Chapter 1

General Introduction
The Human Leukocyte System

The Human Leukocyte Antigen (HLA) system, the human equivalent of the Major Histocompatibility Complex (MHC), possesses the highest degree of polymorphism present in the human genome. The HLA complex is located on the short arm of chromosome six, 6p21.3, and is divided into three main regions: the class I, II and III regions and in addition the extended class I and class II regions. This entire complex encompasses 7.6 Mb, with 421 identified genes to date: 252 are expressed genes, 139 pseudogenes and 30 transcripts without a characterized open reading frame [1].

In the class I region, the classical genes HLA-A, HLA-B and HLA-C, the non-classical genes HLA-E, HLA-F and HLA-G and the pseudogenes HLA-H, HLA-J, HLA-K and HLA-L are located. Within the class II region the classical class II genes HLA-DR, HLA-DQ and HLA-DP, their pseudogenes and the class II-like genes HLA-DO and HLA-DM are present. Genes involved in the class I antigen processing and presentation pathway, the proteasome subunit β (PSMB) and the transporter associated with antigen processing (TAP) genes are also located in this region [2]. The class III region, interspersed between the classes I and II, is the most dense gene region in the human genome, with an average of 8.5 genes per 100 kb. Although many class III genes have unknown function, several of the genes are involved in the innate immune system (i.e. the complement genes C2, C4 and factor B), the immune/inflammatory response (i.e. the cytokines TNF, LTA, LTB and the lymphocyte antigen 6 (Ly6) genes and stress-induced signalling genes (heat shock proteins) [3].

During evolution the HLA region has diversified from ancestral genes. It is presumed that the HLA region has emerged in two rounds of large-scale genomic duplication from a common ancestral region resulting in four paralogous regions [4]. The regions paralogous to the HLA region on the short arm of chromosome 6 are located on chromosomes 1, 9 and 19. Each of these regions has undergone positive and negative genetic selection, explaining why not all genes present in the HLA region are present in one of the other paralogous regions. The region on chromosome 1 is the region most closely related to the HLA region on chromosome 6, with numerous genes similar as in the HLA region. This region contains for example a cluster of genes, the CD1 genes that resemble the HLA genes in structure and function. The CD1 molecules are cell surface glycoproteins and are able to bind antigens from (glyco)lipids and to present them to γδ T-cells, NKT-cells and natural killer (NK) cells [5]. Associations of genes with particular diseases have been described for chromosome 6, but also for genes in the paralogous region on chromosome 1 [5].
Classical HLA genes

The classical class I molecules, HLA-A, HLA-B and HLA-C, are cell-surface glycoproteins consisting of a polymorphic heavy or α chain (~45 kDa) and a monomorphic light or α chain (~12 kDa). Constitutive expression for these classical HLA molecules is observed on most nucleated cells, although the level of expression can differ between tissues and cell types. Higher expression levels are in general detected for HLA-A and HLA-B compared to HLA-C. An increased turnover of HLA-C mRNA and restricted peptide binding has been suggested to reduce HLA-C cell surface expression. The function of the classical HLA class I molecules is binding and presentation of endogenously derived peptides to CD8+ cytotoxic T-cells, eliciting a cellular immune response if a non-self peptide is presented. Stable expression of the class I molecules requires a complex of the heavy and light chain and in addition a bound peptide [6-10].

Genes encoding for the polymorphic heavy chain are located in the HLA class I region and the gene for the monomorphic light chain, beta2-microglobulin (β2m), on chromosome 15. Extracellular and intracellular domains in the HLA class I α chain are defined by different exons. The signal sequence encoded by exon 1, is necessary for the correct transport of the HLA class I heavy chain into the endoplasmatic reticulum (ER) after which the signal sequence is cleaved off. The extracellular domains, α1, α2 and α3, are encoded by respectively exons 2, 3 and 4. The α1 and α2 domains form the peptide binding groove of the HLA class I molecule and define the polymorphism in the peptide-binding groove. The peptide-binding groove presents peptides to CD8+ T-cells and binds the T-cell receptor. The α2 domain and the α3 domain, an immunoglobulin-like domain, interact with the CD8 co-receptor on CD8+ T-cells. The hydrophobic transmembrane domain is encoded by exon 5, the hydrophilic cytoplasmic tail is encoded by exons 6 and 7 and the termination codon is present in exon 8 for HLA-A and HLA-C. In HLA-B, exons 6 and 7 encode the cytoplasmic tail and the termination codon [7, 11-13].

The classical class II molecules HLA-DR, HLA-DQ and HLA-DP are cell-surface molecules composed of a polymorphic α chain (~33-35 kDa) and a polymorphic β chain (~26-28 kDa). A restricted expression pattern is observed for the class II molecules. Constitutive expression is observed on B cells, whereas expression can be induced by IFN-γ on antigen presenting cells, dendritic cells, macrophages, monocytes, Langerhans cells and vascular endothelial cells. Upon activation T-cells are also able to express class II molecules. Stable class II complexes bind exogenously derived peptides and present those to CD4+ T-cells. If non-self peptides are bound, presentation may elicit a humoral immune response [7, 14].

The α and β chain are respectively encoded by an A and a B gene in the class II region. HLA-DR can have more than one polymorphic β chain and more than one B gene, depending on the haplotype. The α chain is encoded by 4 exons
and the β chain by 5 exons. In both chains exon 1 encodes for the signal sequence and exons 2 and 3 for the extracellular domains α1, α2 and β1, β2. In the α chain the hydrophobic transmembrane domain and the hydrophilic intracellular domain are encoded by exon 4, while in the β chain the transmembrane domain is encoded by exon 4 and the intracellular domain by exon 5. The peptide-binding groove of the class II molecule, formed by the α1 and β1 domain of both chains, is able to present peptides to CD4+ T-cells and interacts with the T-cell receptor. The α2, β1 and β2 domains are able to interact with the CD4 co-receptor on CD4+ T-cells [7, 11, 12, 15-17].

The organization of the peptide-binding groove is overall the same for HLA class I and class II molecules. The groove is composed of 6 pockets in the class I molecule and 5 pockets in the class II molecule. The impact of each pocket on peptide binding can differ, depending on the polymorphism present. Amino acids in respectively the α1/α2 domains and the α1/β1 domains are involved in the formation of these pockets [7]. The peptides to be presented differ in length as class I peptides are shorter (8-10 amino acids) and are attached with their ends in the peptide-binding groove compared to class II peptides that are longer (12-24 amino acids) and attached with amino acids in the middle of the peptide to the peptide-binding groove, so that the ends of a class II peptide are located outside the peptide-binding groove [18].

**HLA polymorphism**

HLA polymorphism has been established during million of years evolution in which exposure to the (pathogenic) environment has shaped the HLA region as it is nowadays, with the capacity to distinguish between self and non-self and to improve the survival of species. The extent to which HLA class I and II genes have diversified differs extremely between species. For example, the Syrian hamster is monomorphic for his HLA class I genes, but polymorphic for his class II genes whereas the cheetah is monomorphic for all HLA genes. In human and mice, on the contrary, the class I and II polymorphisms is considerable [19].

The majority of the HLA polymorphism is located within the peptide-binding groove, giving rise to the ability to present a divers amount of peptides. However, polymorphism can also be observed in other exons (not involved in the peptide-binding groove), introns and untranslated regions (UTRs) and can result in non-coding polymorphism, synonymous or non-synonymous substitutions, stop codons, alternative splicing and aberrant expression [20-33]. Several mechanisms account for the origin of HLA polymorphism: point mutations, gene-conversion and recombination [18]. For example, the HLA-A69 allele evolved from a recombination between HLA-A68 and one of the HLA-A2 alleles [34], and the HLA-B42 allele evolved from HLA-B7 and HLA-B8 alleles [35]. An unique interlocus
recombination is observed for the HLA-B*4601 allele, that for the majority exists of the HLA-B*1501 allele, while the amino acids at positions 66 though 76 are derived from the HLA-Cw*0102 allele [36]. These mechanisms led to a numerous number of HLA class I and class II alleles for which names have been assigned by the World Health Organization (WHO) HLA Nomenclature Committee. The IMGT/HLA database (version 2.11) (http://www.ebi.ac.uk/imgt/hla/) harbours 1403 class I alleles, 781 class II alleles and 96 alleles of HLA related and non-HLA related genes. The major contributors for the class I alleles is HLA-B and HLA-DRB1 for class II [12].

The different loci with their extensive polymorphism result in high numbers of HLA haplotypes, defined as "the combination of HLA alleles from different HLA loci on a single chromosome in an individual" [1]. As a result of linkage disequilibrium, defined as "the non-random association of HLA alleles at adjacent HLA loci", linkage between alleles of different loci exists. The frequencies of these HLA haplotypes are higher or lower than was expected based upon the frequencies of the individual alleles of the loci. Linkage disequilibrium exists among HLA class I alleles, among class II alleles and between class I and II alleles. Based upon segregation analysis in family studies and independent non-family studies, linkage disequilibria have been determined [37-45].

**HLA nomenclature**

The HLA Nomenclature Committee, under the auspices of the WHO, was founded to formulate guidelines for the official assignment of HLA antigens and alleles [46]. Serological methods initially detected a few HLA specificities for class I, the broad specificities and public epitopes, and these specificities received names based upon the gene locus and followed by a number, e.g. HLA-A2 or HLA-A9. Later on, splits or private epitopes for several broad specificities could be identified, e.g. the HLA-A23 and HLA-A24 splits belong to the broad specificity HLA-A9. The introduction of cellular methods added additional information in the identification of HLA antigens. This technique not only detected more subtypes within a single class I antigen, that generally could not be detected by serology, but revealed also the existence of another group of HLA antigens, the class II antigens. This resulted in complexities in the correct definition of HLA antigens. Monoclonal antibodies, biochemical and molecular approaches were subsequently developed in order to discriminate HLA antigens intra- and interlocus [11, 18].

Guidelines for the assignment of HLA antigens and alleles, as formulated by the WHO, resulted in the following nomenclature: the name of the locus is followed by an asterisk, indicating that the HLA allele was determined at molecular level. The 1st and 2nd digits identify the allele family, in general corresponding with the serological group to which the allele belongs (HLA-A*02 alleles belong to
the serological antigen HLA-A2). The 3\textsuperscript{rd} and 4\textsuperscript{th} digits indicate non-synonymous substitutions in exons, changing thus the amino acid composition of the antigen. The 5\textsuperscript{th} and 6\textsuperscript{th} digits give a synonymous substitution in an exon and sequence polymorphism within introns, the 5\textsuperscript{´}-untranslated (UTR) or 3\textsuperscript{´}-UTR is given by the 7\textsuperscript{th} and 8\textsuperscript{th} digits [12]. In addition, 5 extra suffixes can be added for HLA alleles designating their expression status: “N”, “L”, “S”, “C” and “A” [47-49]. An “N”-suffix indicates an HLA null allele that according to the HLA Nomenclature Committee is defined as “a sequence defect in an HLA allele or gene which prevents normal expression of the product at the cell surface (it does not necessarily imply that no internal partial product is made which might be a T-cell target)” [48]. The suffix “L” indicates that the allele “is associated with unusual expression levels” [47]. The suffix “S” denotes “an allele specifying a protein that is expressed as a soluble secreted molecule but is not present on the cell surface”. No alleles have been described to date for the suffixes “C”, “an allele product present in the cytoplasm but not at the cell surface” and “A”, “aberrant expression if there is some doubt whether the protein is expressed” [49].

**HLA typing**

Since the discovery of the HLA system HLA typing methods have been developed to discriminate alleles intra- en interlocus and these methods can be used in a variety of disciplines, i.e. in solid organ and haematological transplantations, disease-association studies, anthropology, epidemiology and forensics. Several factors should be taken into consideration when performing HLA typing regarding the material, the required resolution and the choice of typing system. In family transplantation low-resolution typing by serology may provide sufficient information to select an appropriate family donor, while in unrelated donor settings, the donor should be typed by high-resolution. Three levels of resolution can be discriminated: low-resolution, obtained with serological methods or sequence-specific priming (SSP) mimicking serological equivalents, and intermediate and high-resolution methods with molecular-based approaches.

The complement-dependent cytotoxicity test (CDC) is used for obtaining a low-resolution class I HLA typing. Polyclonal antisera, with known HLA specificities, are mixed with viable cells from the sample with unknown HLA typing. Recognition by the polyclonal sera, reflected by the lysis of the cells to be tested, indicates that the same HLA typing is present in the cells to be tested. For class II the mixed lymphocyte reaction (MLR) was available in which stimulator cells with known HLA typing were mixed with responder cells with unknown HLA typing. Proliferation of the responder cells indicates that the HLA typing of the responder cells is not the same as the stimulator cells. Although serological typing is able to provide information regarding the presence of particular HLA antigens, shortcomings are
the need for viable cells and the limited ability to provide a detailed HLA typing. Polymorphism present deep within the peptide-binding groove is in general not accessible for the polyclonal antisera [11].

The introduction of molecular-based approaches paved the way for a straightforward identification of the class I and II allelic variants, defining the alleles that belong to the same serological defined group. Three techniques will be discussed: sequence-specific oligonucleotide probing (SSOP), sequence-specific priming (SSP) and sequencing based typing (SBT). All three methods use a polymerase chain reaction (PCR) as the initial basis to selectively amplify the entire or partial HLA locus of interest. Subsequently, additional probing or sequencing can be performed to identify the allele or alleles individually.

The PCR template generated with the SSOP method is fixed to a membrane and labelled oligonucleotide probes, specific for different polymorphic sequences, are hybridized to the membrane. The opposite is also possible, called the reverse dot blot approach and this approach is more widely used than the conventional dot blot approach since one hybridization step is required for typing of one sample. The oligonucleotide probes are fixed to the membrane and a labelled PCR product is subsequently hybridized to the membrane. The obtained hybridization pattern will result in an HLA typing [11]. Similar to the concept of SSOP, an oligonucleotide micro-array has recently been introduced for HLA-DRB1 genotyping [50].

The SSP method is based on specific PCRs only. One or more sequence specific PCRs are developed for each allele or allele-group that generates a pattern of positive and negative PCR product fragments with defined sizes. Depending on the HLA locus investigated, the number of PCRs to be performed varies. The combination of the results of the specifically amplified fragments results in an HLA typing. Probes and primers used in the SSOP and SSP approach are in general located at polymorphic positions to discriminate between the different HLA alleles or allele groups within a locus [11].

The SSP and SSOP method are useful for the identification of serological equivalents. For the identification of allelic variants belonging to the serological equivalents, the number of primers/probes required for the identification is much higher compared to the number required for identification of the serological equivalents. For new alleles, primers or probes may have to be added, changed or relocated and in addition implemented into the SSP and SSOP method. New polymorphisms present within existing PCR products will likely not be determined with the SSP and SSOP method. The resolution obtained with these methods is in most cases an intermediate resolution (two-digit level). To obtain a higher typing resolution a higher number of hybridization or SSP reactions are required.

In 1977, Sanger et al. initiated a sequencing method based upon dideoxy chain termination to determine nucleotide sequences of bacteriophage DNA [51].
Throughout the years refinements have been realized, e.g. automated sequencing, resulting in the SBT method that is now considered to be the golden standard for allele identification. In the context of HLA typing, the SBT method was initially developed to confirm HLA typings obtained with SSP/SSOP and solve the difficult samples. Direct sequencing approaches were developed subsequently to directly identify the alleles. The International Histocompatibility Working Group (IHWG; http://www.ihwg.org/) has initiated protocol development for SBT of the HLA class I and II loci during International Workshops, resulting in a Technical Manual with standardized strategies for HLA typing. These protocols enable to investigate coding and non-coding sequences to obtain insight into the polymorphic character of HLA. Software for sequence data analysis and allele assignment were developed (SBTengine®; Genome Diagnostics, Utrecht, The Netherlands) [52].

As for SSOP and SSP, the SBT method starts with the amplification of HLA alleles of interest. The PCR amplification primers are located in introns where possible. In the next step, sequencing reactions determine the exact sequence in the alleles and computerized-based analysis assigns the correct HLA alleles in a sample. Knowledge of polymorphism of different exons may be required to identify the alleles. In heterozygous sequencing separation of alleles is sometimes required if ambiguous allele combinations are obtained. New polymorphic positions, if present in the amplified product, are easily recognised with the SBT method.

**Influence of HLA polymorphism**

HLA polymorphism is present in all coding and non-coding regions of an HLA gene and is determined by different techniques. The density of polymorphic positions varies between the different regions, with the peptide-binding groove by far the most polymorphic region in the HLA molecule. Polymorphism in the HLA molecule influences the peptide binding, the interaction of HLA molecules with T-cells/NK-cells, the expression level of HLA molecules and the transport of the HLA molecules to the cell surface. Polymorphism in the peptide-binding groove, composed of 5 or 6 pockets, enables the HLA molecules to present a wide range of different peptides to T-cells and B-cells. The chemical composition of the pockets of an HLA molecule has a direct impact on the structural composition and charge of the peptide binding-groove. Thus, polymorphism affects peptide-binding, T-cell binding and T-cell reactivity.

Depending on the amino acids present in the peptide-binding groove, different outcomes are observed: i) differences in amino acid with minimal effect on peptide binding, indicating that the same peptides can be bound [53], ii) binding of different peptides [53-57], that results in a distinct T-cell recognition and reactivity pattern [54, 56], iii) influencing the strength of the peptide-binding [58], iv) although differences in amino acids in HLA are present, the peptide
binding is the same, but interacts in a different way with the HLA molecules and induces in this way different T-cell repertoires [59, 60] or v) based upon the amino acid present in the peptide, the peptide is loaded onto the HLA molecule in a TAP-dependent or TAP-independent pathway [61, 62].

NK-cells interact with HLA molecules via killer-cell immunoglobulin-like receptors (KIR) present on the NK-cells. These KIRs either inhibit or activate lysis of cells, depending whether the KIR ligand is present on HLA molecules that are expressed on the cell surface. HLA-C molecules, one of the major ligands, can be divided into two groups based upon polymorphism present at positions 77-83 in the α2 domain for interaction with particular KIRs, while the Bw4 epitope (positions 77-83) is required for the interaction between KIRs and HLA-B molecules. In addition to these positions, other positions outside the Bw4 epitope also contribute to the HLA-B/NK cell interaction [63-65].

HLA cell surface expression can be affected by polymorphisms in the genomic sequence of an HLA allele. At present 68 null alleles, 1 soluble allele and 4 low expressed alleles are identified (IMGT/HLA database, version 2.11) [12]. Loss of expression is caused by nucleotide substitutions leading to an immediate premature stop codon [66-68], insertions or deletions causing a frame shift with consecutive stop codon [69-73] and substitutions or deletions affecting the splicing machinery [74-76]. The low expressed alleles "HLA-A*0101null" [77] and HLA-A*02010102L [78] are both caused by polymorphism in the enhancer B element in the promoter region, and in the HLA-A*24020102L allele, an intronic substitution in the splice acceptor site caused low expression levels [70, 79]. A non-synonymous substitution in codon 164 resulted in improper folding and loss of expression for the HLA-A*3014L (GenBank AY323229) and the HLA-B*39010102L has a two-nucleotide deletion in the 5'-UTR causing the low expression level (GenBank AB091216). The transmembrane region, encoded by exon 5, is not included in the HLA-B*44020102S due to a single nucleotide difference in intron 4 at the acceptor splice site leading to a soluble molecule [80].

**Functionality of HLA polymorphism**

Immune surveillance of HLA molecules by T- and NK-cells ensures the elimination of invading pathogens during a viral infection. In autoimmune diseases on the other hand, the immune system attacks healthy cells of a person, thereby initiating an immune response with subsequent damage. Tumour cells developed mechanisms to escape immune surveillance by T- and NK-cells, e.g. through down-regulation of HLA expression. From a tumour’s point of view, it would be advantageous to downregulate expression of those HLA molecules that could present tumour-associated antigens (TAA). In leukaemic blasts a downregulated expression was observed for HLA-A molecules and HLA-B alleles with a Bw6 epitope
while expression of HLA-B alleles with a Bw4 epitope was not downregulated, indicating that in this way leukaemic blasts could both escape from T and NK-cell immune surveillance [81] and in this way promote tumour growth. In solid organ and haematological transplantations, matching for HLA polymorphisms between donor and recipient has shown essential. Mismatching may influence peptide presentation, initiating an immune response, leading to graft rejection and graft-versus-host disease (GvHD), and resulting in lower survival rates and increased mortality, thus influencing the outcome of transplantation.

**HLA and transplantation**

Initially only serological methods were available for identification of HLA antigens in patient and donor. Typing for HLA at high-resolution level with molecular-based approaches provides more insight in HLA polymorphism compared to typing at low-resolution level with serological methods. Studies have shown that mismatching at antigen level has worse transplant outcomes when compared to mismatching at allelic level [82-84]. Finding a matched donor is complicated by extensive HLA polymorphism, linkage disequilibrium and constraints on time limits [82]. Identification of permissible and immunogenic mismatches was performed in kidney transplantations. The presence of permissible mismatches resulted in kidney graft survival rates equally or better than in matched donor-recipient couples. Significant higher graft survival rates were observed compared to the immunogenic mismatches, for which graft survival was poor [85].

Investigation of the influence of HLA class I and II allele matching on graft rejection, GvHD, relapse, survival and mortality rates in stem cell transplantation have not resulted in clear conclusions regarding the exact role of the different HLA loci [86-91]. Some single class I or single class II mismatches in stem cell transplantation give clinical outcomes almost as good as HLA matched recipient/donor couples [92]. Mismatches at position 116 in the class I molecules however, involved in peptide binding, result in an increased risk for developing severe acute GvHD and transplant related mortality compared to other polymorphisms involved in peptide binding [93].

In HLA-DPB1 matched kidney cadaver retransplants, survival is improved compared to mismatched retransplants [94]. Epitope matching for HLA-DPB1 in cadaver kidney transplantation, based upon hypervariable regions A through F in the β1 domain of the DP molecule, indicated that specific DPB1 hypervariable regions were associated with improved graft survival. Matching for specific hypervariable regions seemed to be functionally more relevant than allelic matching [95].

The role of HLA-DP in stem cell transplantation is not yet understood and provide contradictory results [86, 89, 91, 96, 97]. Evaluation of matching for HLA-DPB1 in stem cell transplantations showed that HLA-DPB1 matching had either no
effect on the outcome [86, 89], or improved survival [91], or increased relapse rates [96], or was associated with GvHD [91, 96, 97]. If specific amino acids of HLA-DPB1 are considered, mismatching of amino acid position 69, involved in peptide binding, was associated with increased helper T lymphocytes precursor frequencies in stem cell transplantations [98]. In patients with a single HLA mismatch at the DP locus, T-cells restricted for the mismatched DPB1 allele were observed in the patients [99-101]. Functional studies with DPB1 restricted T-cells predicted three groups of HLA-DPB1 alleles with different immunogenicity for the HLA-DP restricted T-cells. From these functional studies an algorithm was developed to predict permissive and non-permissive HLA-DPB1 allele combinations [102]. Non-permissive HLA-DPB1 mismatches increased acute grade II to III GvHD frequency and transplantation related mortality in patients matched for HLA-A, -B, -Cw, -DRB1 and -DQB1 at allele level.

**HLA and autoimmune disease**

Several hypotheses have been postulated that may explain HLA associations in autoimmune diseases. Presentation of pathogenic peptides or biochemically modified self-peptides, selection of T-cells in the thymus that are specific for HLA molecules presenting self-peptides, modification of the peptide binding groove, altering the peptide-binding specificity and linkage of an HLA gene/allele to a predisposing non-HLA disease gene are some examples [103]. Associations with HLA are either identified as susceptible or protective. The majority of autoimmune diseases is associated with HLA class II genes, and only a few associations have been reported with HLA class I genes.

The first recognized association between HLA and an autoimmune disease, and one of the strongest associations, was the observation that almost all patients with ankylosing spondylitis have HLA-B27 while the B27 only in a few percent of healthy individuals was detected. Other associations with class I genes are HLA-B51 in Behçet's syndrome, HLA-Cw6 with psoriasis vulgaris and HLA-A*2902 with birdshot retinitis, which shows an association as strong as observed for B27 and ankylosing spondylitis [104-108].

For associations with class II loci, it is often difficult to determine whether the DR or the DQ locus is primarily involved due to the extensive linkage disequilibrium between both loci. This is the case in insulin-dependent diabetes mellitus (IDDM, type 1 diabetes) and rheumatoid arthritis (RA). Depending on the individual DR or DQ alleles or the combination of DR/DQ alleles in a patient present, either susceptible or protective DR-DQ associations can be defined. It is thought that the DR/DQ association in RA and IDDM depends on respectively shared epitopes in RA-associated DRB1 alleles and on individual amino acids in IDDM-associated DR and DQ alleles [103, 109-112]. In sarcoidosis a DRB1-DQB1
haplotype association is observed but it was not clear whether the DRB1*150101 allele or the DQB1*0602 allele is primarily involved in the disease [113].

Celiac disease is one of the few autoimmune diseases for which the cause has been explained. Patients diagnosed with this disease are intolerant to gluten and related products. The disease is associated with two DQ molecules that contribute both directly in the pathogenesis of celiac disease: the DQ2 molecule (DQA1*0501,DQB1*0201) and the DQ8 molecule (DQA1*0301,DQB1*0302) [114]. CD4+ restricted T-cells for the DQ2 and DQ8 molecules have been isolated from the intestine from celiac patients, recognizing gluten peptides [115, 116]. Elucidation of the peptide-binding characteristics of the DQ2 molecule indicated that the structural requirements of the DQ2 molecule did not fit entirely possible candidate gluten-derived peptides [117]. However, the breakthrough came with the finding that peptides derived from gluten were in general first deamidated by an enzyme called tissue transglutaminase, before they were presented by the DQ2 and DQ8 molecules and subsequently eliciting an immune response [118].

With the availability of the entire sequences of the human genome as performed by the Human Genome Project, information has become available on genes other than the HLA genes in the HLA region. Refinement of the HLA associations observed in (autoimmune) diseases may result in the identification of other genes in the HLA region that are primarily involved in the disease pathogenesis.

**HLA related and non-HLA related diversity**

The entire HLA region, including the extended regions, has been sequenced by the Human Genome Project, and has revealed an enormous amount of gene information. Six gene clusters and six gene superclusters can be discriminated in the HLA and its extended regions [1]. A cluster is defined as "a cluster of genes with three or more paralogous genes or pseudogenes present within 1 Mb". The six clusters are: solute carrier cluster, vomeronasal-receptor cluster, tumour necrosis factor cluster, lymphocyte antigen cluster, heat shock cluster and HLA class II cluster. A supercluster is defined as "clusters with additional related gene(s) outside the core cluster, but within the HLA and its extended region". The six superclusters are: histone supercluster, HLA class I supercluster, tRNA supercluster, butyrophilin supercluster, olfactory-receptor supercluster and the zinc-finger supercluster [1]. The genes in the HLA and extended regions have several biological functions, with an emphasis on immunological functions. Those genes with immunological functions can be categorized into groups: i) antigen processing and presentation; ii) immunoglobulin superfamily; iii) inflammation; iv) leukocyte maturation; v) complement cascade; vi) non-classical MHC molecules; vii) immune regulation and viii) stress response [1].
Chapter 1

An example of HLA related genes are the MIC genes, located in the class I region, structurally related to the class I molecules, but not able to present peptides or interact with β2m [119, 120]. Two functional genes, MICA and MICB, and five pseudogenes, MICC, MICD, MICE and MICF, have been identified [121]. Particularly MICA displays a high degree of polymorphism, with currently 60 identified MICA alleles [12]. MICA and MICB are stress-inducible molecules are can be recognized by NK-cells and T-cells, expressing the NKG2D receptor [122-124]. Different cell types constitutively express MICA and expression can be induced in cells upon stress, for example after viral infection or tumour transformation [123, 125-128]. The precise function of MICA in the immune response cascade is unknown.

In addition to the HLA, HLA related and non-HLA related genes, repeated sequences are located throughout the HLA region, named microsatellites or short tandem repeats (STR) that can be used as markers. These microsatellite markers are located primarily in non-coding regions, consisting of repeated lengths of 2 to 6 basepairs and are in linkage disequilibrium with the classical HLA loci [129, 130]. Since the majority of microsatellite markers are located in non-coding regions, they do not have a direct correlation with gene expression. Microsatellite markers are used in the context of different purposes: as follow-up for monitoring mixed chimerism status after stem cell transplantation, for the mapping of predisposing genes in disease association studies, in population genetics and for evolutionary aspects of population genetics, in forensics and for determination of microsatellite instability in tumours. Several research groups have established a microsatellite marker map for the HLA region with information regarding the microsatellite markers within the HLA region [131-137]. These reports harbour information concerning the length of the repeat, the number of different alleles observed for the marker, the localization of the marker within the HLA region and the primer sequences used to amplify the microsatellite marker.

Disease-associations were initially based on HLA locus/allele associations. Since it is difficult to determine if the HLA gene itself or a gene in linkage disequilibrium with the associated HLA allele or haplotype is involved, further refinement is required. The identification of the gene responsible for the disease can be refined and defined with microsatellite analysis. For example, Behçet disease is associated with HLA-B*5101 [107], but analysis of microsatellite markers located in the vicinity of the HLA-B locus identified an associated region of 46 kb between the MICA gene and HLA-B, indicating that within this region the candidate gene for Behçet disease might be located [138]. MICA analysis revealed associations between Behçet disease and the MICA-A6 repeat and the MICA*009 allele [139-141], HLA-B*51 and MICA-A6/MICA*009 are in linkage disequilibrium; it is therefore difficult to indicate which is the primary disease factor. The same
holds for psoriasis vulgaris, initially associated with HLA-Cw6, but microsatellite association analysis indicated a region spanning 111 kb telomeric from HLA-C to be a susceptibility locus with candidate genes for psoriasis [142]. In other diseases microsatellite analysis has been applied either as a first screening or as a refinement for existing HLA associations to identify susceptibility regions: non-melanoma skin cancer [143], insulin-dependent diabetes mellitus [144, 145], rheumatoid arthritis [146] and idiopathic arthritis [147]. In contrast, very few information is available regarding HLA associations in cancer.

**Head and neck squamous cell carcinoma**

Head and neck squamous cell carcinoma (HNSCC) originates from epithelial cells that line the upper aerodigestive tract. In HNSCC different tumour locations can be distinguished: the oral cavity includes lips, buccal mucosa, alveolar ridges, retromolar trigone, floor of the mouth, anterior two thirds of the tongue, hard palate and gingivae. Laryngeal tumours originate from supraglottis, glottis and subglottis and nasopharyngeal tumours from the nasopharyngeal tonsils. The oropharynx includes soft palate, tonsillar pillars, palatine tonsils, posterior tonsillar pillars, uvula, base of the tongue and vallecula whereas hypopharynx includes the pyriform sinus, cricoid cartilage, thyroid cartilage, hyoid bone and posterior part of the pharynx wall, major (parotid, submandibular, submaxillary and sublingual glands) and minor salivary glands and the ear (external, middle and internal) [148]. The oropharynx, hypopharynx and nasopharynx are grouped together under the name pharynx. Tumours occur most frequently in the larynx, oral cavity and oropharynx (http://www.iKCnet.nl). Nasopharyngeal tumours on the other hand are very rare in the Netherlands whereas in the Mediterranean region and Asia higher incidences are observed for this type of tumour [149, 150]. In general more males are affected than females with HNSCC, although the number of females increases due to changes in their life style in the second half of the previous century. Classification of the tumours occurs via a standardized system, the TNM system that includes 4 stages (I, II, III and IV). This staging is based upon the size of the tumour, presence or absence of lymph node metastases and presence or absence of distant metastases [148].

Although the numbers of all new HNSCC cases diagnosed per year have increased between 1989 and 2003 (1950 new cases in 1989 to 2408 new cases in 2003), the overall incidence of HNSCC per 100,000 inhabitants has not changed dramatically during these years according to the Dutch Comprehensive Cancer Centres (Vereniging van Integrale Kankercentra, VIKC; http://www.iKCnet.nl). In 1989 an overall incidence of 14.1 per 100,000 inhabitants was observed to 13.7 per 100,000 in 2003. If the different HNSCC tumour locations are considered separately in the context of their incidence, differences are observed. The overall
Chapter 1

Incidence of tumours in the oral cavity, and in the pharynx has increased with respectively 44% and 15% in the period 1989-2003, whereas the overall incidence for larynx, lip, nasopharynx/ear has decreased (28%, 33% and 25% respectively). The overall incidence for tumours of the salivary glands did not change. A distinction based upon sex shows that the ratio males/females diagnosed with HNSCC was 3.1 in 1989 and 2.2 in 2003, indicating that the number of females diagnosed has increased during this period. When HNSCC tumour locations are considered in relation to sex during the period 1989-2003, incidences for tumours of the oral cavity, the lip, the larynx and the salivary glands have increased in females while they have decreased in males. The largest increase in females is observed for tumours of the lip: in 1989 12.4% of the lip tumours were present in females while in 2003 this percentage was increased to 25.8% (http://www.iKCnet.nl).

The 10-year survival of HNSCC differs also between tumour locations, their sub-locations and the sex of the patient. Patients with a lip tumour have a good prognosis, 90% of patients are still alive after 10 years; although patients with small tumours and TNM stage I or II have higher survival rates than patients with TNM stage III or IV. For tumours of the oral cavity approximately 50% of the patients are still alive after 10 years, but this depends on the sub-location of the oral cavity tumour, the TNM stage and the age of the patient. A worse prognosis is seen in patients with a tumour in the pharynx; approximately 30% is still alive after 10 years. Increasing age, sub-locations and higher TNM stage are not advantageous for survival of the patients. Sixty percent of patients with a laryngeal tumour are still alive after 10 years; for this type of tumour being a female, older than 75 years, sub-location and higher TNM stage contribute to a poor survival (http://www.iKCnet.nl).

Two major risk factors are correlated with the development of HNSCC. The first is excessive alcohol consumption and the second is smoking tobacco, particularly cigarettes. These two factors confer each an individual risk, but in combination they provide a highly increased risk for HNSCC development [148, 151-154]. Depending on the location of the HNSCC tumour and the sex of the patient, differences in risk contribution can be observed for alcohol and tobacco consumption. Patients who present with advanced stage of disease are heavy drinkers and smokers [151-153]. The rising incidence of females diagnosed with HNSCC is attributable to the fact that their smoking and/or drinking habits have increased in the second half of the 20th century. Several other factors are also potential risk factors for the development of HNSCC: long exposure to sunlight (lip tumours), occupational exposure to nickel, asbestos, and wood, familiar and genetic factors and several viruses [148, 155-157].

Current therapies for treatment of HNSCC are surgery, radiotherapy and chemotherapy, either alone or in combination. Several factors should be taken
into account for the choice of treatment: tumour location, TNM stage and absence or presence of (lymph node) metastases and the preservation of the organ [158]. Tumours that are detected in an early phase can be treated well with surgery or radiotherapy but over time it has become clear that an increasing number of patients present with advanced stage of HNSCC when diagnosed at the hospital [159]. Advanced stage tumours are then treated with surgery or radiotherapy in combination with chemotherapy, which is either given before, during or after surgery or radiotherapy [158, 160]. For these patients, cure is not always feasible; therefore this combination therapy is also given as palliative treatment.

In spite of many years of research, HNSCC remains a very heterogeneous disease for which the precise biological and molecular mechanisms involved are not entirely elucidated. Characterization of these mechanisms involved might eventually lead to a better treatment for the patients. The occurrence of chromosomal aberrations, (epi)genetic changes in tumour suppressor genes and oncogenes, with subsequent altered expression levels, have been observed in HNSCC. The tumour suppressor gene p53 has a high mutation frequency in HNSCC's, resulting in aberrant expression [161, 162]. In addition, loss of heterozygosity (LOH) of the short arm of chromosome 17, on which the p53 is located, is also observed [163]. Loss of the chromosomal region 9p, related with the altered functioning of the p16/p14ARF gene, loss of chromosomal region 3p, with the candidate tumour suppressor genes FHIT and RASSF1A, and amplification of the 13q region, with oncogenes as bcl-1, int-2, hst-1 and proto-oncogene cyclin D1, are observed in HNSCC. Protein and mRNA EGFR overexpression is observed in many HNSCCs and thought to correlate with poor prognoses [163]. Recently, a gene expression predictor profile was identified that allows detection of those HNSCC tumours at risk for development of metastases [164]. A total of 102 predictor genes were identified with this approach, a number of them already known for their involvement in metastases but the majority still unknown and functionally involved as extracellular matrix components, in cell adhesion, cell death and cell growth.

**Head and neck squamous cell carcinoma and HLA**

Genes located on several chromosomes are involved in the carcinogenic process occurring in HNSCC. HLA and other genes within the HLA region on the short arm of chromosome 6 are also important for clearance of the HNSCC tumour cells. Associations with HLA genes have mainly been investigated in patients diagnosed with nasopharyngeal carcinoma (NPC). These studies show that only certain HLA-A and HLA-B alleles are associated with NPC [165-168]. In Asian patients for example, the HLA-A*02-B*46 haplotype confers an increased risk for development of NPC [165, 166]. In addition to HLA-A*02 and HLA-B*46, HLA-
A*11 was identified as a protective allele for the development of NPC [166, 167]. In Caucasian HNSCC patients, the relation between HLA genes and prognosis was investigated and it has been shown that the presence of either HLA-A11 or HLA-DR6 was associated with decreased survival rates among these patients [169]. Whether these associations are directly involved in the carcinogenic process of HNSCC is not yet known, but due to the phenomenon of linkage disequilibrium in this region, it cannot be excluded that other genes, linked to these HLA genes/haplotypes, are the true genetic factors involved.

Aberrant expression of HLA molecules on tumour cells is a well-known phenomenon in tumours. Aberrant HLA class I expression can be classified into groups [170-172]. Monomorphic, locus specific, allele specific and anti-\(\beta_2m\) antibodies were used in primary HNSCC tumours and their lymph node metastases to determine HLA expression [173-175]. Besides the presence of diverse HLA expression patterns in HNSCC, these studies as well demonstrated the important advantage of allele specific antibodies over monomorphic and locus specific antibodies. A striking observation was the difference in HLA expression patterns between primary tumours and lymph node metastases in several patients [174], indicating that knowledge about the expression status is crucial for immunotherapeutical intervention. Several molecular mechanisms have been identified for the aberrant expression patterns [171]. Sequencing based mutation analysis of the \(\beta_2m\) gene indicated that \(\beta_2m\) mutations are rare in HNSCC [176, 177] while in melanomas and colorectal tumours “hotspot” mutations in \(\beta_2m\) have been described [177, 178]. In laryngeal tumours, among other tumour types, loss of HLA-B44 antigen expression was observed more frequently than expression loss of HLA antigens. This was not correlated with a decreased HLA-B*44 allele frequency in the patients [179]. The loss of HLA-B44 antigen expression in colon carcinomas was due to a defective tapasin transcription [180]. Other mechanisms that can contribute to the loss of expression are LOH of chromosomes 6 and 15. It should however be kept in mind that with determination of LOH no discrimination can be made between amplification and loss of chromosomal regions or genes. Additional techniques as fluorescent in-situ hybridisation (FISH) analysis and immunohistochemistry should be used to validate the LOH results [175, 181-183]. HLA expression loss at both the mRNA and protein level can be caused by hypermethylation of the HLA genes themselves. HLA-A, -B and -C have numerous locations where DNA methylation can occur [184]. Aberrant expression of genes involved in the antigen processing and presentation pathway for class I molecules has also been investigated as cause of HLA class I expression loss [174, 178, 185, 186].

Determination of HLA diversity provides information on the functional role of HLA diversity in health and disease, explaining e.g. whether aberrant HLA
expression arises as a consequence of HLA diversity or as a consequence of the involvement of other genes in the HLA region. The availability of this information is useful to acquire knowledge about HLA diversity and functionality in several biological processes. This thesis has been performed to understand if, and how, HLA diversity is functionally involved in the pathogenesis of HNSCC.
Chapter 1

Outline of the thesis

The human leukocyte antigen (HLA) system on the short arm of chromosome 6, the human equivalent of the major histocompatibility complex (MHC), is the most polymorphic region of the human genome and includes three regions: HLA class I, II and III. Within these regions many HLA, HLA-related and non-HLA related genes are located, which have an important function in immune responses. Disease associations with HLA alleles have been described and aberrant HLA expression patterns are observed in tumours. The rationale of this thesis is therefore to determine if and how HLA genes, HLA-related and non-HLA related genes in the HLA region are involved in the pathogenesis of HNSCC.

The role of genes, within the HLA region, in the pathogenesis of head and neck squamous cell carcinoma (HNSCC) can be influenced by the diversity within a gene, or by expression differences. Diversity can be determined with DNA-based techniques, whereas the expression of genes can be determined by using specific monoclonal and/or polyclonal antibodies. The scarcity of available allele specific antibodies limits a broad applicability for the identification of HLA molecules by immunohistochemical methods. RNA-based approaches provide information regarding the diversity, expression and functional status of genes in tumours at transcriptional level and partially compensate for the lack of the limited availability of antibodies. Aberrant HLA expression patterns are present in HNSCC. Exploring the HLA diversity and the functionality of HLA diversity in HNSCC provides insight in the pathogenesis of HNSCC.

HLA gene diversity is determined both at the allele and antigen level for which molecular and serological approaches are widely used. For detection of transcripts, RNA sequencing based typing (SBT) protocols for the class I genes HLA-A, -B and -C and the class II gene HLA-DPB1 were developed. The diversity of HLA, HLA-related and non-HLA related genes within the HLA region was correlated with the pathogenesis of HNSCC.

An RNA SBT approach was developed to evaluate the influence of polymorphisms on RNA transcription. Specific amplification for each of the class I RNA transcripts, subsequently followed by sequencing the entire coding region, was performed for HLA expressed and non-expressed alleles present in reference cell lines. RNA expression of these alleles was evaluated considering the amount of RNA present in the reference cell lines (chapter 2).

The influence of nucleotide polymorphism on RNA expression was examined with the developed class I RNA SBT protocol for HLA-A in a leukaemia patient. A new nucleotide polymorphism affected RNA transcription and cell surface expression of one of the HLA-A alleles (chapter 3).

Few antibodies are available for diversity and expression analysis of HLA-DPB1 antigens. Analysis with RNA approaches provides information regarding
the expression status of DPB1 and in addition DPB1 diversity is obtained. More than 100 HLA-DPB1 alleles have been identified so far, only for a few alleles sequence information is available for the DPB1 regions extending exon 2. An RNA SBT approach was developed to investigate the extended polymorphism and the functional relevance of the extended polymorphism (chapter 4).

HLA-A, -B, -C, -DRB1 and -DQB1 polymorphism was investigated in HNSCC patients and in controls to determine whether particular HLA alleles or haplotypes were associated with HNSCC. The results observed point towards associations with HLA-B*40 and B*35 alleles and the HLA-B*40-DRB1*13 haplotype (chapter 5). The frequency of the associated HLA-B-DRB1 haplotype was low in the patients, hence this could not be the only risk factor involved in HSNCC.

A potential candidate gene, the MICA gene, was further investigated for its role in the pathogenesis of HNSCC. MICA is in linkage disequilibrium with HLA-B and is involved in the immune response. Nucleotide and repeat polymorphism is present throughout the gene. The repeat polymorphism, present in the transmembrane region of MICA, and expression levels of MICA were determined with a monoclonal antibody against MICA. The majority of tumours expressed MICA protein, no correlation was found between the repeat and MICA expression (chapter 6).

A subsequent efficient screening of the MHC region with microsatellites in the class I, II and III regions, using DNA pools from patients with different tumour locations, identified associated regions with candidate genes, particularly in the oral cavity. mRNA expression of 18 selected genes from these regions was evaluated in oral cavity tumour tissue compared to the surrounding healthy tissue. MICA mRNA expression was increased significantly in tumour tissue. In addition, hydroxysteroid (17-beta) dehydrogenase 8 (HSD17B8) mRNA expression was significantly decreased in tumour tissue, especially in those tumours without lymph node metastases, while the receptors retinoid X receptor β (RXRβ) and NOTCH4 showed a not significant but decreased RNA expression in tumour tissue (chapter 7).

The general discussion reflects on the potential functional implication of the results of this thesis (chapter 8).
References


17. König R, Huang LY, and Germain RN, MHC class II interaction with CD4 mediated


19. Lawlor DA, et al., Evolution of class-I MHC genes and proteins: from natural

20. Reinders J, et al., Extended HLA-DPB1 polymorphism: an RNA approach for HLA-

21. Voorter CE, Swelsen WT, and van den Berg-Loonen EM, B*27 in molecular
diagnostics: impact of new alleles and polymorphism outside exons 2 and 3. Tissue

22. Swelsen WT, Voorter CE, and van den Berg-Loonen EM, Sequence analysis of exons
1, 2, 3, 4 and 5 of the HLA-B5/35 cross-reacting group. Tissue Antigens, 2002.
60(3): p. 224-34.

23. Swelsen WT, Voorter CE, and van den Berg-Loonen EM, Polymorphism of intron 4

24. Swelsen WT, Voorter CE, and van den Berg-Loonen EM, Ambiguities of human
leukocyte antigen-B resolved by sequence-based typing of exons 1, 4, and 5.

25. Swelsen WT, et al., Elucidation of exon 1, 4, and 5 sequences of 39 infrequent HLA-

26. van der Vlies SA, Voorter CE, and van den Berg-Loonen EM, There is more to HLA-C
169-77.

27. Voorter CE, Swelsen WT, and van Den Berg-Loonen EM, Intron sequences of HLA-

28. Cereb N, Hughes AL, and Yang SY, Locus-specific conservation of the HLA class I

29. Cereb N, et al., Nucleotide sequences of MHC class I introns 1, 2, and 3 in humans

30. Elsner HA and Blasczyk R, The phylogenies of introns 4-7 demonstrate an
inconsistent pattern between human leukocyte antigen-C group topologies. Tissue

31. Dunn PP, et al., HLA-DQB1 sequencing-based typing using newly identified conserved

32. Pascual M, et al., Complete characterization of the DQB1 first exon polymorphism.

33. Maffei A, et al., Polymorphism in the 5’ terminal region of the mRNA of HLA-
DQA1 gene: identification of four groups of transcripts and their association with
Chapter 1

67. den Hollander N, et al., Identification of a new HLA-B null allele, B*1817 N, in a
Chapter 1


Malkki M, et al., MHC microsatellite diversity and linkage disequilibrium among common HLA-A, HLA-B, DRB1 haplotypes: implications for unrelated donor hematopoietic transplantation and disease association studies. Tissue Antigens,
Chapter 1

144. Lie BA, et al., The predisposition to type 1 diabetes linked to the human leukocyte antigen complex includes at least one non-class II gene. Am J Hum Genet, 1999. 64(3): p. 793-800.


Chapter 1


