	N	96	1	E	139	0.0034089	K	182	0.2464669	H
11	0.1375453	G	54	1	I	97	1	R	140	1	R	183	1	G
12	1	R	55	1	K	98	1	E	141	1	K	184	0.1813238	D
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	1	R	143	1	R	186	1	L
15	0.0384615	S	58	1	L	101	1	D	144	0.5741254	V	187	1	D
16	0.3257044	I	59	1	G	102	1	H	145	1	T	188	0.4442975	T
17	1	L	60	1	K	103	1	R	146	1	G	189	1	G
18	1	E	61	1	K	104	1	R	147	1	P	190	1	G
19	0.0758028	K	62	1	D	105	1	R	148	0.5698299	P	191	0.5596499	S
20	1	W	63	1	R	106	1	K	149	0.2928912	T	192	1	L
21	1	V	64	1	D	107	1	D	150	1	G	193	1	G
22	1	E	65	1	G	108	1	L	151	1	D	194	1	F
23	1	E	66	1	E	109	1	E	152	1	V	195	1	P
24	1	R	67	1	G	110	1	N	153	1	N	196	1	W
25	1	N	68	1	A	111	1	K	154	1	P	197	1	D
26	0.4689954	R	69	1	P	112	1	?	155	1	H	198	1	?
27	0.1813238	I	70	1	P	113	0.0752808	R	156	1	E	199	1	L
28	0.0758028	E	71	0.0065463	T	114	1	Q	157	1	G	200	1	F
29	0.4187479	D	72	1	K	115	1	L	158	0.0023109	G	201	1	P
30	1	L	73	0.5711644	R	116	0.0367206	T	159	0.0586001	S	202	1	P
31	1	E	74	1	S	117	0.2432838	G	160	1	R	203	1	E
32	1	K	75	1	R	118	1	G	161	1	G	204	1	P
33	0.5596499	E	76	1	S	119	1	G	162	1	A	205	1	P
34	1	L	77	1	D	120	1	K	163	1	P	206	1	F
35	1	R	78	1	H	121	0.1573923	S	164	1	G	207	1	S
36	1	K	79	1	M	122	1	L	165	1	G	208	1	P
37	0.1120458	E	80	1	E	123	1	S	166	1	G	209	1	Q
38	1	R	81	0.5700476	I	124	0.6176103	R	167	1	?	210	1	S
39	1	K	82	1	D	125	1	E	168	1	V	211	1	C
40	1	K	83	1	H	126	1	E	169	1	P	212	1	R
41	1	L	84	1	?	127	1	E	170	0.5704663	S	213	0.2464669	P
42	1	K	85	1	P	128	1	G	171	1	M	214	1	Q
43	1	?	86	1	G	129	1	E	172	0.1120458	Q	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-B*44

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	1	I	87	1	K	130	1	L	173	1	G
2	1	S	45	1	E	88	1	K	131	1	R	174	0.7115139	V
3	1	R	46	0.5508605	D	89	1	P	132	0.6860398	K	175	1	P
4	0.5978741	S	47	0.1353735	D	90	0.5607143	H	133	1	L	176	1	E
5	1	G	48	0.3328812	H	91	1	K	134	0.3328812	A	177	1	S
6	0.2354368	A	49	0.5920919	P	92	1	G	135	0.2354368	R	178	1	P
7	1	K	50	1	W	93	0.3540846	G	136	1	E	179	1	F
8	1	K	51	1	L	94	1	F	137	1	D	180	0.0347630	A
9	0.0403266	R	52	1	G	95	0.1622925	T	138	1	E	181	1	R
10	1	R	53	1	N	96	1	D	139	0.1622925	E	182	0.3556579	T
11	0.0752808	D	54	0.5920919	V	97	0.3328812	E	140	1	R	183	1	G
12	0.5914911	G	55	1	K	98	1	E	141	0.6940242	E	184	0.4943107	D
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	0.4943107	R	143	1	R	186	1	L
15	0.3556579	D	58	0.2110788	L	101	1	E	144	0.3447755	V	187	1	D
16	0.5513911	V	59	1	G	102	1	H	145	1	T	188	0.0572071	V
17	1	L	60	1	K	103	1	R	146	1	G	189	1	R
18	1	E	61	1	K	104	1	R	147	1	P	190	1	G
19	0.1164969	Q	62	1	E	105	1	R	148	0.5096951	P	191	0.125	V
20	1	W	63	0.4187479	K	106	1	K	149	0.3289625	T	192	0.3328812	H
21	1	I	64	1	D	107	1	T	150	1	G	193	1	G
22	0.1475001	S	65	1	G	108	1	L	151	1	D	194	1	F
23	0.2068347	G	66	1	E	109	1	E	152	1	V	195	1	P
24	1	R	67	1	G	110	1	N	153	1	N	196	0.4943107	W
25	0.5607143	R	68	1	A	111	1	K	154	1	P	197	1	D
26	0.125	L	69	1	P	112	0.0132698	K	155	1	L	198	0.6886562	I
27	0.4943107	I	70	1	P	113	0.125	V	156	1	G	199	1	L
28	1	E	71	1	T	114	1	Q	157	1	G	200	1	F
29	0.0559845	E	72	1	K	115	1	L	158	0.2354368	D	201	1	P
30	1	L	73	0.7275844	R	116	0.0721130	S	159	0.4187479	R	202	0.2354368	V
31	1	E	74	1	P	117	0.5075132	G	160	1	R	203	1	?
32	1	R	75	1	R	118	1	G	161	1	G	204	1	P
33	0.7017677	E	76	0.0559845	T	119	1	G	162	1	A	205	1	P
34	1	S	77	1	D	120	1	K	163	1	P	206	1	F
35	1	R	78	1	R	121	0.2354368	I	164	1	G	207	1	S
36	1	K	79	1	M	122	1	L	165	1	G	208	1	P
37	0.0405482	E	80	1	E	123	1	S	166	1	G	209	1	Q
38	0.4943107	R	81	0.7291322	V	124	0.5547089	K	167	0.125	F	210	1	S
39	1	K	82	1	D	125	0.2354368	G	168	1	V	211	1	C
40	1	K	83	1	H	126	1	E	169	1	P	212	1	R
41	0.0914093	I	84	1	?	127	1	K	170	0.125	?	213	0.2110788	P
42	1	R	85	1	S	128	1	G	171	0.5978741	L	214	1	Q
43	1	R	86	0.5607143	R	129	1	E	172	1	Q	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-B*49

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	1	L	87	1	K	130	1	L	173	0.0673076	D
2	1	S	45	1	E	88	1	R	131	0.4228027	R	174	1	I
3	1	Q	46	0.4439532	E	89	1	P	132	0.2929908	K	175	1	P
4	0.1903527	A	47	0.0395596	E	90	0.3487647	H	133	1	L	176	1	E
5	1	G	48	0.0113781	H	91	0.2464669	K	134	1	A	177	1	S
6	0.3447934	S	49	1	S	92	1	R	135	0.0673076	D	178	1	P
7	0.3831093	R	50	1	W	93	0.1306945	D	136	0.2992142	D	179	1	Y
8	0.0673076	D	51	1	L	94	1	F	137	1	D	180	0.2929908	S
9	0.0092400	K	52	1	G	95	1	T	138	1	E	181	1	R
10	1	R	53	1	N	96	1	E	139	0.5540427	K	182	0.0673076	?
11	0.0134361	E	54	0.0696639	I	97	0.0673076	R	140	1	R	183	1	G
12	1	D	55	1	K	98	1	E	141	1	E	184	0.0673076	G
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	0.5896782	Q	143	1	R	186	1	L
15	0.5920919	E	58	1	I	101	1	D	144	0.0130164	T	187	1	D
16	0.4389210	T	59	1	G	102	1	H	145	1	T	188	0.2379739	T
17	1	L	60	1	K	103	1	R	146	1	G	189	0.0673076	V
18	1	E	61	1	?	104	1	R	147	1	P	190	1	G
19	0.0673076	?	62	0.1306945	E	105	1	R	148	0.6770233	P	191	0.2992142	G
20	1	W	63	1	R	106	1	K	149	0.4393122	V	192	0.3952580	R
21	1	V	64	1	D	107	1	D	150	1	G	193	1	G
22	0.0199840	S	65	1	G	108	1	L	151	1	D	194	1	F
23	1	A	66	0.1306945	G	109	1	E	152	1	V	195	1	P
24	1	R	67	1	E	110	1	N	153	1	N	196	0.2992142	W
25	0.3487647	R	68	1	A	111	1	K	154	1	P	197	1	D
26	0.2653287	R	69	1	P	112	0.1903527	E	155	1	H	198	1	L
27	0.0673076	?	70	1	P	113	0.3434795	K	156	1	E	199	1	L
28	1	E	71	1	T	114	1	Q	157	1	A	200	1	F
29	0.5920919	D	72	1	K	115	1	L	158	0.0673076	L	201	1	P
30	1	L	73	0.1974818	R	116	0.4367155	S	159	0.3487647	P	202	0.4527778	S
31	1	E	74	1	P	117	0.6268765	S	160	1	R	203	1	D
32	0.5920919	K	75	1	R	118	1	G	161	1	G	204	1	P
33	0.6088319	E	76	0.5920919	T	119	1	G	162	1	A	205	1	P
34	1	S	77	1	E	120	1	K	163	1	P	206	1	L
35	1	R	78	1	H	121	0.3487647	R	164	1	G	207	1	S
36	0.1903527	K	79	1	M	122	1	L	165	1	G	208	1	P
37	0.0015169	V	80	1	D	123	1	S	166	1	G	209	1	Q
38	0.2992142	R	81	0.1919315	I	124	1	K	167	1	?	210	1	S
39	1	R	82	1	D	125	1	G	168	1	A	211	1	C
40	1	K	83	1	P	126	1	E	169	1	P	212	1	R
41	0.2653287	I	84	1	?	127	0.4389210	K	170	0.6735451	S	213	1	P
42	1	R	85	1	S	128	1	E	171	0.1116249	L	214	1	Q
43	0.5920919	K	86	1	R	129	1	E	172	1	L	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-B*50

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	1	I	87	1	K	130	1	L	173	0.0452760	G
2	1	S	45	1	E	88	1	R	131	0.0961538	L	174	0.2027393	V
3	0.2639151	R	46	0.0150131	D	89	1	P	132	0.0451099	K	175	1	P
4	0.2639151	A	47	0.5059451	D	90	0.1016584	H	133	1	L	176	1	E
5	1	E	48	0.0238874	H	91	0.0452760	K	134	1	T	177	1	S
6	0.6847133	S	49	0.1839058	L	92	1	G	135	0.0961538	D	178	1	P
7	0.7221343	K	50	1	W	93	0.1839058	D	136	1	D	179	1	F
8	0.2639151	R	51	1	L	94	1	F	137	1	D	180	0.0961538	?
9	0.3367948	G	52	1	G	95	0.0112213	S	138	1	E	181	1	R
10	1	R	53	1	N	96	1	E	139	0.4634067	E	182	0.1839058	S
11	0.0961538	?	54	0.1348802	V	97	0.6005875	K	140	1	R	183	1	G
12	0.0238874	D	55	1	K	98	1	E	141	0.6828696	K	184	0.0961538	G
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	0.3509269	Q	143	1	R	186	1	L
15	0.1839058	N	58	0.5181611	I	101	1	D	144	0.7221343	V	187	1	D
16	0.0281937	T	59	1	G	102	1	H	145	1	A	188	0.3243272	L
17	1	L	60	1	K	103	1	R	146	1	G	189	0.0961538	V
18	1	E	61	1	K	104	1	R	147	1	P	190	1	G
19	1	R	62	1	E	105	1	K	148	0.5181611	Q	191	0.3117868	N
20	1	W	63	1	K	106	1	K	149	0.3228590	V	192	0.0961538	L
21	1	I	64	1	D	107	1	D	150	1	G	193	1	G
22	0.0415528	S	65	1	G	108	1	L	151	1	D	194	1	F
23	0.1094259	A	66	0.0084017	G	109	1	E	152	1	V	195	1	P
24	1	R	67	1	G	110	1	N	153	1	N	196	1	?
25	0.1016584	R	68	1	A	111	1	K	154	1	P	197	1	D
26	0.6347107	R	69	1	P	112	0.0238874	E	155	1	H	198	0.6211815	L
27	0.3367948	Q	70	1	P	113	0.6096321	R	156	1	G	199	1	L
28	1	E	71	1	A	114	1	Q	157	1	G	200	1	F
29	0.1094259	E	72	1	K	115	1	L	158	0.0961538	L	201	1	P
30	1	L	73	0.1115678	R	116	0.3188872	S	159	0.2810347	S	202	0.7396183	S
31	1	E	74	1	A	117	0.5059451	A	160	1	R	203	0.0961538	E
32	1	R	75	1	R	118	1	G	161	1	G	204	1	P
33	0.0961538	?	76	0.0961538	P	119	1	G	162	1	A	205	1	P
34	1	L	77	1	D	120	1	K	163	1	P	206	1	F
35	1	R	78	0.0961538	P	121	0.0452760	K	164	1	G	207	1	S
36	1	R	79	1	M	122	1	L	165	1	G	208	1	P
37	0.1704423	I	80	1	E	123	1	S	166	1	G	209	1	Q
38	1	R	81	0.4447041	I	124	0.7412131	K	167	1	F	210	1	S
39	1	R	82	1	D	125	1	G	168	1	V	211	1	C
40	1	K	83	1	P	126	1	E	169	1	P	212	1	R
41	0.3505969	I	84	1	A	127	0.5678352	K	170	1	N	213	1	T
42	1	R	85	1	P	128	1	E	171	0.2076979	M	214	1	Q
43	0.2850992	R	86	1	R	129	1	E	172	0.2639151	Q	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-B*51

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	1	I	87	1	K	130	1	L	173	0.0515309	S
2	1	S	45	1	E	88	1	R	131	0.6605051	E	174	0.7757286	V
3	0.5488292	R	46	0.8138922	D	89	0.5711644	P	132	0.2307692	?	175	1	P
4	0.1323639	A	47	0.3527413	E	90	0.1323639	F	133	1	L	176	1	E
5	1	G	48	0.5872457	N	91	0.6605051	G	134	0.5711644	T	177	1	S
6	0.4100074	G	49	0.0515309	T	92	1	R	135	0.2307692	V	178	1	P
7	0.4392858	R	50	1	W	93	0.3481699	E	136	0.3258347	E	179	1	Y
8	0.0792542	K	51	1	L	94	1	F	137	1	D	180	0.0502469	S
9	0.2307692	?	52	1	G	95	1	S	138	1	E	181	1	R
10	1	R	53	1	N	96	0.6197657	D	139	0.0642269	K	182	0.3481699	H
11	0.1323639	R	54	1	V	97	0.2307692	R	140	1	R	183	1	G
12	0.2307692	E	55	1	K	98	1	E	141	0.3546290	K	184	0.3258347	D
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	0.3258347	K	143	1	R	186	1	L
15	0.2897567	E	58	0.6605051	L	101	1	E	144	0.2307692	A	187	1	D
16	0.0591359	V	59	1	G	102	1	H	145	1	T	188	0.7284698	L
17	1	L	60	1	K	103	1	R	146	1	G	189	0.5711644	R
18	1	E	61	0.2307692	K	104	1	R	147	1	P	190	1	G
19	0.2307692	?	62	0.0515309	D	105	1	K	148	0.4100074	R	191	0.0913782	D
20	1	W	63	1	R	106	0.4100074	M	149	0.0484862	P	192	0.0513940	Q
21	1	V	64	1	D	107	0.2307692	D	150	1	G	193	1	G
22	0.2307692	V	65	1	G	108	1	L	151	1	D	194	1	F
23	0.4911907	A	66	1	D	109	1	E	152	1	V	195	1	P
24	1	R	67	1	E	110	1	N	153	1	N	196	0.3258347	W
25	0.0515309	N	68	1	A	111	1	K	154	1	P	197	1	D
26	0.1061890	R	69	1	P	112	0.2307692	?	155	0.2307692	H	198	0.0360103	L
27	0.3258347	I	70	1	P	113	0.0515309	?	156	1	E	199	1	L
28	1	E	71	1	T	114	1	Q	157	1	A	200	1	F
29	0.7275844	D	72	1	K	115	1	L	158	0.0515309	S	201	1	P
30	1	L	73	0.1623972	K	116	0.4855976	S	159	0.0375252	T	202	0.8144306	S
31	1	E	74	0.3320885	A	117	0.2770656	G	160	1	R	203	1	?
32	0.7275844	R	75	1	R	118	1	G	161	1	G	204	1	P
33	0.3705315	E	76	0.0375252	A	119	1	G	162	1	A	205	1	P
34	0.2307692	S	77	1	E	120	1	K	163	1	P	206	1	L
35	1	R	78	0.1931056	R	121	0.2272025	K	164	1	G	207	1	S
36	0.1323639	R	79	1	M	122	1	L	165	1	G	208	1	P
37	0.0515309	M	80	1	D	123	1	S	166	1	G	209	1	Q
38	1	K	81	0.0018713	I	124	0.1021037	R	167	1	?	210	1	S
39	0.5488292	R	82	1	D	125	1	G	168	1	A	211	1	C
40	1	K	83	0.2307692	A	126	1	E	169	1	P	212	1	R
41	0.6934767	V	84	1	?	127	1	K	170	0.6034946	S	213	1	P
42	0.5488292	K	85	1	S	128	1	G	171	0.6808798	L	214	1	Q
43	0.2728318	R	86	0.3320885	R	129	1	E	172	0.5488292	L	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-B*52

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	1	L	87	1	K	130	1	L	173	1	S
2	1	S	45	1	E	88	0.1486183	K	131	0.0769230	L	174	0.1792156	I
3	1	Q	46	0.7149635	E	89	1	T	132	0.6056558	K	175	1	P
4	1	T	47	0.2886244	D	90	0.6759134	L	133	1	L	176	1	E
5	1	G	48	1	?	91	0.2775404	K	134	1	T	177	1	S
6	0.0769230	P	49	0.2153934	S	92	1	R	135	1	K	178	1	P
7	0.4377657	K	50	1	W	93	0.5918608	G	136	1	D	179	1	Y
8	1	D	51	1	L	94	1	F	137	1	D	180	0.5678352	H
9	0.1486183	Y	52	1	G	95	1	T	138	1	E	181	1	R
10	1	R	53	1	N	96	1	D	139	0.0769230	?	182	0.1486183	S
11	0.5280959	E	54	0.0898052	I	97	0.5918608	K	140	1	R	183	1	G
12	0.0052277	N	55	1	K	98	1	E	141	0.3452353	E	184	1	D
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	0.5939083	Q	143	1	R	186	1	L
15	0.5678352	D	58	1	I	101	1	E	144	0.1208478	T	187	1	D
16	0.1486183	L	59	1	G	102	1	H	145	1	T	188	0.5918608	L
17	1	L	60	1	K	103	1	R	146	1	G	189	1	T
18	1	E	61	1	?	104	1	R	147	1	P	190	1	G
19	1	?	62	1	E	105	1	R	148	0.6759134	P	191	0.1945643	D
20	1	W	63	1	R	106	1	K	149	0.0281937	P	192	0.0769230	L
21	1	V	64	1	D	107	1	D	150	1	G	193	1	G
22	0.0769230	D	65	1	G	108	1	L	151	1	D	194	1	F
23	0.2615417	A	66	1	E	109	1	E	152	1	V	195	1	P
24	1	R	67	1	E	110	1	N	153	1	N	196	0.2153934	?
25	1	K	68	1	A	111	1	K	154	1	P	197	1	D
26	0.0876636	R	69	1	P	112	0.2344638	R	155	1	S	198	0.3477077	I
27	0.4389210	A	70	1	P	113	0.1945643	K	156	1	E	199	1	L
28	0.1486183	D	71	1	T	114	1	Q	157	1	A	200	1	F
29	0.2615417	E	72	1	K	115	1	L	158	0.2153934	E	201	1	P
30	1	L	73	0.1931056	R	116	0.3890473	Q	159	0.2775404	R	202	0.7225385	A
31	1	E	74	1	P	117	0.3537654	S	160	1	R	203	0.0052277	D
32	0.5914911	R	75	1	R	118	1	G	161	1	G	204	1	P
33	0.3455567	E	76	0.5914911	T	119	1	G	162	1	A	205	1	P
34	1	S	77	0.0769230	D	120	1	K	163	1	P	206	1	L
35	1	R	78	1	R	121	0.2775404	K	164	1	G	207	1	S
36	1	R	79	1	M	122	1	L	165	1	G	208	1	P
37	0.4851955	I	80	1	D	123	1	S	166	1	G	209	1	Q
38	1	R	81	0.1936581	V	124	1	R	167	1	?	210	1	S
39	1	R	82	1	D	125	1	E	168	1	A	211	1	C
40	1	K	83	0.2153934	P	126	1	E	169	1	P	212	1	R
41	0.5987428	I	84	1	G	127	1	E	170	1	S	213	1	P
42	1	K	85	1	S	128	0.1486183	G	171	0.5280959	M	214	1	Q
43	1	K	86	1	R	129	1	E	172	1	Q	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-B*55

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	1	L	87	1	K	130	1	L	173	1	D
2	1	S	45	1	E	88	1	K	131	0.5093353	R	174	1	I
3	1	Q	46	0.5102688	E	89	1	T	132	0.2852875	K	175	1	P
4	1	A	47	1	S	90	1	F	133	1	L	176	1	E
5	1	G	48	1	H	91	1	A	134	1	A	177	1	S
6	0.2184466	R	49	1	S	92	1	R	135	1	I	178	1	P
7	0.0705750	R	50	1	W	93	1	G	136	1	E	179	1	F
8	0.0571321	R	51	1	L	94	1	F	137	1	D	180	0.4247572	A
9	0.1306945	R	52	1	G	95	1	T	138	1	E	181	1	R
10	1	R	53	1	N	96	1	D	139	1	I	182	1	S
11	0.1663554	E	54	1	I	97	1	E	140	1	R	183	1	G
12	1	D	55	1	K	98	1	E	141	1	E	184	1	G
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	1	K	143	1	R	186	1	L
15	1	E	58	1	I	101	1	D	144	1	A	187	1	D
16	0.0382748	A	59	1	G	102	1	H	145	1	A	188	0.2184466	L
17	1	L	60	1	K	103	1	R	146	1	G	189	0.0382748	T
18	1	E	61	1	K	104	1	R	147	1	P	190	1	G
19	1	?	62	1	E	105	1	K	148	0.0192307	?	191	0.0192307	R
20	1	W	63	0.0758028	K	106	1	K	149	0.1839058	A	192	0.5491038	Q
21	1	V	64	1	D	107	1	T	150	1	G	193	1	G
22	1	S	65	1	G	108	1	L	151	1	G	194	1	F
23	1	A	66	1	E	109	1	E	152	1	V	195	1	P
24	1	R	67	1	G	110	1	N	153	1	N	196	1	?
25	1	N	68	1	A	111	1	K	154	1	P	197	1	D
26	1	L	69	1	P	112	0.5317401	K	155	1	H	198	1	K
27	0.1306945	V	70	1	P	113	1	R	156	1	G	199	1	L
28	1	D	71	1	P	114	1	Q	157	1	G	200	1	F
29	1	E	72	1	K	115	1	L	158	1	E	201	1	P
30	1	L	73	0.3950709	K	116	0.4393203	A	159	1	P	202	1	S
31	1	E	74	1	P	117	0.2018297	A	160	1	R	203	1	E
32	1	R	75	1	R	118	1	G	161	1	G	204	1	P
33	1	?	76	0.0942867	V	119	1	G	162	1	A	205	1	P
34	1	S	77	1	D	120	1	K	163	1	P	206	1	L
35	1	R	78	1	P	121	0.0192307	Y	164	1	G	207	1	S
36	1	R	79	1	M	122	1	L	165	1	G	208	1	P
37	0.0382748	M	80	1	D	123	1	S	166	1	G	209	1	Q
38	1	K	81	1	V	124	0.4951456	R	167	1	F	210	1	S
39	1	R	82	1	D	125	1	E	168	1	A	211	1	C
40	1	K	83	1	P	126	1	E	169	1	P	212	1	R
41	1	I	84	1	G	127	1	E	170	1	?	213	1	T
42	1	R	85	1	P	128	1	G	171	1	L	214	1	Q
43	1	K	86	1	R	129	1	E	172	1	L	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-B*57

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	1	L	87	1	K	130	1	L	173	1	G
2	1	S	45	1	E	88	0.1306945	K	131	0.1003727	R	174	1	I
3	1	Q	46	0.0190227	E	89	1	P	132	0.5918297	K	175	1	P
4	1	A	47	1	E	90	0.1880773	L	133	1	L	176	1	E
5	1	G	48	1	H	91	1	K	134	0.1903527	A	177	1	S
6	0.5876094	R	49	1	S	92	1	R	135	1	V	178	1	P
7	0.6701407	K	50	1	W	93	1	G	136	1	E	179	1	Y
8	1	R	51	1	L	94	1	F	137	1	D	180	0.1348802	H
9	0.0696639	R	52	1	G	95	0.3487647	S	138	1	E	181	1	R
10	1	R	53	1	N	96	0.3487647	E	139	0.3406653	R	182	0.1306945	S
11	0.2464669	D	54	0.3952815	I	97	0.1848093	K	140	1	R	183	1	G
12	1	D	55	1	K	98	1	E	141	1	K	184	1	G
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	0.0354776	R	143	1	R	186	1	L
15	0.5920919	E	58	0.3952815	L	101	1	E	144	0.3208206	T	187	1	D
16	0.1306945	L	59	1	G	102	1	H	145	1	T	188	0.2316543	V
17	1	L	60	1	K	103	1	R	146	1	G	189	1	R
18	1	E	61	1	?	104	1	R	147	1	P	190	1	G
19	0.0673076	E	62	1	D	105	1	R	148	0.2929908	S	191	0.0673076	V
20	1	W	63	1	R	106	0.1306945	M	149	0.3831093	T	192	0.1903527	H
21	1	V	64	1	D	107	1	D	150	1	G	193	1	G
22	0.4367155	N	65	1	G	108	1	L	151	1	D	194	1	F
23	1	A	66	1	D	109	1	E	152	1	V	195	1	P
24	1	R	67	1	E	110	1	N	153	1	N	196	1	?
25	0.3487647	R	68	1	A	111	1	K	154	1	P	197	1	D
26	0.5896782	R	69	1	P	112	0.0673076	Q	155	1	S	198	0.2379739	L
27	0.2464669	Q	70	1	P	113	0.0673076	V	156	1	E	199	1	L
28	1	E	71	1	A	114	1	Q	157	1	A	200	1	F
29	0.0403266	D	72	1	K	115	1	L	158	0.1974818	G	201	1	P
30	1	L	73	0.6486228	K	116	0.2464669	R	159	0.0514617	Q	202	0.0467323	A
31	1	E	74	1	P	117	0.4528145	A	160	1	R	203	0.0673076	?
32	0.5920919	K	75	1	R	118	1	G	161	1	G	204	1	P
33	0.6088319	E	76	0.0673076	?	119	1	G	162	1	A	205	1	P
34	1	L	77	1	E	120	1	K	163	1	P	206	0.0673076	F
35	1	R	78	0.4389210	R	121	0.0673076	F	164	1	G	207	1	S
36	1	R	79	1	M	122	1	L	165	1	G	208	1	P
37	0.0673076	L	80	1	D	123	1	S	166	1	G	209	1	Q
38	1	K	81	1	V	124	1	K	167	0.0673076	F	210	1	S
39	1	R	82	1	D	125	1	E	168	1	A	211	1	C
40	1	K	83	0.1903527	P	126	1	E	169	1	P	212	1	R
41	0.2992142	L	84	1	A	127	0.4389210	K	170	0.6605051	N	213	0.3952815	P
42	1	R	85	1	S	128	0.1306945	G	171	0.4798330	M	214	1	Q
43	0.5920919	K	86	1	R	129	1	E	172	1	L	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

HLA-B*58

Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA	Po	p-value	AA
1	1	M	44	1	L	87	1	K	130	1	L	173	1	N
2	1	S	45	1	E	88	1	R	131	0.5102688	K	174	1	I
3	1	Q	46	1	D	89	1	I	132	1	?	175	1	P
4	1	A	47	0.4998132	D	90	0.0382748	A	133	1	L	176	1	E
5	1	E	48	1	H	91	0.1306945	G	134	0.0571321	A	177	1	S
6	1	P	49	1	S	92	1	G	135	1	K	178	1	P
7	1	E	50	1	W	93	1	E	136	0.0942867	D	179	1	Y
8	1	R	51	1	L	94	1	F	137	1	D	180	0.4247572	A
9	0.4953323	N	52	1	G	95	1	S	138	1	E	181	1	R
10	1	R	53	1	N	96	1	E	139	0.2012696	K	182	0.1306945	H
11	0.1663554	E	54	1	I	97	1	E	140	1	R	183	1	G
12	1	D	55	1	K	98	1	E	141	0.3175877	E	184	1	D
13	1	R	56	1	G	99	1	R	142	1	R	185	1	G
14	1	E	57	1	I	100	1	Q	143	1	R	186	1	L
15	1	E	58	1	L	101	1	D	144	1	A	187	1	D
16	0.1486183	T	59	1	G	102	1	H	145	1	T	188	1	I
17	1	L	60	1	K	103	1	R	146	1	G	189	0.0192307	G
18	1	E	61	1	?	104	1	R	147	1	P	190	1	G
19	1	?	62	1	E	105	1	R	148	1	S	191	0.3334578	S
20	1	W	63	1	K	106	1	K	149	0.4958924	V	192	0.0571321	H
21	1	V	64	1	D	107	1	A	150	1	G	193	1	G
22	0.4951456	N	65	1	G	108	1	L	151	1	D	194	1	F
23	0.2354368	A	66	1	D	109	1	E	152	1	V	195	1	P
24	1	R	67	1	E	110	1	N	153	1	N	196	0.0942867	W
25	1	R	68	1	A	111	1	K	154	1	P	197	1	D
26	1	R	69	1	P	112	0.5093353	R	155	1	H	198	1	?
27	1	M	70	1	P	113	0.1306945	N	156	1	E	199	1	L
28	1	D	71	1	A	114	1	Q	157	1	A	200	1	F
29	1	D	72	1	K	115	1	L	158	0.2522404	A	201	1	P
30	1	L	73	1	K	116	0.0758028	N	159	0.0758028	T	202	1	S
31	1	E	74	0.0758028	P	117	0.4998132	A	160	1	R	203	1	?
32	0.2354368	K	75	1	R	118	1	G	161	1	G	204	1	P
33	1	?	76	1	T	119	1	G	162	1	A	205	1	P
34	1	L	77	1	E	120	1	K	163	1	P	206	1	F
35	1	R	78	1	H	121	0.1456310	S	164	1	G	207	1	S
36	1	R	79	1	M	122	1	L	165	1	G	208	1	P
37	0.5102688	V	80	1	E	123	1	S	166	1	G	209	1	Q
38	1	R	81	0.4247572	I	124	1	R	167	1	?	210	1	S
39	1	R	82	1	D	125	1	E	168	1	A	211	1	C
40	1	K	83	1	P	126	1	E	169	1	P	212	1	R
41	1	L	84	1	?	127	1	E	170	0.4100074	N	213	1	T
42	1	R	85	1	S	128	1	E	171	1	M	214	1	Q
43	1	K	86	1	G	129	1	E	172	1	L	215	1	*

Po: position of amino acid on L-HDAg; AA: amino acid with the lowest p-value

Abbreviations

aa	amino acid
Ab	antibodies
ADAR	adenosine deaminase acting on RNA
AdV	adenovirus
Ag	antigen
ALT	alanine aminotransferase
APC	antigen presenting cell
APC (Dye)	Allophycocyanin
AST	aspartate aminotransferase
BFA	Brefeldin A
bp	base pairs
CD	cluster of differentiation
CHB	chronic hepatitis B
CHC	chronic hepatitis C
CHD	chronic hepatitis D
CMV	Cytomegalovirus
CPEB3	cytoplasmic polyadenylation element binding protein 3
CTL	cytotoxic T lymphocyte
CTLA-4	cytotoxic T-lymphocyte-associated Protein 4
DC	dendritic cell
ddH2O	double-distilled water
DHBV	duck hepatitis B virus
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxy ribonucleotide triphosphate
e.g.	for example
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
ETV	Entecavir
FACS	fluorescence activated cell sorting
FCS	fetal calf serum
FITC	Fluorescein isothiocyanate
HAV	hepatitis A virus
HBs Ag	hepatitis B surface antigen
HBV	hepatitis B virus
HCC	hepatocellular carcinoma
HCV	hepatitis C virus
HDV	hepatitis delta virus
HIV	human immune deficiency virus
HLA	human leukocyte antigen
ICS	intracellular cytokine staining

IEDB	immune epitope database
IFN	interferon
IgG	immunglobulin G
IgM	immunglobulin M
IL	interleukin
JAK-STAT	Janus kinase-signal transducer and activator of transcription
kb	kilo base
L-HD Ag	large hepatitis delta antigen
MHC	major histocompatibility complex
Min.	minute
µg	microgram
µl	microliter
ml	milliliter
ML	maximum likelihood
M-MLV RT	Moloney Murine Leukemia Virus Reverse Transcriptase
mRNA	messenger ribonucleic acid
MSA	multiple sequence alignment
NAs	nucleot(s)ide analogs
NK	natural killer
nt	nucleotide
ORF	open reading frame
PBMCs	peripheral blood mononuclear cells
PBS	phosphate buffered saline
PCR-SSO	polymerase chain reaction-sequence specific oligonucleotide
PD-1	programmed cell death protein-1
PD-L1	programmed cell death protein ligand-1
Peg-IFN	pegylated interferon
PRRs	pattern recognition receptors
RdRp	RNA dependent RNA polymerase
RNP	ribonucleoprotein
Sec.	second
siRNA	small interfering RNA
SNP	single nucleotide polymorphisms
SVP	subviral particle
TCR	T cell receptor
UDPS	ultra-deep pyrosequencing
UV	ultraviolet
WHsAg	woodchuck hepatitis surface antigen
WHV	woodchuck hepatitis virus
WT	wild type

List of figures

List of tables

Acknowledgement

First, I would like to express my sincere thanks to my advisor, Prof. Dr. Ulrike Protzer, for giving me the opportunity to join her lab and her wonderful group. I very much appreciate her support during these years.

Foremost, a special thanks goes to my mentor/advisor, Prof. Dr. Michael Roggendorf, for trusting me in the first place, and for introducing me to this project. His continuous support and guidance helped me immensely during my research and the writing of this dissertation.

I would also like to acknowledge my co-advisor, Prof. Dr. rer. nat. Dmitrij Frischmann, for inspiring discussions and critical questions which helped to strengthen this dissertation. I also sincerely thank his group members for technical and bioinformatical advice.

In particular, I am grateful to PD. Dr. Christoph Neumann-Haefelin and his PhD student, Michael Kiraithe, not only for fruitful discussions, but also for critical input, including valuable samples and data from patients with resolved HDV infection.

I have to thank Dr. Anna Kosinska for critical discussions during my project and Dr. Melanie Fiedler for providing the first patient data supporting my hypothesis of HDV-specific HLA-B*27-restricted CD8 T cell response.

I am thankful to Dr. Tanja Bauer and the whole immunmonitoring group for sophisticated assistance. I am also grateful to all the faculty members of the institute of virology TU München for the kind atmosphere.

I sincerely thank Prof. Dr. Jörg Timm for enlightening discussions and specific help in analysis and interpretation of the data.

I would also like to thank all the anonymous patients and the healthy volunteers who trusted this project and us by generously donating their valuable blood samples for our study. Of course, I thank all of the nurses in the blood drawing facilities of our collaborating centers.

Last but not the least, I would like to thank my family: my parents and my brothers (especially Majid) for supporting me spiritually throughout the writing of this thesis and my life in general.