DISSECTING GENETIC SUSCEPTIBILITY TO GLUTEN SENSITIVITY: HLA-LINKED RISK FACTORS IN COELIAC DISEASE AND DERMATITIS HERPETIFORMIS

KATI KARELL

ACADEMIC DISSERTATION

To be publicly discussed, with permission of the Faculty of Science of the University of Helsinki, in the Nevanlinna Auditorium of the Finnish Red Cross Blood Transfusion Service, Kivihaantie 7, 00310 Helsinki, on April 25th 2003, at 12 o’clock.
CONTENTS

LIST OF ORIGINAL PUBLICATIONS ................................................................. 5
ABBREVIATIONS ....................................................................................... 6
ABSTRACT ................................................................................................. 7
REVIEW OF THE LITERATURE ................................................................. 9
Human Immunology and HLA ................................................................. 9
  Intestinal mucosal immunology ................................................................. 9
  Human leucocyte antigen, HLA ................................................................. 10
  Disease associations .................................................................................. 17
Genetics of multifactorial (‘complex’) diseases ....................................... 18
  Complex diseases .................................................................................... 18
  Genetic studies ....................................................................................... 19
Coeliac disease, gluten sensitive enteropathy ............................................ 23
  Clinical picture ....................................................................................... 23
  Diagnosis ............................................................................................... 25
  Prevalence ............................................................................................. 25
  Dermatitis herpetiformis ....................................................................... 26
  Pathogenesis .......................................................................................... 27
  Genetics ................................................................................................. 29
  Finnish studies on coeliac disease .......................................................... 32
AIMS OF THE STUDIES .......................................................................... 34
STUDY SUBJECTS AND ETHICAL PERMISSIONS .................................... 35
Coeliac disease patients .......................................................................... 35
  Multiplex families .................................................................................. 35
  Simplex families .................................................................................... 35
  Monozygotic twin pairs ......................................................................... 35
  Hungarian coeliac disease patients ....................................................... 36
  Coeliac disease patients from European Genetics Cluster .................... 36
Samples in study I ...................................................................................... 36
METHODS ................................................................................................................. 37
DNA extraction ........................................................................................................ 37
Microsatellite markers ............................................................................................ 37
Determination of HLA alleles ('HLA typing') ......................................................... 38
Haplotyping and statistical analysis ....................................................................... 38
RESULTS AND DISCUSSION .................................................................................. 39
Microsatellite markers as a practical tool for HLA typing .................................... 39
Same HLA haplotypes confer genetic susceptibility to both intestinal and dermatological manifestation of gluten sensitivity ................................................. 40
Extended B8;DR3;DQ2 haplotype predispose to coeliac disease ....................... 43
Only a few patients with coeliac disease lack the known HLA susceptibility alleles. A multinational European study ......................................................... 46
CONCLUDING REMARKS ....................................................................................... 49
ACKNOWLEDGEMENTS ......................................................................................... 50
REFERENCES ........................................................................................................... 52
ORIGINAL PUBLICATIONS
LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original articles, which are referred to in the text by their Roman numerals:


* These authors have contributed equally.

These articles have been reproduced with the permission of the copyright holders.
### ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AH</td>
<td>ancestral haplotype</td>
</tr>
<tr>
<td>cCD</td>
<td>classical coeliac disease</td>
</tr>
<tr>
<td>CD</td>
<td>coeliac disease</td>
</tr>
<tr>
<td>CEPH</td>
<td>Centre d’Etude du Polymorphisme Humain</td>
</tr>
<tr>
<td>CYP21B</td>
<td>cytochrome 21B-gene</td>
</tr>
<tr>
<td>DH</td>
<td>dermatitis herpetiformis</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>ESPGAN</td>
<td>European Society for Pediatric Gastroenterology and Nutrition</td>
</tr>
<tr>
<td>eTG</td>
<td>epidermal transglutaminase</td>
</tr>
<tr>
<td>HFE</td>
<td>haemochromatosis</td>
</tr>
<tr>
<td>HLA</td>
<td>human leucocyte antigen</td>
</tr>
<tr>
<td>IFN-γ</td>
<td>interferon-γ</td>
</tr>
<tr>
<td>IL</td>
<td>interleukin</td>
</tr>
<tr>
<td>LOD</td>
<td>logarithm of odds</td>
</tr>
<tr>
<td>Mb</td>
<td>megabase</td>
</tr>
<tr>
<td>MHC</td>
<td>major histocompatibility complex</td>
</tr>
<tr>
<td>MICA,-B</td>
<td>MHC class I chain-related A, -B</td>
</tr>
<tr>
<td>NK</td>
<td>natural killer cell</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>TAMRA</td>
<td>tetramethyl rhodamine</td>
</tr>
<tr>
<td>TAP1,-2</td>
<td>transporter associated with antigen processing 1, -2</td>
</tr>
<tr>
<td>TDT</td>
<td>transmission/disequilibrium test</td>
</tr>
<tr>
<td>Th</td>
<td>T helper lymphocyte</td>
</tr>
<tr>
<td>TNF</td>
<td>tumor necrosis factor</td>
</tr>
<tr>
<td>tTG</td>
<td>tissue transglutaminase</td>
</tr>
</tbody>
</table>
Gluten sensitivity or coeliac disease is a relatively common multifactorial enteropathy, where the main environmental factor is the gluten fraction of wheat, barley and rye. The gluten intake may cause variable symptoms in genetically predisposed individuals, classically affecting the gastrointestinal tract but also causing e.g. dermatological, neuronal, or autoimmune manifestations. The only confirmed genetic susceptibility locus is HLA-DQ, located on chromosome 6. Approximately 90% of the patients possess the HLA-DQ2 molecule encoded by the DQA1*05 and DQB1*02 alleles and the rest of the patients carry the HLA-DQ8 encoded by the DQA1*03 and DQB1*0302 alleles.

This thesis is composed of five studies focusing on the role of HLA genes in different manifestations of gluten sensitivity. Firstly, a practical and economical tool for HLA typing was developed using HLA-linked microsatellite marker alleles (I). The alleles proved to be in strong linkage disequilibrium with certain HLA DR;DQ haplotypes in the Finnish population. This method was used in the following studies with gluten sensitive patients, their family members and controls.

To estimate the relative importance between genetic and non-genetic factors on susceptibility to intestinal and dermatological manifestation of gluten sensitivity, six monozygotic twin pairs were examined (II). The results indicated that the same genetic background can result in variable manifestations. The probandwise concordance for gluten sensitivity as a broad entity was surprisingly high, 0.91. No differences in predisposing HLA haplotypes were found either in 25 sib pairs with one having the intestinal and the other the dermatological form of the disease, or in 85 sib pairs with identical manifestation types (III). Similar results with no major differences in HLA distribution between patients with different manifestations were obtained for Hungarian index patients (III).

The HLA-DQ2 heterodimer is the major predisposing genetic factor for coeliac disease, but still only a minority of DQ2 carriers at population level are affected. To look for possible differences in DQ2 positive haplotypes, a segregation study in families with a DQ2 homozygous parent was performed (IV). The DQ2 positive haplotypes transmitted to affected offspring were more conserved than the non transmitted DQ2 haplotypes; the most significant statistical difference was in the frequency of the HLA-A*01 allele (p=0.007). To reveal additional risk factors within the HLA region, coeliac patients without the predisposing HLA-DQ2 were collected from a large European coeliac disease material consisting of 1008 families (V). This study definitely confirmed that coeliac disease without the known predisposing DQ alleles (DQA1*05, DQB1*02, or DQB1*0302) is very rare: only four (0.04%) such patients could be found. These four patients had no similarities in their HLA alleles, symptoms, age of onset or family history.

The present studies demonstrate that the HLA molecules DQ2 and DQ8, or either the DQA1*05 or DQB1*02 allele predispose to both the intestinal and dermatological manifestation of coeliac disease, and that patients without these genetic risk factors...
are very rare. Our results further suggest, that from a genetic point of view, these two forms in fact can be regarded as different manifestations of a single disease entity. Finally, we provided evidence that the HLA region may contain a novel non-DQ risk factor for coeliac disease, but we were not able to pinpoint its location. As similar results have been obtained in other HLA-associated diseases, the complexity of the HLA-disease associations may be higher than previously assumed.
HUMAN IMMUNOLOGY AND HLA

The immune system has a life-saving task protecting the human body from destruction by harmful invaders such as bacteria or parasites, maintaining tolerance to self-structures and destroying old, mutated or virulent cells. The system is partly complete at birth, but the specificity increases throughout life as a consequence of coming into contact with foreign antigens. The major populations of cells responsible for the increasing specificity, so called adaptive immunity, are T and B lymphocytes, often simplified into T and B cells.

Intestinal mucosal immunology

The gastrointestinal tract is continuously exposed to different dietary antigens, pathogens and resident microflora. Especially the intestine has to balance between controlling the access of potentially harmful antigens and at the same time trying to absorb necessary nutrients. The outermost part of the intestine is a mucosa, which covers circa 400 m² as a single cell epithelial layer. The epithelium is spread over villi, finger-like structures. The villus epithelium contains mainly enterocytes and intraepithelial lymphocytes, both which reside above the basement membrane. Under the basement membrane is lamina propria, connective tissue containing several cell types e.g. T and B lymphocytes and dendritic cells. The villi are separated by crypts of Lieberkühn (Fig. 1). The essential nutrients may easily pass the thin epithelia by different transport mechanisms but the peptide antigens recognised as foreign are mainly caught up by antigen presenting cells.

Figure 1. Epithelial layer covers intestinal villi. The epithelium consists of enterocytes and intraepithelial lymphocytes (IELs), mostly cytotoxic T cells. Under the basement membrane, in lamina propria, there are more lymphocytes, mainly T helper cells and B cells. The surface epithelium is continuous with that in crypts of Lieberkühn, and embedded in lamina propria. Picture according to Halttunen, 2002.
The places for T and B lymphocytes to mature are secondary lymphoid tissues, in the intestine called Peyer’s patches. There B cells start producing antigen-specific immunoglobulins, mainly IgA, and T lymphocytes develop into CD8+ cytotoxic T cells or activate and differentiate into Th1 or Th2 type of CD4+ T helper cells. The activation of T cells requires two separate signals. First is the interaction between the T cell receptor-CD3 complex and the human leucocyte antigen (HLA) molecule of the antigen presenting cell. The second signal is produced by costimulatory and adhesion molecules, which encounter ligands on T cells, CD28/CD80 and CD86, CD2/LFA-3 (leucocyte functional antigen), LFA-1/ICAM-1 (intercellular adhesion molecule-1). The subset specific differentiation of CD4+ T cells depends on cytokines. Interleukine (IL)-4 induces Th2 cell differentiation whereas IL-12, IL-18 and IFN-γ (interferon-γ) generates Th1 cells. The physiological difference of matured T helper cells appear on their cytokine production; Th1 produces mainly IFN-γ, induces inflammation and promotes cytotoxic immune responses while Th2 is involved in B cell activation and humoral immune responses with the production of IL-4, IL-5, IL-9 and IL-13 (1;2). The matured CD4+ T cells home from Peyer’s patches back to lamina propria and stay there mediating specific adaptive immune defence. The lymphocytes within the epithelial layer are mostly CD8+ cytotoxic T cells whose cytokine profile is determined by the activation signal, mainly resulting in the production of Th1 type cytokines.

A particular population of CD8+ T cells bears a γδ T cell receptor. In the periphery they are rare, only about 5% of T cells, but in the gut they comprise even 10% of intraepithelial lymphocytes (3;4;5). The mechanism they apply in the immune response differs from that of T cells with the ‘standard’ αβ receptor. The γδ T cells recognise stress-associated MICA and MICB molecules (6) and secrete cytokines mediating both innate and adaptive immunity (7). They may exceptionally develop in the gut independently of the thymus and appear in the gut wall before more common T cells with αβ T cell receptors (8). The γδ T cells may also take part in the genesis of oral tolerance.

Oral tolerance is a typical character of mucosal immunity. It is a systemic unresponsiveness to peripheral antigen challenge like food antigens, thus preventing otherwise devastating hypersensitivity reactions. The detailed mechanisms of oral tolerance are unknown but at least T cells are known to mediate it by active suppression or by regulatory cells, by going into anergy or by becoming deleted (9).

**Human leucocyte antigen, HLA**

In T cell activation the peptide antigen is presented by the human leucocyte antigen (HLA) molecule to the T cell receptor. The genes coding for HLA molecules are located on chromosome 6, in special HLA region at 6p21.3.

The HLA region compasses approximately 4 Mb of DNA and it is one of the gene richest segments within the human genome as it contains more than 200 genes. The region is often split into three regions based on the functional characteristics of genes within each class, named from telomeric class I, class III and class II (Fig. 2).
The genes for classical HLA molecules are located within class I and class II region, but many other genes involved in immune regulation are found scattered throughout the HLA region. (10;11)

**Classical HLA molecules**

Major histocompatibility (MHC) molecules were first discovered in mice while studying tissue rejections following transplantations. Hence the name. Later similar molecules have been observed in every vertebrate from zebra fish to humans. In man MHC is called the human leucocyte antigen, HLA. (12;13)

The molecules involved in antigen presentation to T cells are called classical HLA molecules. They are divided into two groups: HLA-A, -B and -C are transcribed from class I region and they present mainly endogenously synthesised peptides like viral or tumour peptides to CD8+ cytotoxic T cells while HLA-DR, -DQ and -DP from class II present predominantly exogenous antigens as bacterial peptides to CD4+ T helper cells. This functional division of peptide presentation ensures that all deleterious antigens get proper treatment; cells with mutated DNA are destroyed immediately by cytotoxic T cells and a specific B cell immunoglobulin repertoire is created for exogenous antigens in addition to clearance of antigens aided by T helper cells.

---

**Figure 2.** The map of HLA region on chromosome 6 (according to Klein and Sato, 2000).
HLA class I molecules

Most nucleated cells express the HLA class I molecules, with exceptions such as cells from the central nervous system, sperm, oocytes, placenta and endothelial cells of lens. The first three-dimensional structure of an HLA class I molecule, HLA-A2 was illustrated in 1987 (14). Since then a similar structure has been observed in all studied class I molecules (Fig. 3a).

The HLA class I molecule is composed of two polypeptide chains: an α chain encoded by the HLA locus and a lighter β₂-microglobulin encoded by chromosome 15. The α chain is an integral membrane protein of approximately 43 kD in size. Its extracellular portion has three domains of which α1 and α2 form the functionally important structure: the peptide binding groove (Fig. 3c). The groove consists of eight variable β-sheet and two α-helical structures and it is capable of binding the peptide antigen and presenting it to T lymphocytes. Briefly, intracellular peptide antigens are cut into peptides of 8-11 amino acids in length. They bind to the groove and following transportation of the complex onto the cell membrane, are presented by HLA molecules to T cells. An important feature is the extensive genetic variation of the HLA molecules (see chapter below), which is enriched in the amino acids forming the groove. Consequently, the exact structure of HLA molecules and importantly their capacity to bind peptides and present them to the immune system differ between individuals. (15)

Figure 3. Simplified illustrations of classical HLA molecules from class I region (A) and class II region (B). A peptide binding groove with α-sheets and β-helixes of HLA-A*02 molecule (C) pictured from above (according to Bjorkman et al 1987). β₂m = β₂-microglobulin, TM = transmembrane

HLA class II molecules

HLA-DR, -DQ and -DP molecules are part of the immunoglobulin superfamily like class I molecules with related structure and function. However the distribution of class II molecules is more limited as these integral membrane glycoproteins are expressed mainly on professional antigen presenting cells: on B cells, macrophages and dendritic cells and after cytokine induction on epithelial cells and fibroblasts.

The class II molecules are also heterodimers but both chains are encoded from the HLA region, hence there are always two genes for class II molecules, such as DQA
and DQB, or DPA and DPB. There can be even more genes like DRB1-DRB8, but of those only DRB1, DRB3, DRB4 and DRB5 are functional, the others being pseudogenes. The chains, \( \alpha \) and \( \beta \), consist of four extracellular domains noncovalently detached (Fig. 3b). \( \alpha 1 \) and \( \beta 1 \) form a groove similar to class I molecules with \( \alpha \)-helixes and \( \beta \)-sheets. The peptide binding groove of class II molecules differs from that of class I molecules in its ability to bind longer peptides, starting from 13 amino acids and going even as high as 24. \( \alpha 2 \) and \( \beta 2 \) domains are membrane proximal domains supporting the peptide-binding platform. (15;16)

**Nomenclature and polymorphism**

Observation of leucocytes agglutinating with antisera from transfused individuals lead to the discovery of the HLA system in the 1950’s. At the beginning, the names of different alleles were given on the basis of serological typing. However, now with DNA sequence data available the allele names are based on sequenced genes. The more detailed typing methods have increased the number of different alleles dramatically. For example, the serological HLA-A2 specificity actually includes 68 alleles (HLA-A*0201 – *0260 / www.anthonyanlan.org.uk/HIG/lists/class1list.html 20.11.2002) as defined by their DNA sequence. The allele name always has the name of the region (HLA), locus (A) and the particular group of alleles (*02), usually also the name of the specific allele, such as HLA-A*0201 which means allele 01 of the HLA-A*02 group (Table 1). In addition, there can be information of silent mutations (fifth and sixth digits) or mutations outside of the coding regions (seventh and eighth digits). Exceptional HLA gene expression is indicated by letters: N for null allele, L for low expressed allele, S for soluble secreted molecule, C for cytoplasmic product and A for allele with an aberrant expression. In cases where numbers exceed 99, a second number series will be introduced (17). An attempt to simplify scientific text has actually increased confusion, as often the HLA class II heterodimers are called by their serological equivalents instead of writing both alleles encoding molecules e.g. DQA1*0501 and DQB1*0201 form a serological specificity called DQ2, which is also the name for the molecule encoded by DQA1*0201 and DQB1*0202.

**Table 1.** The nomenclature of the HLA alleles is based on the sequenced data and HLA gene expression. First comes the name of the region followed by the particular locus with indication of the allele group. The specific allele is identified by the third and the fourth numbers which may be followed by a mark for aberrant gene expression (see text) or for a synonymous mutation (fifth and sixth numbers) or for a mutation outside of the coding region (seventh and eighth numbers).

<table>
<thead>
<tr>
<th>Nomenclature</th>
<th>Indication</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA</td>
<td>gene region</td>
</tr>
<tr>
<td>HLA-A</td>
<td>particular locus</td>
</tr>
<tr>
<td>HLA-A*02</td>
<td>group of alleles</td>
</tr>
<tr>
<td>HLA-A*0201</td>
<td>specific allele</td>
</tr>
<tr>
<td>HLA-A*020101</td>
<td>allele with synonymous mutation</td>
</tr>
</tbody>
</table>
Classical HLA genes are among the most polymorphic ones in humans (18). At least 250 HLA-A, 488 HLA-B, 118 HLA-C, 387 HLA-DRB, 53 HLA-DQB1, 99 HLA-DPB1 alleles had been identified worldwide by August 2002 (www.anthonynolan.com/HIG/30.7.2002). The polymorphism is created by point mutations, gene conversions and recombinations. At the amino acid level the most important changes take place in the hypervariable regions encoding the amino acids which form the peptide binding groove. These mutated hypervariable regions allow binding and presentation of new and different peptides to T cells increasing the potential of the immune response. The mutation rate within the HLA genes is relatively high (10⁻⁸)(19) and interestingly, very few mutations within the coding region appear to be silent compared to the mutations on non-coding regions. This speaks in favour of evolutionary force towards amino acid variation.

The huge number of different alleles gives rise to a variety of HLA combinations, and thus to different and versatile immune responses between individuals. The diversity is even increased as a typical feature with HLA alleles is the codominant expression, meaning that the alleles from both parents are expressed at the same time, giving a minimum number of twelve different classical HLA molecules in a heterozygous person. The variety is also increased by HLA-D molecules encoded in trans e.g. DQA1 inherited from the mother is combined with paternal DQB1. In addition, individuals with DRB1*03, *11, *12, *13 or *14 usually have an expressed allele in the DRB3 locus, those with DRB1*04, *07 or *09 express the DRB4 locus and those with DRB1*15 or *16 alleles express the DRB5 locus (Fig. 4). The peptides encoded by the DRB3, DRB4 and DRB5 genes combine with the DRA protein to form DR52, DR53 or DR51 molecules, respectively. (20)

Annual meetings of the WHO Nomenclature Committee for Factors of the HLA System, monthly updates in journals as well as online, help researchers to keep up to date with HLA nomenclature. The founding of new alleles is important knowledge especially for clinical tissue typing laboratories, where the matching of transplants is carried out. Since the importance of HLA typing was originally demonstrated by Patel et al (1968) (21) for the selection of kidney donors, all patient - donor pairs are

![Figure 4](image-url)
now routinely matched for their HLA-A, -B, and -DR alleles for solid organ transplantation. For stem cell transplantation the requirements are much stricter and basically the matching is done for every classical HLA loci.

Other genes of the HLA region

Soon after genomic cloning of the HLA gene region (22) it was clear that there is a high number of other genes located near and between the classical HLA genes. Many of those genes are involved in the immune regulation in addition to the classical HLA genes. Some of them are directly involved in maturation of HLA class I or II molecules such as TAP1 and TAP2 (transporter associated with antigen processing 1 and -2) presenting antigens to class I molecules and LMP2 and LMP7 (low molecular weight protein 2 and -7), which are parts of proteasomes. Proteasomes degrade antigens into peptides, which bind to the class I molecules. Tapasin from class II anchors the class I heterodimers into TAP1 and TAP2 to wait for antigen before moving to the surface. Other molecules encoded on the HLA region and aiding class II heterodimers to mature are e.g. HLA-DM which facilitates antigen binding to class II molecules and its negative regulator HLA-DO. (20)

The HLA region also contains genes for the complement components C2, C4 and factor B, for the cytokines TNFa (tumor necrosis factor), LTA and LTB (lymphotoxin A and -B) and for the heat shock protein HSP70. Cells expressing HLA-E or HLA-G (mainly on placental cells) derived from class I region are not killed by NK cells, while members of the MIC gene family, MICA and MICB (MHC class I chain related gene A and B), activate killing. MICs have a similar structure to HLA class I molecules but without β2-microglobulin. They have no capacity for antigen presentation but instead they serve as a ligand to NK cells, γδ T cells and some CD8+ T cells (23;24). Another molecule structurally related to class I molecules, is HFE (haemochromatosis), which is responsible for iron metabolism. Its gene is located in the extended HLA region, 4.6 Mb telomeric from the classical HLA border, but considered as an HLA molecule because of its structural similarity. Another gene not even structurally related to HLA is the cytochrome 21B-gene, CYP21B, whose product controls intermediate steps in mineralocorticoid and glucocorticoid biosynthesis (25).

The genes are scattered throughout the HLA region, the HLA class III region being the most gene-rich region with approximately one gene per 15 kb (26). In addition to functional genes, HLA contains numerous pseudogenes, repeated stretches called microsatellites and genes without a yet identified functions.

Linkage disequilibrium and ancestral HLA haplotypes

Strong linkage disequilibrium is a special feature within the HLA region. It is a phenomenon linking two alleles so that they occur together more frequently than expected. For example the gene frequencies for both HLA-A*01 and HLA-B*08 are about 16% in Finland (27). If their inheritance was random and independent of each other the frequency of the combination of A1 and B8 would be about 2.5% (16% x 16%) but in reality it is about 7.7% (28), meaning alleles occur more commonly together than expected.
A set of alleles, which are inherited en bloc, is called a haplotype. The conserved allele combinations within HLA region are called extended or ancestral haplotypes (AH). The ancestral haplotypes are relatively similar between populations and have been claimed to cover even 30% of the known haplotypes in Caucasoids (29). However their frequencies are population specific as a result of population history and selection pressure.

One well known ancestral haplotype is HLA-A1;B8;DR3;DQ2 i.e. AH 8.1. The AH 8.1 is very common in Caucasoids with a frequency of 4.7% in Finns (28). It has been shown to extend exceptionally long, even 6 Mb telomeric from HLA-A until microsatellite D6S276 (30) and on centromeric side HLA-DPB*0101 is often part of it (31). This is exceptional as DPB1 is rarely in linkage disequilibrium with other class II genes. The AH 8.1 is well studied as many diseases associate to it and it has special immunological character e.g. apoptosis impairment (32), lower cytokine production and an increase of IgM producing B cells (31).

The reasons for the conservation of ancestral haplotypes are still unknown. There has been speculation of the advantages of maintaining allelic combinations and thus providing special phenotypes. The AHs may also lack some recombination-promoting sequences, so called hotspots, preventing changes or they are simply so new that there has not been enough time for changes to occur yet (31).

Crossovers cause interlocus recombinations and thus create new haplotypes. Crossovers are restricted to certain places, hotspots. The hotspots split the genome into relatively stable blocks with high linkage disequilibrium. The length of the blocks in Caucasoids is quite constant, about 50-60 kb and the hotspot regions cover circa 5 kb. However the length of blocks varies between population and e.g. extending much less in Nigerians (33). This can be due to selection and population history: size and amount of admixture. The length of blocks varies also between chromosomal locations and in the HLA region some exceptionally conserved blocks i.e. ancestral haplotypes can be seen.

The well known hotspots within the HLA class II region are located in four places: between HLA-DOA and RING3 (really interesting new gene 3), between DOB3 and DOB1 and around TAP2 (34;35) and in the vicinity of DMB (36). On the other hand there is almost a total lack of recombinations between DQ and DR and DP and DOA. In the class I region there are two hotspots namely around P5 and HLA-E, maybe also between HLA-C and OTF3 (octamer-binding transcription factor 3) and in centromeric of MICA (37) leaving at least one unbreakable region between HLA-B and -C (38).

Hotspots were noticed to be partly haplotype specific in mouse (39) and later same phenomenon was observed in humans with the rare HLA-B60;Cw*0304 haplotype (40) and with the HLA-B47;DR7 haplotype (41) both having recombinations occurring between the TNFa and D6S273 microsatellite marker.
Disease associations

Over 500 diseases are associated either with classical HLA alleles, with some other genes within the HLA region or with whole HLA haplotypes. Many of them are autoimmune disorders, multifactorial diseases with the HLA gene being only one of numerous predisposing or protecting factors. (Table 2)

An example of a disease with a specific predisposing HLA allele is ankylosing spondylitis which is associated with HLA-B*27 so strongly that HLA-B typing is used even as a diagnostic tool; the B27 increases relative risk by 150 compared to people not having that specific HLA-B allele. Alleles can also be protective like HLA-B*53 is for malaria caused by infection of Plasmodium falciparum. People with the B53, especially homozygous for the B53 resist malaria well or suffer mild symptoms compared with non-B53 carriers. The B53 is relatively common in sub-Saharan people but rare in non-African populations. Defects in genes regulating other than obvious immune function but still located in or near the HLA region may cause iron-storage disease, haemochromatosis (HFE gene) or congenital adrenal hyperplasia (CYP21B gene). (42;43)

The disease associations may be relatively complex e.g. in multiple sclerosis and type I diabetes. Multiple sclerosis seems to have one strong predisposing HLA DR;DQ haplotype as well as other minor, maybe population specific protective and predisposing haplotypes (44). The most predisposing haplotype combination associated with type 1 diabetes is DR4;DQ8/DR3;DQ2, where the predisposing part is the combination of DQA1*03 (of DQ8 molecule) and DQB1*02 (of DQ2 molecule). However both DQ8 and DQ2 molecules themselves are also predisposing in a single dose. It is noted that in addition to DQ molecules DR alleles also have an effect on the development of type 1 diabetes. Different DRB1*04 alleles have varying risk, DRB1*0405 and *0401 being the most predisposing followed by *0402. HLA-DRB1*0406 and *0403 are dominantly protective alleles besides DQ6 molecule. In addition to these class II alleles, over 20 putative diabetes-predisposing loci have been reported, of which class I alleles HLA-A*2401 and A*3002 are some of the latest. (45-48)

<table>
<thead>
<tr>
<th>Class</th>
<th>Class II</th>
<th>other</th>
</tr>
</thead>
<tbody>
<tr>
<td>B27</td>
<td>Ankylosing spondylitis</td>
<td>DR4, DQ8, DQ2*</td>
</tr>
<tr>
<td>BS3</td>
<td>Malaria</td>
<td>DR2,DR3</td>
</tr>
<tr>
<td>BS1</td>
<td>Bechet's syndrome</td>
<td>DR4</td>
</tr>
<tr>
<td>A2902</td>
<td>Birdshot retinitis</td>
<td>DR2, DQ6</td>
</tr>
<tr>
<td>Cw6</td>
<td>Psoriasis vulgaris</td>
<td>DQ2*, DQ8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DQ6</td>
</tr>
</tbody>
</table>

*DQ2 is considered here as a combination of DQA1*05 and DQB1*02.
CAH = Congenital adrenal hyperplasia, SLE = Systemic lupus erythematosus

Table 2. Numerous diseases are associated with the HLA region, either with classical class I or class II genes or with other loci.
GENETICS OF MULTIFACTORIAL (‘COMPLEX’) DISEASES

Complex diseases

For the development of a multifactorial disease a set of genes and certain environmental factor(s) are needed (Fig. 5). The number of the susceptibility genes may be large and even differ between individual patients. In addition to the locus heterogeneity, the predisposing alleles at a specific locus may vary and the genes interact with each other epistatically. Also incomplete penetrance, a non-uniform occurrence of a disease between individuals with the same set of susceptibility genes, or phenocopies, the same phenotype resulting from different set of genes, are typical features of complex diseases. As the multifactorial disorders do not follow the classical Mendelian laws, the onset of diseases may not be readily predicted from the family history. (49)

The effect of the environment varies between diseases. Sometimes the environmental factors, at least the major ones with direct effect, are known, like gluten in coeliac disease or smoking in lung cancer. Increased incidence of asthma has been explained by a lower burden of parasites causing the immune system to turn against normally neutral antigens like pollen or even the body itself inflicting autoimmune diseases. Bacterial DNA, viruses and parts of micro-organisms have also been associated with the disease development; as adjuvants they may stimulate the immune system into overactivation and hence induce autoimmune disease in susceptible individuals.

Most autoimmune diseases are complex disorders with a strong HLA association. They are chronic conditions caused by the breakage of immunologic tolerance to self-antigens. Many of the autoimmune diseases are more common in females

Figure 5. Many factors influence the development of complex disease. There can be numerous different genetic factors (1-4) in addition to environmental and cultural factors (5). The combination of genetic factors may vary between people and genes themselves may influence each other epistatically (6). The statistical methods rely on markers (7) that can be used to locate predisposing genetic factors (8).
than males and on average affect 3-5% of the general population. The diseases can be classified as organ specific like type 1 diabetes where pancreatic islet cells are destroyed or systemic ones which affect the whole body such as systemic lupus erythematosus. (50)

**Genetic studies**

Attempts to solve the predisposing genes behind complex diseases have proven to be a hard task as the genes with small effects vary between similarly affected people. An essential assumption to overcome these obstacles is a carefully chosen study material or methods. A difficulty is to get as many genetically homogenous patients as possible, but still have a sufficiently large number of patients.

**Choosing material**

To diminish the effect of the environment or other genes, the chosen patients should be as similar as possible for the genetic, environmental and cultural risk factors. The genetic homogeneity is maximum when studying monozygous twins and they are valuable samples for measuring the overall effect of genes versus the environment by concordance. In an attempt to find predisposing genes the monozygotic twins are useless. Instead dizygotic twins, siblings or relatively homogenous populations have been widely used as too much heterogeneity exists in most general population samples. After the gene has been roughly located the transracial studies may be helpful in pinpointing the exact predisposing factor.

**Twins**

In classical twin studies the concordance and discordance rates of monozygotic and dizygotic twin pairs are compared to define the relative effects of genes and environmental factors for the outbreak of a disease. This clearly is one of the first points in any study of the genetics of a disease. Casewise concordance rates for typical complex diseases are variable: they vary between 25-86 % for monozygotic twins and are significantly lower (0-20 %) for dizygotic twins (Table 3)(51-53). When the concordance rate in monozygotic twins is high, as is the case in late onset diabetes (54) or coeliac disease the genes can be assumed to play a more important role than the environmental factors. If there is no difference in concordance rates between monozygotic and dizygotic twins, the role of genes must be assumed to be low and the effect of a shared environment high. Usually diseases occurring in older people have a higher concordance compared with ones affecting children (55).

It is important to remember that not even monozygotic twins are ever genetically totally identical, a percent difference is allowed. This difference comes from point mutations, skewed X- chromosomal inactivation, non-disjunctions and from gene rearrangements of Ig and T cell reseptors. In addition to the genetic differences, there may be differences in methylation and in the levels of protein translation. Monozygotic twins also commonly share a similar childhood environment, and pressure from the surroundings may link the twins more closely together than other siblings adding to the shared environmental factors. (56)
Population isolates

An attempt to decrease patients heterogeneity and reduce the complexity of polygenic disorders has brought population isolates into science (57). People from isolates have less environmental and genotype variation, the latter meaning less allelic and locus heterogeneity and longer conserved haplotypes caused by linkage disequilibrium. Hence less markers and fewer families are needed to study compared to analysis with outbred populations. Isolated populations also have relatively more recessive disorders and a usually well documented history with known pedigrees. These reasons have made isolates the most important tool in genetic studies of inherited disorders. The drawbacks compared to mixed populations include a limited number of patients and occasionally negative attitudes towards participating in studies.

The Finnish population has been used widely in the search for genes for genetic disorders. Finns have a typical founder effect caused by a few ancient invaders who settled in with a limited number of genotypes. Throughout the thousands of years with relatively few immigrants, some bottlenecks and times of rapid growth, genetic drift has had time to modify the gene pool seen in Finnish people today. A typical character is the Finnish disease heritage with about 40 diseases observed with a higher frequency in Finland than elsewhere in the world and commonly caused by a major ‘Finnish’ gene mutation. On the other hand, the population has much less certain relatively common inherited diseases such as cystic fibrosis and phenylketonuria. The Finnish population is also favoured in studies because of high quality public healthcare and church records dating from the 17th century. All this enables collections of large pedigrees with accurate and well preserved medical records. (58;59)

Other populations actively applied in genetic mapping studies include small isolated groups of Amish, Hutterites and Mennonites from the USA, genetically peculiar Basques from the Basque Country, Bedouins from the Middle East and North Africa with a family study of nonsyndromic deafness, well studied Icelanders with the on-going deCODE project, Sardinians from their isolated island and Jews, especially Ashkenazi Jews. In the need for even smaller subisolates and more homogeneous

Table 3. Concordance rates of monozygotic (MZ) and dizygotic (DZ) twins for various autoimmune diseases are low compared with concordance in coeliac disease.

<table>
<thead>
<tr>
<th>Disease</th>
<th>MZ</th>
<th>DZ</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coeliac disease</td>
<td>86</td>
<td>20</td>
<td>Greco et al 2002</td>
</tr>
<tr>
<td>Crohn’s disease</td>
<td>33</td>
<td>12</td>
<td>Thompson et al 1996</td>
</tr>
<tr>
<td>Graves disease</td>
<td>36</td>
<td>0</td>
<td>Cooper et al 1999</td>
</tr>
<tr>
<td>Multiple sclerosis</td>
<td>34</td>
<td>6</td>
<td>Cooper et al 1999</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>25</td>
<td>7</td>
<td>Cooper et al 1999</td>
</tr>
<tr>
<td>Systemic lupus erythematosus</td>
<td>36</td>
<td>3</td>
<td>Cooper et al 1999</td>
</tr>
<tr>
<td>Type 1 diabetes</td>
<td>46</td>
<td>13</td>
<td>Cooper et al 1999</td>
</tr>
<tr>
<td>Ulcerative colitis</td>
<td>27</td>
<td>6</td>
<td>Thompson et al 1996</td>
</tr>
</tbody>
</table>
people with wide genealogical databases, people from Quebec, Costa Rica, parts of Holland and Finland have been studied. (59;60)

Phenotypes

If the disease is rare it might be impossible to collect a sufficient number of patients from any population isolate. Then one can try to limit the genetic heterogeneity by choosing patients with only certain strictly restricted phenotypes as it can be thought that more similar phenotypes would rise from a more similar set of genetic factors. By dividing migraine patients according to their phenotypes, two different predisposing loci for the different subtypes have thus far been found (61;62).

It is important to define all patient data accurately from the beginning to allow later grouping of patients based on e.g. age of onset, manifestation type or degree of difficulty, sex or inheritance. Still the results may be unreliable as there might be phenocopies resulting in similar phenotypes from different sets of genes or a specific mutation may not cause the same symptoms. This is typical for the CYP21B gene deficiency as the siblings carrying the same mutations may have variable phenotypes ranging from totally symptomless to severely affected ones (63).

Choosing methods

Statistical methods are based on measuring the correlation between the disease phenotype and genotype. The basic methods used for defining the predisposing genetic loci by correlation are linkage and association methods.

Microsatellite markers and defining genotypes

The genotypes can be defined by genetic markers, classically by blood grouping, later with restriction fragment analysis (RFLP), short tandem repeats or by single-nucleotide polymorphisms, SNPs. SNPs are especially suited for analysis where dense marker maps are needed such as in candidate gene studies as there is about one SNP every 1.9 kb (64).

Microsatellites are short tandem repeats, ranging from 2-6 nucleotides, the most common being (CA)_n repeat. Sequences can be repeated from tens to hundreds of times. Microsatellites are widely used in genotyping as they are common, found approximately one in every 10 kb, and throughout the genome (65). These markers are informative if they have more than five alleles and a heterozygosity value exceeding 0.75. On average markers in the HLA region have 8.5 alleles and more than half have a heterozygosity value over 75% (66).

Microsatellite marker genotyping is relatively easy with automated electrophoresis. The microsatellite allele frequencies are population specific. The different alleles are produced by either unequal crossing over or slippage during DNA replication. Usually the number of repeats changes by one, either increasing or decreasing, and more commonly changing in already long sequences (67;68). The role of microsatellites, if any, is unknown but they are assumed to condense DNA packing,
regulate gene activation or to act as recombination hotspots (69). The increase in the number of repeats is known to affect the probability of developing a fragile X syndrome (70).

Linkage studies

Parametric linkage has been used successfully when mapping genes causing simple Mendelian diseases (71). It is a powerful and practical method when large families with at least two affected individuals are available and certain genetic parameters, such as penetrance, dominance/recessiveness, recombination fractions, trait and marker allele frequencies can be accurately measured. As these parameters are seldom known, or may not actually ever be defined for complex diseases, the method is not very effective in multifactorial diseases.

Linkage study is based on the assumption that affected family members share the same predisposing gene. As the patients with the shared predisposing gene seldom have recombinations in the vicinity of the disease gene, with linkage studies the predisposing gene is tried to be found with the help of markers located next to the disease gene. This conserved region is assumed to have less recombinations indicated by the theta value, $\theta < 0.5$. The linkage results are plotted to a logarithm of odds (LOD, a lod score). A lod score larger than 3 is considered significant for a single tested locus and a LOD >3.3 for genome-wide analysis (72).

In non-parametric methods the penetrance is not defined, but the allele sharing of relatives with a similar phenotype is measured. This can be done within families with parents and two affected siblings (affected sib-pair method) or between other relatives with the help of inheritance vectors (non-parametric linkage analysis). Both methods have numerous variations.

In genome-wide scans predisposing genes are tried to be found by using evenly spaced genetic markers covering the whole genome. Following the initial genome screen, so called candidate regions with the highest lod scores are narrowed down with other methods. Genome wide scans are usually performed with non-parametric sib pair methods in a particular population and they need to be replicated in other populations, even in old outbreed populations (73).

Association studies

The idea of association studies is to find alleles which are inherited by patients more commonly than by controls. The shared region indicated by shared alleles can be expected to contain the predisposing factor for the disease.

The case-control approach compares the allele frequencies of unrelated patients and their matched controls. Of the family-based tests the simplest method is the transmission/disequilibrium test, TDT, where inheritance from heterozygous parents is traced to affected offspring and the number of transmissions and non-transmissions are plotted in a contingency table (74). The advantage of a family-based TDT is the lack of a population stratification error but on the other hand,
collecting families is time-consuming and not always possible. Nowadays there are a wide variety of association tests either based on the case-control or family studies.

Tests based on association are used to narrow down the candidate regions revealed by linkage studies or when the functional candidate gene approach is used. Then numerous markers are genotyped densely covering a small precise region. The result will pinpoint the specific region with the highest linkage disequilibrium between the disease locus and its nearest marker. Unfortunately the use of dense marker maps increases type I error. Hence it is important to have unaffected well matched controls showing that the linkage disequilibrium is due to the proximity to the disease locus and not because it is common in the general population (75).

COELIAC DISEASE, GLUTEN SENSITIVE ENTEROPATHY

Coeliac disease (CD) (OMIM 142800) is a multifactorial enteropathy, where the main environmental risk factor is gluten. Gluten is a storage protein in cereals and the ingestion of this protein from wheat, barley or rye may cause mucosal lesions in the small bowel in genetically susceptible individuals. Mucosal injury usually affects the patients’ quality of life, and in the long run increases the risk for associated diseases.

Aretaeus from Cappadocia was the first to describe a gastrointestinal condition resembling coeliac disease already in the second century A.D., but the first thorough description of clinical symptoms by doctor Gee dates back to the year 1888. It took another 60 years and World War II before the role of gluten as the trigger for coeliac symptoms was solved. Dutch paediatrician Dicke noticed a recovery of coeliac patients during a shortage of bread in the war time and the reoccurrence of symptoms after cereals were reintroduced to diet (76). Since then the only permanent treatment for patients has been a strict gluten free diet. (77)

Clinical picture

Classical coeliac disease is an enteropathy where gluten ingestion destroys the architecture of the small intestine causing a variety of symptoms. In children the symptoms follow the introduction of cereals into diet. The classical gastrointestinal symptoms in paediatric patients are chronic diarrhoea or steatorrhoea, vomiting, abdominal distension and pain, oedema and pallor. Adult patients may present episodic diarrhoea, weight loss, abdominal discomfort, bloating and flatulence. Today most of the diagnosed patients only have mild symptoms.

The various complications have been shown to be very common in coeliac disease. In paediatric patients untreated CD gives rise to e.g. impaired growth, pubertal delay, anaemia with different mineral deficiencies, defects in dental enamel and behavioural disturbances. Many adults have iron-deficiency or other anaemia, lactose intolerance, osteoporosis, bleeding disturbances or aphthous ulcerations. These patients may totally lack classical gastrointestinal symptoms.

Coeliac disease has special risk groups: the first degree relatives of CD patients and individuals with certain physical conditions or diseases. Diseases which are
overrepresented in CD patients or where CD is overrepresented (Table 4) include many HLA-DR3 or -DQ2 associated autoimmune and gastrointestinal disorders as well as problems affecting the nervous system, bones, or reproduction (78-80). Often patients with associated conditions have a symptomless, so called silent form of CD, which is only found in screenings.

The magnitude of screening studies is shown by a large Italian paediatric study which reported that for every diagnosed patient with CD symptoms as many as seven new patients with a silent form can be found (81). Only on a continuous gluten free diet the symptoms of other complications might ease or disappear improving the quality of life (82), the risk for additional disorders such as malignancy or severe liver diseases return to the level of healthy people (83;84) and child patients may catch-up growth (85). The older the person is and hence the longer the time of gluten exposure at the time of diagnosis, the more possibilities he has for the outbreak of additional autoimmune disorders (86).

As screening studies have revealed an enormous number of people with only mild symptoms or a totally symptomless disease the picture of CD has had to shift from classical coeliac enteropathy to an autoimmune syndrome with different manifestations seen in any part of body. However, CD cannot be considered unambiguously as a classical autoimmune disease, as the symptoms and autoantibodies disappear after gluten removal.

Table 4. Coeliac patients have a diverse repertoire of symptoms and conditions and in addition they may have various autoimmune disorders more commonly than the general population.

<table>
<thead>
<tr>
<th>Symptoms and conditions</th>
<th>Autoimmune disorders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abdominal distention</td>
<td>Addison's disease</td>
</tr>
<tr>
<td>Abdominal pain</td>
<td>Alopecia areata</td>
</tr>
<tr>
<td>Arthralgia</td>
<td>Autoimmune thyroiditis</td>
</tr>
<tr>
<td>Ataxia</td>
<td>Juvenile chronic arthritis</td>
</tr>
<tr>
<td>Borborygmia</td>
<td>Primary biliary cirrhosis</td>
</tr>
<tr>
<td>Dental enamel defects</td>
<td>Selective IgA deficiency</td>
</tr>
<tr>
<td>Diarrhea</td>
<td>Sjögren's syndrome</td>
</tr>
<tr>
<td>Down syndrome</td>
<td>Type 1 diabetes</td>
</tr>
<tr>
<td>Epilepsy</td>
<td></td>
</tr>
<tr>
<td>IgA glomerulonephritis</td>
<td></td>
</tr>
<tr>
<td>Loose stool</td>
<td></td>
</tr>
<tr>
<td>Miscarriages and infertility</td>
<td></td>
</tr>
<tr>
<td>Osteopenia</td>
<td></td>
</tr>
<tr>
<td>Osteoporosis</td>
<td></td>
</tr>
<tr>
<td>Peripheral neuropathy</td>
<td></td>
</tr>
<tr>
<td>Small bowel lymphoma or cancer</td>
<td></td>
</tr>
<tr>
<td>Steatorrhea</td>
<td></td>
</tr>
</tbody>
</table>
Diagnosis

For the diagnosis of coeliac disease, it is mandatory to take a small intestinal biopsy. Changes of mucosa lesions are classified into five categories (Fig. 6) according to the stages of villus atrophy, hyperplasia of crypt cells and the occurrence of intraepithelial lymphocytes (87). For the CD diagnosis no less than stage 3 lesion is required. In the past it was the custom to take three biopsies to ensure CD diagnosis: one before treatment, the second during a gluten free diet and the third after a challenge with gluten. The mucosa of the intestine should have shown signs of abnormality in the first and last biopsy. Now, repeated biopsies with gluten challenge are only used with children diagnosed under the age of two or if there are any doubts about the initial diagnosis (88).

Figure 6. Spectrum of consecutive mucosal stages seen in gluten sensitivity according to Marsh, 1992: type 0 also called pre-infiltrative is the pattern of healthy mucosa but it may already present anti-gliadin antibodies, type 1 has infiltration of intraepithelial lymphocytes, type 2 in addition to intraepithelial lymphocytes also crypt hypertrophy, type 3 flat destructive with hypertrophic crypts, lamina propria swelling and a flattened surface epithelium and type 4 is a totally atrophic mucosa with possibilities of lymphomas.

In addition to the clinical picture and mucosal lesions detectable in biopsies, patients have specific antibodies well-suited for diagnostics. According to the latest revised ESPGAN (European Society for Paediatric Gastroenterology and Nutrition) criteria ‘the finding of circulating antibodies at time of diagnosis and their disappearance when the patient is taking a gluten free diet add weight to the diagnosis’ (88). These antibody tests were first based on serum antigliadin and antireticulin antibodies, but now more sensitive and specific antiendomysial and tissue-transglutaminase antibody tests are displacing them (89-92).

Prevalence

Susceptibility to coeliac disease may be pictured as an iceberg (Fig. 7). The iceberg consists of people with a basic genetic susceptibility for CD, but of whom the majority will never be affected. Some people, so called latent cases, have coeliac antibody pattern and an increased number of intraepithelial γδ T cells but no signs of overt mucosal lesions (93). Patients with the silent form are also symptomless but already have coeliac manifestation in the intestine as well as detectable antibodies. The small visible part of the iceberg consists of patients with clinical CD (94).
The technical and economical possibilities to screen risk groups or sub populations with antibody tests as well as the wider knowledge of the symptoms among clinicians has revealed more of the coeliac iceberg. Previously coeliac disease was estimated to affect 1:1000 Caucasian based on the number of people with clinical manifestations, but now the screening studies have obtained prevalences between 1:300 to 1:100 (95-97). So far the highest population prevalence reported has come from Algeria where the frequency of the antiendomysial antibody was detected in 5.6% of Saharawi children (98). Prevalences of this world-wide disorder however vary between populations due to differences e.g. in dietary patterns, infant feeding habits and susceptibility gene frequencies.

Dermatitis herpetiformis

One of the diseases which is strongly linked to CD is dermatitis herpetiformis (DH), previously called Durhing’s disease (99-101). In Finland it is already accepted that DH is only one manifestation type of gluten sensitivity and hence patients with DH can be considered as coeliac patients. In this thesis too, the entity for CD covers both the intestinal as well as dermatological manifestation type, the classical CD (cCD) refers only to the intestinal form.
DH is classified as a dermatological disorder where patients have itchy blisters located symmetrically in the skin of elbows, knees and buttocks, sometimes also in the scalp and face. 10% of the patients complain about intestinal problems similar to classical CD patients. In microscopical examinations granular IgA deposits seen in healthy papillary dermis are the hallmark of DH, but all patients also have cCD antibody repertoire, an increased number of intraepithelial lymphocytes in intestinal mucosa, and 75% have mucosal changes revealed by an intestinal biopsy (102;103).

In addition to similar symptoms, antibodies and mucosal morphology seen in DH and cCD there is also the same, and the only, permanent treatment, gluten free diet, for both manifestations. Adherence to the diet alleviates the skin and intestinal symptoms and also lowers risk for the outburst of other associated complications in DH patients (104-106). Similar genetic factors also influence the onset of the two manifestations of gluten sensitivity. Nearly every fifth of first degree relatives has either the skin or intestinal form of gluten sensitivity independently of the manifestation type of the family’s proband (107). Classical CD and DH have also high, over 80% incidence among monozygotic twins and the twins may even be discordant for the type of symptoms caused by gluten sensitivity (53;108-111).

Despite many similarities there are also notable differences. DH is a rarer disorder with a prevalence of 1:1500 compared with intestinal CD which is about ten times more common in Finland (112). The distributions of sex and the age of onset are also different as the intestinal manifestation is seen predominantly in females of all age groups, whereas DH occurs typically in men of the age 30-40 (106). The symptoms on the skin react more slowly to gluten withdrawal than intestinal mucosa; it can take weeks to months after starting a gluten free diet before DH patients may decrease the drug intake for irritating skin symptoms. On gluten challenge the dermatological symptoms appear after approximately 1,5 months and IgA deposits on the skin stay visible for a long time; they can be detected as long as 13 years after starting the proper diet (113;114). For patients with only intestinal symptoms the intestinal mucosa heals completely within a couple of years (115). If recovery takes a longer time it can be considered as refractory coeliac sprue with a poor prognosis (116).

Pathogenesis

The pathogenic mechanisms causing coeliac disease have been unveiled slowly and are still in many parts unsolved. However CD could be thought to be a comparatively easy disorder to study as the main environmental risk factor, gluten, is known, particular HLA-DQ genes are shown to play an important role, there is no lack of patients with confirmed diagnosis and the affected organ is relatively easy to study even in vitro.

The reaction cascade in the small intestine of a patient starts as gluten peptides transpass the intestinal epithelia of villus and enters the lamina propria. There gluten antigens are recognised and caught by antigen presenting cells. As the CD specific antigen presenting HLA molecules, DQ2 and DQ8, preferentially bind peptides with negatively charged amino acids, the positive glutamine residues of gluten
antigens are deaminated to glutamic acid. This deamination is catalysed by an enzyme tissue transglutaminase (tTG) (117;118). Tissue transglutaminase can also cross-link gluten residues to itself to produce antigen-tTG-complexes (119).

The antigen is presented to CD4+ T cells which after activation start to differentiate and proliferate into Th1 and Th2 types (Fig. 8). Th1 cells producing IFNγ and TNFα activate cell-mediated immune regulation and cytokine cascade causing epithelial proliferation and crypt cell hyperplasia (120). Th2 cells responsible for humoral response activate B cells to produce antibodies towards tTG and gluten. The blockage of tTG by autoantibodies can have a broad effect related to epithelial destruction and lesion formation. (121-123)

![Figure 8](image)

**Figure 8.** Gluten intake triggers the pathogenesis in coeliac patients. Antigen presenting cells (APC) present gluten peptides and gluten-transglutaminase derived peptides in their HLA molecules to T cell receptors. This induces T cell development either to Th1 type, cell-mediated immunoregulation with activated cytotoxic T lymphocytes (CTL) and production of various cytokines, or Th2 type, triggering B cells to produce antibodies to gluten and transglutaminase. Antibodies, cytokines and cytotoxic T cells cause villus atrophy, hyperplasia of crypt cells and intestinal lesion formation.

The result is a mass of apoptotic epithelial cells, the destruction of the intestinal villus and crypt cells proliferation typically seen in small bowel biopsies of coeliac patients (see Fig. 6). As the permeability of mucosal lesions is increased, gluten antigens may freely enter and cause more damage. Lesions also have an abnormal number of γδ intraepithelial lymphocytes which still have an unknown role, but may take part in lesion formation.

Patients with a dermatological manifestation of coeliac disease, dermatitis herpetiformis, also have tissue transglutaminase as an autoantigen but in addition
antibodies against epidermal transglutaminase (eTG) (124). This autoantigen catalysing posttranslational modifications shows conservation in enzymatically relevant domains with tTG and the same antibodies can cross-react with them. Sardy et al (2002) (124) hypothesised that DH patients first develop silent CD as a result of low-affinity anti-tTG antibodies, but the continuing gluten provocation would raise cross-reactive antibodies against eTG, too. Antibody-eTG complexes could later provoke the visible dermatological symptoms.

**Genetics**

Coeliac disease has a clear genetic component and there is currently increasing evidence that it can be regarded as a multifactorial disease where both environmental and various genetic factors determine the overall susceptibility. The genetic predisposition has been obvious although until to recently no systematic and thorough twin studies had been published; the earlier studies suggested a concordance of 50-80% (125). Also, the prevalences of 10-20% for first degree relatives of probands have been reported in various family studies (107;126;127). The relative risk for siblings (l\(_s\)) to become affected is about 30 (128) compared to the general population. Finally, also speaking for the genetic component is the fact that CD shows strong association with certain HLA alleles.

**HLA class II DQ2 and DQ8**

Already thirty years ago CD was noted to be associated with the HLA region, first with class I alleles, HLA-A\(_1\) and -B\(_8\) (129;130). Later when HLA typing techniques advanced it was observed that the primary association might be with HLA class II alleles DR3 (131;132), DR5 and DR7 (133;134), and finally with DQ2 alleles (135). The association of HLA-DQ2 heterodimer with CD, which is now believed to be genuine, was reported about fifteen years ago (136).

Tosi et al (1983) (135) using serological data and finally Sollid et al (1989) (136) using DNA techniques, established that a particular HLA-DQ2 heterodimer confers primary risk to CD (Fig. 9). This DQ2 heterodimer is encoded by the \(DQA1\)*05 and \(DQB1\)*02 alleles. It is often called DQ2 but it should be noted that DQ2 refers to any polymorphic allele combination equivalent to serological DQ2 specificity and this nomenclature may cause confusion. These \(DQA1\)*05 and \(DQB1\)*02 alleles can be located either in a single haplotype, in *cis*, with \(DRB1\)*03 (DR3) or in different haplotypes, in *trans*, with \(DRB1\)*11 or *12 (DR5), occasionally also with \(DRB1\)*13 (DR6), and \(DRB1\)*07 (DR7). In Northern Europe, about 90% of coeliac patients carry the DR3;DQ2 haplotype, but in southern Europe, the *trans* combination, i.e. DR5(DR6)/DR7 heterozygotes, is more commonly present in patients, most probably as a result of different frequencies of the haplotypes in the general populations (133;137-140). The DQ2 heterodimers in DR3 and DR5(DR6)/DR7 cases are not totally identical, however (Fig. 9); with DR3 the heterodimer is encoded by the \(DQA1\)*0501 and \(DQB1\)*0201 alleles, but in *trans* by \(DQA1\)*0505 and \(DQB1\)*0202 alleles. The differences in DQ\(\alpha\) chains by one residue in the leader peptide and in DQ\(\beta\) chains by one amino acid residue in the membrane proximal domain (141) are assumed to be of no functional significance.
Determination of HLA-B alleles in CD patients have revealed that the haplotypes in coeliac patients are often the well known ancestral haplotypes B8;DR3;DQ2 or B18;DR3;DQ2, where the only common region is the DR3;DQ2 part (142;143). While sequencing DQA1*05 and DQB1*02 alleles of CD patients (144) and DH patients no (145) particular ‘CD-mutations’ were detected. Also DQ2 segregation studies within families revealed no skewing (146). As a conclusion there is relatively strong genetic evidence that the frequent HLA-DQ2 is the major predisposing factor for CD.

Further evidence for the primary role of DQ2 has come from functional studies. T cells, which are specific for gliadin (soluble part of gluten of wheat), were isolated from the intestinal mucosa of CD patients. They were restricted by HLA-DQ2, that is, they were activated if the antigen was presented by DQ2 molecules encoded by DQA1*05 and DQB1*02 alleles (147).

As the common DQ2 molecule seen in about every fifth Caucasian proved to be the major risk factor, its dose effect has also been studied. It seems that a higher number of patients are heterozygous for DR3;DQ2/DR7;DQ2 (148) or homozygous for DR3;DQ2 (140) than expected from the general population frequencies. Thus a double dose of the DQB1*02 allele may increase the relative risk for the onset and influence the manifestation type of disease or the age of onset (149-152). Recently Vader et al (2002) (153) published that also the DQ2 molecule encoded by DQA1*02 and DQB1*02 alleles can present gluten antigens partly explaining the over-representation of DR3;DQ2/DR7;DQ2 patients.

The HLA association in HLA-DQ2 negative patients has been a matter of some debate recently. Patients negative for HLA-DQ2 usually seem to carry the DR4;DQ8 haplotype encoded by DRB1*04; DQA1*0301; DQB1*0302. The frequency of DR4;DQ8 positive patients represents only a minor part (5-12%) of patients (154;155).
with an interesting exception of Chilean patients who have a higher frequency of DQ8 than DQ2 (156), again possibly reflecting the general population frequencies.

The actual susceptibility gene in DQ2 negatives has been under debate with possibilities of the DQ8 molecule, different DRB1*04 alleles or the DRB4 gene, which is expressed only in DRB1*04, DRB1*07 and DRB1*09 positive haplotypes. A major claim has been that if the predisposing locus was the DRB4 gene, it would explain the association with DR4;DQ8 haplotypes as well as the high risk effect of DR3;DQ2/DR7;DQ2 combination. In addition, the DRB4 molecule (DR53) has also been showed to bind gliadin-derived peptides (157), possibly explaining some earlier reports of gliadin-specific T cells observed to be restricted by DR7 and DR53 (158) haplotypes which carry the DRB4 gene. Speaking strongly against the role of DRB4 are the facts that there is no excess of DR4;DQ7 or DR9;DQ9 haplotypes in CD patients, nor a different distribution of DRB4 alleles between patients and controls (159) and not all DQ2 negative patients have the DRB4 gene (160).

There is no systematic data of the effect of different DRB1*04 alleles in CD susceptibility. Different DRB1*04 alleles are known to be either predisposing or protective to type 1 diabetes (42) and hence, they may modify the effect of primary DQ association in this disease. However the most realistic option for the predisposing role is left to DQ8 as it can bind gliadin (161;162).

Search for other genes

Both HLA-DQ2 and -DQ8 are common in all European populations with a combined frequency of 20-30% (163;164). It has been estimated that the HLA locus contributes 40% at most of the familial risk to CD (128;165). Also a moderate concordance of 30-50% in HLA identical siblings (166) and a probandwise concordance of 20% in dizygotic twins (53) suggest the existence of other predisposing genetic factors.

Genome wide scans with nuclear families or sib pairs have revealed numerous chromosomal regions possibly predisposing to coeliac disease (167-173), but follow up studies have seldom confirmed the original results (174;175). The only indisputable region is still HLA, whereas the others may be either minor, population specific loci or false positive results. All are common problems in genetic analyses in multifactorial diseases. Studies focusing on genes selected by proven or assumed functional evidence have had similar difficulties; different results in different populations are reality. The HLA region contains numerous interesting immunorelated genes, but their association with or linkage to coeliac disease has usually been attributed to linkage disequilibrium within the region or the results have not been replicated in other studies (see Results and Discussion, pages 45-46). Candidate genes outside of the HLA region now considered as the most promising ones are a cluster on chromosome 2q33, including genes for the T cell costimulatory molecules CTLA4 (cytotoxic T lymphocyte-associated 4), CD28 (176-178), and ICOS (inducible co-stimulator) (179). No indication of linkage or association has been found for other interesting candidate genes, such as T-cell receptors (180), tissue transglutaminase genes (181), B7 genes which encode ligands for CD28 and CTLA4 (182), or the IL12B gene, modifying expression of IL-12 (183).
The *IL12B* gene is located in the region 5q31-33, which is considered one of the most interesting regions detected from genome scans. The 5q31-33 region includes a cytokine gene cluster for IL-3, -4, -5, -9 and IL-13. Another interesting region, 11q22 towards telomere, harbours the *IL-10 receptor* gene, genes encoding three CD3 complex chains and genes for matrix metalloproteinase-1 and -3, but at least matrix metalloproteinase-1 and -3 promoters are not shown to associate with coeliac disease (184).

As the population specific linkage studies have very conflicting results, new methods are developed to combine the data already available. In meta analysis the results of genome scans are pooled together whereas in mega analysis the results are divided into chromosomal regions and these are ranked in ascending order before analysing the combined data. With a huge amount of information the effect of population specificities and typing errors are reduced and the results can be considered more accurate than in individual genome scans. The first combined studies have already been accepted by European Journal of Human Immunology (Meta-analysis by Babron M-C, Clerget-Darpoux R, Nilsson S, Ascher H, Ciclitira PJ, Sollid LM, Partanen J, Greco L and the European Genetics Cluster on Coeliac Disease; Mega-analysis by Nilsson S, Adamovic S, Naluai ÅT, Babron M-C, Wahlström J, Ascher H, Ciclitira PJ, Clerget-Darpoux R, Sollid LM, Partanen J, Greco L and the European Genetics Cluster on Coeliac Disease).

**Finnish studies on coeliac disease**

Coeliac disease has been intensively studied in Finland and the first doctoral thesis about ‘Duodeno-jejunal histology in the malabsorption syndrome in infants’ was already published in the year 1966 by Pekka Kuitunen (185). Since then there has been 31 thesis from Helsinki, Tampere, Turku, Oulu and Kuopio Universities (Table 5). The subjects of these thesis have covered various aspects of coeliac research; epidemiology widening the disease spectrum from strictly classical coeliac disease to silent and latent manifestations and to dermatological form, pathogenesis from cell cultures or biopsy studies to different antibodies, subsets of T cells and adenovirus infections and genetics with research on HLA and HLA-unlinked genes. Also CD associated disorders or manifestations seen e.g. in teeth, have been topics of research as well as nutritional aspects of coeliac patients.

**Table 5.** Doctoral thesis about coeliac disease published in Finland since 1966.

<table>
<thead>
<tr>
<th>Author (year)</th>
<th>Title of thesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pekka Kuitunen (1966)</td>
<td>Duodeno-jejunal histology in the malabsorption syndrome in infants</td>
</tr>
<tr>
<td>Pirkko Immonen (1968)</td>
<td>Levels of the serum immunoglobulins γA, γG and γM in the malabsorption syndrome in children</td>
</tr>
<tr>
<td>Erkki Savilahti (1973)</td>
<td>Suoliston immunoglobulinint larsuusilassä</td>
</tr>
<tr>
<td>Timo Reunala (1977)</td>
<td>Dermatitis herpetiformis; genetic aspects and gluten-free diet treatment</td>
</tr>
<tr>
<td>Matti Vuoristo (1979)</td>
<td>Cholesterol metabolism in adult coeliac disease</td>
</tr>
<tr>
<td>Jorma Kokkonen (1980)</td>
<td>Lasten mahalaukun toiminnalliset häiriöt; tutkimus ohutsuoliperäistä malabsorbiota, atopiaia, diabetesta ja Sjögrenin syndromaa</td>
</tr>
<tr>
<td></td>
<td>sairastavien lasten mahalaukon rittyistoiminnasta ja morfoloja</td>
</tr>
<tr>
<td>Author</td>
<td>Title</td>
</tr>
<tr>
<td>------------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>Aira Laihinen (1986)</td>
<td>Kuutaja ja raapiminen ihosairauksien kroonistajina: psykosomaattinen tutkimus</td>
</tr>
<tr>
<td>Eeva Vainio (1986)</td>
<td>Antigliadin antibodies in dermatitis herpetiformis and gluten-sensitive enteropathy</td>
</tr>
<tr>
<td>Liisa Aine (1986)</td>
<td>Dental enamel defects and dental maturity in children and adolescents with coeliac disease</td>
</tr>
<tr>
<td>Olavi Häälström (1989)</td>
<td>IgA-class reticulin antibody test in coeliac disease and dermatitis herpetiformis</td>
</tr>
<tr>
<td>Timo Klemola (1989)</td>
<td>Immunoglobulini A:n puutos lapsipotilailla</td>
</tr>
<tr>
<td>Kati Holm (1993)</td>
<td>Latent and silent coeliac disease</td>
</tr>
<tr>
<td>Pekka Collin (1994)</td>
<td>Associated diseases and survival in coeliac disease</td>
</tr>
<tr>
<td>Mikael Kuitunen (1995)</td>
<td>Intestinal permeability to human α-lactalbumin, bovine β-lactoglobulin, mannitol and lactulose in healthy infants and coeliac disease patients</td>
</tr>
<tr>
<td>Tero Saukkonen (1996)</td>
<td>Immune response to dietary proteins in children with insulin-dependent diabetes mellitus</td>
</tr>
<tr>
<td>Kristiina Airola (1997)</td>
<td>Matrix metalloproteinases in human skin: expression of interstitial collagenase, stromelysin-1, collagenase-3, and their tissue inhibitors by keratinocytes</td>
</tr>
<tr>
<td>Tarja Kemppainen (1997)</td>
<td>Oat meal as a component of a gluten-free diet, nutrient intakes, nutritional status and osteopenia in coeliac patients</td>
</tr>
<tr>
<td>Anne Polvi (1998)</td>
<td>Genetic susceptibility to coeliac disease: major histocompatibility complex genes in coeliac disease and in its animal model</td>
</tr>
<tr>
<td>Hannu Lähteenoja (1999)</td>
<td>Coeliac disease and oral mucosa</td>
</tr>
<tr>
<td>Sari Ilitanen (1999)</td>
<td>Markers of coeliac disease latency</td>
</tr>
<tr>
<td>Tuula Sontag-Strohm (1999)</td>
<td>Gluten protein composition and quality in bread wheats</td>
</tr>
<tr>
<td>Katri Kaukinen (2000)</td>
<td>Small bowel mucosa in suspected and confirmed coeliac disease</td>
</tr>
<tr>
<td>Satu Suikanen (2000)</td>
<td>Serum antibodies in coeliac disease</td>
</tr>
<tr>
<td>Tommi Leivo (2000)</td>
<td>Basement membrane zone proteins, epithelial integrins and TGF-β system in reepithelialization, dermatitis herpetiformis and psoriasis; modulation by isotretinoin, betamethasone and calcipotriol</td>
</tr>
<tr>
<td>Päivi Holopainen (2002)</td>
<td>Genetic susceptibility to coeliac disease: HLA-unlinked candidate genes</td>
</tr>
<tr>
<td>Timo Örmälä (2002)</td>
<td>Age related changes in the immunology of the intestinal mucosa</td>
</tr>
<tr>
<td>Tuula Halttunen (2002)</td>
<td>Biological functions of coeliac disease antibodies</td>
</tr>
</tbody>
</table>
AIMS OF THE STUDIES

The purpose of this thesis has been

to scrutinise the effect of the HLA gene region on gluten sensitivity, in particular to search novel susceptibility genes in addition to the known HLA-DQ alleles; and
to evaluate the role of the HLA gene region in the different clinical manifestations of gluten sensitivity.
STUDY SUBJECTS AND ETHICAL PERMISSIONS

COELIAC DISEASE PATIENTS

Multiplex families (I,III,IV,V)

Samples from Finnish families with at least two siblings with gluten sensitivity have been collected through advertising in the Finnish Coeliac Society’s newsletter. The collection and the further screening procedures within the families have been described by Mustalahti et al (2002)(186). Briefly, the diagnoses were confirmed by scrutinising hospital records and all voluntary family members were screened for the diagnostic anti-gliadin and anti-endomysial antibodies. A small bowel biopsy was performed after a positive antibody test result. Altogether, 137 families (872 individuals) were collected and their samples were prepared for analysis. Of the 137 families, 110 had at least two affected siblings and they were the principal study set of the present studies.

Every patient has given informed consent and the study has been accepted by the Ethical Committee of Tampere University Hospital with permission number 95173.

In study I, 52 families with microsatellite markers typed on chromosome 5q31 were included. In study III, one sib pair of each 110 families were studied. In study IV, nine families with HLA-DQ2 homozygous parents were included. In study V, one hundred families were screened for DQ2 negativity. These 100 families were those included in the EU Consortium studies.

Simplex families (IV)

An earlier collection of Finnish CD family material consisted of 31 families (131 individuals) with at least one gluten sensitive patient (187). Five families overlapped with the multiplex family material described above (186). All diagnosis were based on ESPGAN criteria (88).

Four families with HLA-DQ2 homozygous parents were included in study IV. The approval to study coeliac disease in these families came from the Ethical Committee of the Tampere University Hospital.

Monozygotic twin pairs (II)

From a total of 1292 dermatitis herpetiformis patients identified during 1969 – 1999 in Finnish dermatological clinics, 6 patients were reported to have a monozygotic twin. The diagnosis of DH were based on the presence of a granular IgA deposit in the skin, and cCD on the demonstration of subtotal or partial villous atrophy in the small intestine and histological response on the gluten free diet treatment. All twin pairs were re-examined in 1999 for adherence to gluten-free diet, presence of the rash or gastrointestinal symptoms and IgA antibodies to gliadin and endomysium.

The study has permission (number 27.10.98/98181) from the Ethical Committee of Tampere University Hospital and every twin gave an informed consent in 1999.
Hungarian coeliac disease patients (III)
The study included 139 unrelated Hungarian patients with either the skin (71 DH patients) or the intestinal (68 cCD patients) form of gluten sensitivity who all had biopsy-confirmed diagnosis. DH patients have been diagnosed by the presence of a blistering rash and granular IgA deposits in the skin. They also underwent jejunal biopsy at diagnosis and were investigated for serum endomysial antibodies. Coeliac patients were diagnosed by jejunal biopsy according to the criteria of the ESPGAN (88) and all were endomysial antibody positive. Samples were collected after discussion with the patients/parents and with the approval of the Ethical Committee of the Heim Pal Children’s Hospital, Budapest.

Coeliac disease patients from European Genetics Cluster (V)
Nine-hundred-and-eight families with gluten sensitive patients were recruited from France, Italy, Norway, Sweden and the UK for a study of DQ2 negative patients. The net result including the Finnish material of one hundred families was altogether 1008 families with 121 DQ2 negative index patients. The foreign patients fulfilled the revised ESPGAN criteria (88) for the diagnosis and material obtained an approval from the regional ethics committees for each participating centre and written informed consent from every participating subject.

SAMPLES IN STUDY I
HLA region-linked microsatellite markers were studied in 131 chromosomes from 49 Finnish families, in 51 patients with congenital adrenal hyperplasia due to 21-hydroxylase deficiency, in 13 Finnish patients waiting for an unrelated bone marrow transplant from a Finnish donor, in 8 Finnish patients waiting for an unrelated bone marrow transplant from a non-Finnish donor, in 58 Finnish bone marrow donors, and in 32 HLA reference cell lines.
METHODS

DNA EXTRACTION

Genomic DNA was extracted from whole blood by either the salting out procedure (188) or by commercial kits (Pel-Freez Clinical Systems, Brown Deer, WI or Qiagen, Hilden).

MICROSATELLITE MARKERS

Altogether nineteen microsatellite markers linked to the HLA region were typed. These markers D6S276, D6S2239, D6S2225, D6S2223, D6S105, MOGc, D6S265, C3_2.11, C2_2.4, MIB, MICA, TNFb, TNFa, D6S273, NOTCH4, DQCARII, DQCAR, G51152, D6S291, are commonly used tools for HLA typing (66;69;189-191). Most of the markers were located within the 3 Mb HLA region, but D6S291 was ~2 Mb centromeric from HLA-DPB1; D6S2223, D6S2225 and D6S2239 in the HFE gene region, D6S105 ~2 Mb telomeric from the HLA-A and D6S276 ~6 Mb telomeric from the HLA-A (Fig. 10). Variable combinations of microsatellite markers were used in different studies.

Figure 10. A map indicating the locations of nineteen typed microsatellite markers and their relation to classical HLA loci. Most microsatellites are located within the 3 Mb segment in the HLA region, but D6S291 is located ~2.5 Mb centromeric from DQB1, and D6S105 ~2 Mb telomeric from HLA-A, D6S2223, D6S2225 and D6S2239 around the HFE region ~4.6 Mb from HLA-A, and D6S276 ~6 Mb telomeric from HLA-A. The map is not on scale.

In addition, eight microsatellite markers from different chromosomes (D1S1589, D9S158, D10S1213, D14S617, D15S642, D16S403, D17S1293 and D18S851) were used for verification of twins genuine monozygosity in study II.

The standard amplification procedures were applied and the length of the PCR products were estimated by single capillary electrophoresis ABI PRISM™310 Genetic Analyzer (Applied Biosystems, Foster City, CA). The TAMRA-labelled size standard GeneScan-500 TAMRA or -2500 TAMRA (Applied Biosystems, Foster City, CA) was simultaneously run with each sample to get allele size estimates. The microsatellite alleles were given as arbitrary migration units and the standardisation between the
studies was archived by re-typing the same patients, as well as using national and international CEPH-controls.

DETERMINATION OF HLA ALLELES (‘HLA TYPING’)

Commercial kits used for DQB1 typings were Dynal AllSet™ SSP DQ “low resolution” (Dynal, Oslo) and LiPA-DQB1 kit (Innogenetics, Ghent). Two microsatellite markers (DQCAR, DQCARII) in Finnish and three (DQCAR, DQCARII, G51152) (189) in Hungarian samples were employed to deduce the DR;DQ haplotypes. More detailed typing of HLA-A, -B, -DRB1-5, -DQA1, -DQB1 and -DPB1 alleles was done by commercial kits by Innogenetics, Ghent; Dynal, Oslo; Olerup SSP, Saltsjöbaden and PEL-FREEZ Clinical Systems, Brown Deer, WI.

Other typing methods utilised in the Finnish simplex families with gluten sensitive patients have been described by Polvi et al (1996)(187), with 21-hydroxylase deficiency family material described by Levo et al (1997)(192) and with patients from France, Italy, Norway, Sweden and the UK by Karell et al (2003)(V).

HAPLOTYPING AND STATISTICAL ANALYSIS

Haplotypes were constructed from individual genotypes manually and by the Genehunter 2.0 program (193). The Genehunter program was also used for Mendel checking and for indicating possible recombinations.

The linkage disequilibrium between HLA alleles and microsatellite markers (study I) was estimated from Finnish haplotypes using the Arlequin program package version 1.1 by Schneider et al (1997)(194) http://anthropologie.unige.ch/arlequin. The test is an extension of the Fisher exact test. Instead of enumerating all possible contingency tables, a Markov chain is used to evaluate the probability. For the actual test the chain length was set at 100 000 and the dememorization at 1000. Evidence for the linkage disequilibrium was regarded as strong if the exact probability obtained was p < 0.0001, weak but possible if 0.0001 < p < 0.01 and no linkage disequilibrium was assumed if p > 0.01.

Estimation of probability that the assumed twins genuinely were monozygotic rather than dizygotic ones (study II), was performed by calculating the number of microsatellites used for identification by probability of 0.25 i.e. the sib pair is identical by change. The casewise concordance rate, Cc, was estimated according to the formula Cc=2C/(2C+D), where C stands for the number of pairs concordant and D for discordant.

The Fisher exact two-tailed test was applied when comparing frequencies of alleles either transmitted or never transmitted to affected sibs from DR3;DQ2 homozygous parents (study IV). It was also used when comparing haplotype frequencies between Finnish and Hungarian gluten sensitive patients (study III).

A simple binomial test was used to compare haplotype frequencies in DQ2 negative patients (study V). A p-value <0.05 was considered significant.
RESULTS AND DISCUSSION

MICROSATELLITE MARKERS AS A PRACTICAL TOOL FOR HLA TYPING (I)

The determination of HLA alleles, often called ‘HLA typing’, by using either commercial kits or standardised assays, is a relatively easy and reliable method now used world-wide in clinical diagnostics in tissue typing laboratories. However, in scientific research projects the emphasis is often directed to more economical, faster and less DNA demanding methods, at least for an initial screening for possible HLA association. Lin et al (1997) (189) demonstrated strong linkage disequilibrium between HLA class II-linked microsatellite markers, DQCAR, DQCARII, and G51152, and the HLA DR-DQ genotypes in the Japanese, Norwegians and New Guineas. We examined if the same phenomenon was seen in the Finnish population. As the genetic background of the Finns is assumed to be relatively narrow, we speculated that microsatellite typing could be used as an alternative method in the deduction of at least the common, or major, HLA haplotypes in this population.

In addition to the three microsatellite markers studied by Lin et al (1997) (189), we tested nine more markers (D6S276, MOGc, D6S265, MIB, TNFb, TNFa, D6S273, NOTCH4, D6S291) distributed over the entire HLA region and even some of the outskirts (Fig. 10). Most of the markers were found to be highly polymorphic when tested in 131 unrelated chromosomes. Therefore they could be regarded as informative genetic markers. Nearly all loci, except for the most distal D6S276 and D6S291, showed evidence of strong linkage disequilibrium with the adjacent loci (Study I, Fig. 2). In addition many microsatellite marker alleles appeared to be associated with only certain HLA alleles (Study I, Fig. 3), indicating that they can be used to deduce many HLA haplotypes.

To compare linkage disequilibrium between Finnish and non-Finnish chromosomes, we estimated the degree of genetic variation in microsatellites associated with HLA-A, -B, -DRB1-5 and -DQB1 matched haplotypes in 21 Finnish patients waiting for a bone marrow transplant and their potential unrelated Finnish (n = 44; range 1 - 6 per patient) and non-Finnish donors (n = 14, range: 1 - 3 per patient). The similarity between Finnish haplotypes was striking after excluding the outermost markers (D6S276 and D6S291) as 48% (21/44) of the patient - donor pairs were identical. The conservation of the microsatellites was highest near the DR-DQ genes where there was practically no variation at all (1 mismatch in 88 chromosomes). In Finnish patient - non-Finnish donor pairs the variation was substantially higher as only one pair (7%) was identical after excluding the outermost markers, but the linkage disequilibrium was still notably strong within the DR-DQ region compared to other regions.

We concluded that the three markers located in the DR-DQ region, DQCAR, DQCARI1 and G51152, can be used to deduce many common DR;DQ haplotypes in the Finnish population (Table 6). This method has been later applied in additional research projects for defining certain HLA-DRB1 alleles and also occasionally for whole DR;DQ haplotypes (195). It also seems likely that many HLA-A and -B alleles can be deduced from the microsatellite data but the large variation in alleles and lower linkage disequilibrium may sometimes hamper the interpretation.
RESULTS AND DISCUSSION

Microsatellite typing is a relatively easy procedure with a simple PCR (polymerase chain reaction) run and measuring the length of the PCR amplicon is now possible to perform using highly automated apparatuses.

However it is unlikely that the actual HLA typing would be replaced by the microsatellite method, since the HLA typing is a widely standardised and accredited method used by co-operating international centres. It is also important to remember that the HLA molecules as such are the crucial determinants for transplantation matching. In principle, microsatellites could be applied for routine HLA typing as extra markers or as the first screen for HLA identical siblings, but this would require more detailed validation of the method. Also, due to a higher mutation rate in microsatellite loci (196), their allelic differences may not always indicate a genuine difference in the HLA genes.

SAME HLA HAPLOTYPES CONFER GENETIC SUSCEPTIBILITY TO BOTH INTESTINAL AND DERMATOLOGICAL MANIFESTATION OF GLUTEN SENSITIVITY (II, III)

The two main manifestations of gluten sensitivity are the intestinal form, here called classical coeliac disease (cCD), and the dermatological form, dermatitis herpetiformis (DH). DH patients have often been included in studies of coeliac disease (170;197) in particular when coeliac disease is defined more broadly, synonymous to gluten sensitivity, as is also done in the present thesis. This is explained by similarities in environmental and genetic risk factors as well as in symptoms (106). However there are some differences in opinion whether DH patients should be included or not.

As the HLA-DQ is known to be the major genetic locus for susceptibility to gluten sensitivity, we wanted to evaluate whether different clinical forms are determined by differences in the HLA loci. First, concordance in monozygotic twins was estimated to evaluate the relative contributions of genetic and environmental factors. Six pairs

### Table 6

<table>
<thead>
<tr>
<th>HAPLOTYPE</th>
<th>DQCAR</th>
<th>DQCARII</th>
<th>G51152</th>
</tr>
</thead>
<tbody>
<tr>
<td>DR3;DQ2</td>
<td>92</td>
<td>201/197/195</td>
<td>213/211</td>
</tr>
<tr>
<td>DR7;DQ2</td>
<td>106/114/112/110/108/92</td>
<td>215/217/213/211/209</td>
<td>211</td>
</tr>
<tr>
<td>DR1;DQ5</td>
<td>96</td>
<td>195/199</td>
<td>220/236</td>
</tr>
<tr>
<td>DR2;DQ5</td>
<td>96</td>
<td>199</td>
<td>211</td>
</tr>
<tr>
<td>DR6;DQ5</td>
<td>100</td>
<td>197</td>
<td>215</td>
</tr>
<tr>
<td>DR2;DQ6</td>
<td>96</td>
<td>197</td>
<td>220</td>
</tr>
<tr>
<td>DR6;DQ6</td>
<td>96</td>
<td>205/207</td>
<td>220/217/215</td>
</tr>
<tr>
<td>DR5;DQ7</td>
<td>110/112/114/116</td>
<td>183/185/191</td>
<td>211</td>
</tr>
<tr>
<td>DR4;DQ8</td>
<td>104</td>
<td>191/197</td>
<td>220/222</td>
</tr>
<tr>
<td>DR8;DQ4</td>
<td>92</td>
<td>195</td>
<td>192</td>
</tr>
</tbody>
</table>
of monozygotic twins were found amongst 1292 probands with DH (II). Second, a study on 25 sib pairs (III) discordant for the manifestation was done in order to define the sharing of HLA haplotypes between siblings within each family. In addition, 85 concordant sib pairs (81 cCD-cCD and 4 DH-DH) were included for comparison. We wanted to focus on the haplotype distribution especially in the families, since the other factors, environmental and genetic, can be assumed to be substantially more homogeneous in each family than in the standard case-control studies.

The basic clinical manifestation and HLA alleles of the six twin pairs are shown in Table 7. Briefly, three pairs (pairs 1, 4, 5) were concordant for the manifestation of DH, two (pairs 2 and 3) only for gluten sensitivity when defined as a more broad entity (i.e. one sibling had DH, the other cCD). Finally, one of the six pairs was discordant for any manifestation, one was affected by DH and cCD whereas the other had no symptoms or any laboratory findings suggesting preclinical forms of this disease. The follow-up period for the twins varied from 3 to 30 years. It is of note that the discordant pair was followed for the shortest time. All the twins had the HLA-DR and -DQ alleles associated with gluten sensitivity; one of the DH-DH concordant pair was homozygous for DR4;DQ8 and the others were positive for DR3:DQ2. Interestingly, the only discordant pair for gluten-sensitivity had no homozygosity in any of the HLA risk loci, unlike all the other pairs. However, otherwise there appeared to be no clear overall association between the HLA alleles and clinical manifestation. Consequently, five of the six pairs were concordant with respect to gluten-sensitivity, giving a casewise concordance rate of 0.91. Then again, the more precise clinical manifestation varied. One of the major problems in twin studies is that there may be positive bias for concordant twins, in particular, when no systematic population-based approach is performed. Nevertheless, our figures can be regarded as reliable, since the casewise concordance rate in a thus-far only population-level screening study by Greco and co-workers (2002)(53) was 0.86, surprisingly similar to ours. These two studies indicate that the concordance rate in gluten sensitivity is very high for a multifactorial disease. In most other multifactorial diseases the rate has been reported to be around 0.25 - 0.46 (Table 3). Thus it can be concluded that the role of genes in gluten sensitivity is crucial - alternatively, it is possible to speculate that the figures can be explained by a very low variation in an environment.

The effect of HLA genes on the manifestation was also studied in families (study III). The distribution of haplotypes inherited identical by state (IBS) was similar between concordant and discordant sib pairs (study III) (Table 8). The distribution of non-shared HLA haplotypes in the 11 discordant pairs, who shared only one haplotype IBS, suggested that no predominant haplotype was enriched in or missing from either group. So, the other haplotype may not modify the manifestation. All patients were found to have the typical predisposing HLA types, the only exception was a single DH patient. It should be noted that originally these families were selected based on having at least two affected siblings. This material may consequently be enriched by various major genetic and environmental susceptibility factors which may mask possible minor effects by other locus within HLA on the clinical outcome.
Table 7. The information of the six monozygotic twin pairs, their gender and birth year, HLA status and clinical data of individual twins (A and B).

<table>
<thead>
<tr>
<th>Twin pair 1</th>
<th>Twin pair 2</th>
<th>Twin pair 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender/birth year</td>
<td>DQ2/DQ8/other</td>
<td>DRB1</td>
</tr>
<tr>
<td>F/51</td>
<td>DQ8/8</td>
<td>0401</td>
</tr>
<tr>
<td>F/65</td>
<td>DQ2/2</td>
<td>0301</td>
</tr>
<tr>
<td>F/58</td>
<td>DQ2/other</td>
<td>0301,1201</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 8. The distribution of the identical by state (IBS) haplotypes did not differ between the discordant (cCD-DH) and the concordant sib pairs (cCD-cCD or DH-DH).

<table>
<thead>
<tr>
<th>cCD-DH</th>
<th>cCD-cCD</th>
<th>DH-DH</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>%</td>
<td>n</td>
</tr>
<tr>
<td>2 IBS</td>
<td>14</td>
<td>56</td>
</tr>
<tr>
<td>1 IBS</td>
<td>11</td>
<td>44</td>
</tr>
<tr>
<td>0 IBS</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

To extend the above results, obtained from the Finnish families, a set of Hungarian patients with cCD or DH was studied. All the Hungarian patients carried at least one of the known predisposing haplotypes: DR3;DQ2, DR4;DQ8, or DQ2 in trans. The distribution of HLA phenotypes was very similar in the cCD and DH groups, thus supporting the family study. However, a statistically significant difference between the DH and cCD groups was noted; the DR3;DQ2/DR5;DQ7 combination was more common in the DH patients than in the cCD patients (26.8% vs. 8.8%, respectively; p = 0.008 by Fisher’s exact test). In other words, DH patients had a double dose of DQA1*05 allele more frequently than the cCD patients.

The double dose of certain HLA alleles may have a stronger effect on DH, but the genuine value of this finding must be left open, since this could also reflect solely the differences in haplotype frequencies between the two study populations. There are obviously many genetic differences between the Finnish and Hungarian populations, although the languages are distantly related. DH
RESULTS AND DISCUSSION

in Hungary is a relatively common disorder and usually seen in paediatric cases, whereas in Finland it is found mainly in adults (106). A peculiar detail is that the one discordant twin pair in study II was the only pair not having a double dose of DQA1*05.

Our results will close down the debate of possible differences in HLA between the intestinal and dermatological manifestations of gluten sensitivity. The haplotypes are not only molecularly similar as suggested by earlier case-control studies (198;199) but they are often indeed identical as could be demonstrated by the present family study. The very high concordance rate in monozygotic twins and HLA similarity even in discordant sib pairs gives the impression of a limited number of genetic factors to gluten sensitivity. However, the few twins found to be discordant clearly demonstrate that environmental factors, in addition to dietary gluten, and small genetic differences between twins still have a detectable effect.

A lot of speculation has been reported about environmental factors influencing the development of coeliac disease. In addition to the primary role of gluten, adenovirus has been suggested as one potential factor. It has segments extremely similar in amino acid sequence to gluten, but the genuine role of adenovirus is not clear (94). Exposure to different antigens leads inevitably to a variable set of T cells which may play a role in the onset of disease. Indeed different sets of T cell receptors were reported in twins even concordant for CD (200). Also customs in breast feeding (201) and the time of introducing solid food to children may have an effect. This was evident in Sweden in the 80’s with a sudden increase in prevalence of paediatric cases with gluten sensitivity when cereals were added into commercial baby food (202). At the same time the age of onset was increased in many other countries (203). Now we are able to witness a similar increase in the number of CD patients among Saharawi refugees, who are fed by cereals instead of their traditional diet (98) indicating that it is not only the timing but also the amount of gluten that affects the onset, maybe by breaking oral tolerance.

EXTENDED B8;DR3;DQ2 HAPLOTYPE PREDISPOSE TO COELIAC DISEASE (IV)

The DR3;DQ2 haplotype is known to predispose to coeliac disease and there is both genetic and functional evidence that the HLA-DQ molecules are the primary susceptibility factors (121). However, although the class II genes are obviously the major associated locus to various autoimmune disease, data have emerged that the other HLA-linked genes may modify their effect on susceptibility (48;204;204;205). Also the overwhelmingly strong evidence in the whole genome screens for HLA and its flanking regions in coeliac disease have stimulated detailed searches of the HLA region in hope of finding novel influencing susceptibility genes. We wanted to address this question by dissecting the DR3;DQ2 haplotypes in CD families and evaluating their relative effects on susceptibility. The homozygous parent TDT-type of approach was used (206) for the first time in coeliac studies by Polvi et al (1997)(207) in which evidence for DPB1*0101 was obtained. In study IV this approach was systematically applied; fourteen DR3;DQ2 homozygous parents were
found and the segregation of the DR3;DQ2 haplotypes, dissected by microsatellite typing, were traced to the affected offspring.

The DR3;DQ2 positive haplotypes transmitted to affected children were very conserved and the majority (15/21, 71%) had alleles HLA-A*01, -B*08 and -DPB1*0101. One third showed conservation even up to loci D6S2239, D6S2225 or D6S2223, which are all located even 4.6 Mb telomeric from HLA-A (Fig. 10). In DR3;DQ2 haplotypes never inherited by the affected children, the conservation was restricted only to the region between G51152 (close to HLA-DQ) and MIB (next to HLA-B). The conservation near DR-DQ genes was expected since the selection was done based on the DR3;DQ2 positivity. Statistically the most significant difference between transmitted and non transmitted haplotypes was the frequency of HLA-A*01 (p = 0.007) but additionally suggestive evidence (p = 0.02 - 0.04) could be found for D6S265*181, C3_2_11*216 and C2_4_4*222 alleles (Fig. 11).

The results for the DR3;DQ2 homozygous parents clearly indicated that there are differences between DR3;DQ2 haplotypes. The majority of the transmitted haplotypes had alleles HLA-A*01, B*08, DRB1*03, DQ2, and often also DPB1*0101, that is, they formed the well-known ancestral haplotype called 8.1 AH. The peculiar 8.1 AH is exceptional in many ways; it is very conserved and associated in addition to coeliac disease with e.g. type 1 diabetes, myastenia gravis and systemic lupus erythematosus, common variable immunodeficiency and IgA deficiency (31). However it is also one of the most common haplotypes in Caucasoids. Hence it is either expected to be relatively young haplotype or to have some survival advantages. Both may be true; the survival advantages have been shown from controversial

Figure 11. The comparison between the DR3;DQ2 haplotypes transmitted (n = 21) and non-transmitted (n = 7) from homozygous parents to coeliac offspring. The most common allele in each locus of the transmitted haplotypes was chosen as the index size and their frequencies were compared. Black line: the frequency in transmitted haplotypes; Dotted line: the frequency in non-transmitted haplotypes. The transmitted haplotypes remained conserved longer than the non-transmitted haplotypes. Statistically the most significant difference was at HLA-A*01 (p = 0.007) and suggestive difference for microsatellite alleles D6S265*181, C3_2_11*216 and C2_4_4*222 (p = 0.02-0.04).
results where 8.1 AH was over-represented in Caucasian nonagenarian males and the haplotype seemed be advantageous in early foetal life (208).

The strong linkage disequilibrium in AH 8.1 may hide the locations of other predisposing factors for CD in addition to DQ2. Our results pinpoint most strongly to the class I region but as the Finnish population is known for its strong founder effect the linkage disequilibrium may cover exceptionally large regions and other risk factor(s) may be situated anywhere within or near HLA. In a recent Italian segregation study of DR3:DQ2 haplotypes the most significant differences between transmitted and non transmitted haplotypes were seen more centromeric than in study IV, in microsatellite markers MIB, MICA and MICB (Bolognese E, Karell K, Percopo S, Coto I, Greco L, Mantovani V, Suoraniemi E, Partanen J, Mustalahti K, Mäki M and Momigliano-Richiardi P, accepted by Tissue Antigens). Those markers are in the vicinity of HLA-B and the highly transmitted alleles are all markers of the ancestral 8.1 haplotype, thus supporting our results. Also a recent multicentre European study indicated weakly that a group of neighbouring loci between D6S273 and MIB possibly may be associated with an increased risk although no single locus with significant evidence was found (Louka AS, Moodie SJ, Karell K, Bolognesi E, Ascher H, Greco L, Momigliano P, Partanen J, Ciclitira PJ, Sollid LM and the European Genetics Cluster on Coeliac Disease, accepted by Human Immunology). Hence, the final proof for the existence as well as the for the location of the putative novel susceptibility gene in HLA require more systematic mapping studies.

Many candidate genes in the HLA region have already been tested for association with CD. However, they are mostly sporadic studies and suffer from the case-control type of approach where selection of proper controls, due to high linkage disequilibrium in HLA, is difficult. The HLA-DP locus has shown to be associated with CD, but obviously due to linkage disequilibrium with DQ, in Finnish (187), Norwegian (209) and French patients (152), whereas no association between DP and CD was seen in Ashkenazi Jews (210). In the USA DPw2 showed a negative association with DH and thus a protecting effect was suggested (211). Also the reported association of TAP1 and TAP2 genes, which are located in the HLA class II region (Fig. 2), seems to be secondary and a consequence of linkage disequilibrium (152;212;213).

More attention has been given to telomeric side with class III and I loci. Next to the TNFα gene, which has a role in immunological regulation and inflammation lies a microsatellite shown to be associated with CD in the Irish population, possibly independently of the A1:B8;DR3:DQ2 haplotype (214). The same allele showed association in Finnish patients but after comparing the association with haplotype-matched controls the difference vanished (215). Also an unequal distribution of heat-shock protein 70 (HSP70-2) alleles seems to result from a strong association with HLA-DR3 and not primarily with CD (216). The A5.1 allele of the MICA gene, has been shown to carry a nucleotide insertion causing it to become a secreted form of the MICA molecule. The predominant expression of MICA molecules on gastrointestinal epithelial makes it an interesting candidate: it could inhibit the activation of γδ T cells or NK cells (217) thus inflicting the onset of CD or at least
causing milder symptoms. The A5.1 allele was claimed to be associated with silent CD independently of the HLA class II alleles (218). Controversially, it has been shown to be a part of the extended DR3;DQ2 haplotype (219;220). Also allele 3 of a microsatellite D6S2223 located telomeric of the HLA class I region (Fig. 10) was found to reduce susceptibility in a recessive manner to both CD and type 1 diabetes (221). The association of D6S2223*3 to some nearby gene raised enthusiasm, but not our study neither the study from the EU Cluster (Louka AS, Moodie SJ, Karell K, Bolognesi E, Ascher H, Greco L, Mornigliano P, Partanen J, Ciclitira PJ, Sollid LM and the European Genetics Cluster on Coeliac Disease, accepted by Human Immunology) have confirmed the first results.

The many genetic studies of coeliac disease have concluded HLA-DQ to be the single confirmed susceptibility factor. The strong linkage disequilibrium complicates studies in the HLA region and hence more effort has been put into other regions and replication studies of potential non-HLA candidate loci. Most probably numerous different loci affect the development of coeliac disease. The affected loci may also vary between populations or manifestation types. To reveal these predisposing loci, subdivision of patients according to stricter clinical phenotypes may be helpful. The basis of subgrouping in CD can be e.g. clinical manifestation type (silent/full/DH), age of onset, or sex. Also, subgrouping according to the HLA status should be of interest. By this way Holopainen et al (2001)(222) found stronger evidence for linkage to chromosomes 5q and 11q in multiplex CD families with at least one DH patient, whereas chromosome 2q was more strongly linked in families with only cCD patients. The role of HLA-DQ2 as a susceptibility factor was suggested to be stronger in females than males. These results call for further studies in which more strictly defined clinical subgroups of CD can be interesting to study specially in non-HLA regions.

ONLY A FEW PATIENTS WITH COELIAC DISEASE LACK THE KNOWN HLA SUSCEPTIBILITY ALLELES. A MULTINATIONAL EUROPEAN STUDY (V)

Multinational European Genetics Cluster on Coeliac Disease enabled the largest ever collection of DQ2 negative coeliac patients. Patients were collected from Finland, France, Italy, Norway, Sweden and the UK. From a total of no less than 1008 families, 121 DQ2 negative patients were found. About half (n = 60) of these carried the DQ8 heterodimer, known to be a predisposing marker in DQ2 negative CD (121), and most of the remaining patients (n = 61) had either the DQB1*02 or DQA1*05 allele which encode for the half of the DQ2 heterodimer. Finally, there were only four (0.4%) patients who did not have any alleles known to predispose to coeliac disease. The results definitely indicate that the HLA susceptibility to CD is restricted to DQ2 and DQ8 alleles and individuals negative for these alleles have only a very low risk to develop CD.

Previously published cases of patients negative for both DQ2 and DQ8 heterodimers have had one part of the DQ2 heterodimer (150;210;223-225) with only nine exceptions (140;155;160;226;227)(Table 9). The similarity in these nine patients has been the over-represented HLA-DQ6 (DQA1*01;DQB1*06), which was seen in 6/9 patients. We did not observe an excess of DQ6 instead 20/61 DQ2 and DQ8
negative patients carried the DQ5 \((DQA1*01;DQB1*05)\) molecule which has previously been claimed to be a predisposing factor (150) as well as a protective one (155). We also noticed an excess of DQ8 homozygous patients (29/60), but the frequencies varied greatly between populations. However, it is intriguing that there are no published papers reporting a surplus of patients heterozygous for DR4;DQ8/DR3;DQ2 if both of these DQ types indeed confer a risk independently.

The four patients negative for the known susceptibility genes did not have any obvious similarities in symptoms, age of diagnosis, or HLA alleles (Table 10); thus, they cannot suggest novel susceptibility factors within the HLA region. They may represent phenocopies, i.e. cases with a genuinely different genetic background. Although strong emphasis was laid on the accurate clinical data in the collaborative study, we must also question the correctness of the diagnoses. Also possible mix up with the samples can not be definitely ruled out. It is notable, that some other disorders have similar symptoms to CD and the biopsies are often open to interpretation.

Table 9. The published coeliac patients without the HLA-DQ2, -DQ8 or either allele encoding the DQ2 molecule.

<table>
<thead>
<tr>
<th>Population</th>
<th>DRB1</th>
<th>DQA1</th>
<th>DQB1</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish</td>
<td>13,1501</td>
<td>0102</td>
<td>0602,0604</td>
<td>Spurkland et al 1992</td>
</tr>
<tr>
<td>Spanish</td>
<td>13,08</td>
<td>0103,0401</td>
<td>0603,0402</td>
<td>Spurkland et al 1992</td>
</tr>
<tr>
<td>French</td>
<td>07,08</td>
<td>0201,0401</td>
<td>0303,0402</td>
<td>Djilali-Saiah et al 1994</td>
</tr>
<tr>
<td>Swedish</td>
<td>07,08</td>
<td>0201,04</td>
<td>0303,04</td>
<td>Ploski et al 1996</td>
</tr>
<tr>
<td>Swedish</td>
<td>0404,14</td>
<td>0301,0101</td>
<td>0304,0503</td>
<td>Ploski et al 1996</td>
</tr>
<tr>
<td>Finnish</td>
<td>13</td>
<td>0103</td>
<td>0603,0603/7</td>
<td>Polvi et al 1998</td>
</tr>
<tr>
<td>Finnish</td>
<td>1302,0801/6</td>
<td>0101/2,0401</td>
<td>0604,0402</td>
<td>Polvi et al 1998</td>
</tr>
<tr>
<td>Spanish</td>
<td>1301,08</td>
<td>0102,0601</td>
<td>0604,0301</td>
<td>Polvi et al 1998</td>
</tr>
<tr>
<td>Spanish</td>
<td>15011</td>
<td>-</td>
<td>0602/3</td>
<td>Garrote et al 2000</td>
</tr>
</tbody>
</table>

Table 10. Genetic and clinical data from the four unique coeliac patients without HLA-DQ2, -DQ8 or either allele encoding DQ2 molecule found in the multinational European study.

<table>
<thead>
<tr>
<th>Patients</th>
<th>Italy 24</th>
<th>Italy 25</th>
<th>Italy 26</th>
<th>UK 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRB1</td>
<td>14,01</td>
<td>0804,0102</td>
<td>0401,0701</td>
<td>08,01</td>
</tr>
<tr>
<td>DQA1</td>
<td>01,0104</td>
<td>0401,0101</td>
<td>0303,0201</td>
<td>0401,01</td>
</tr>
<tr>
<td>DQB1</td>
<td>05,05</td>
<td>0402,0501</td>
<td>0301,0303</td>
<td>04,0501</td>
</tr>
<tr>
<td>Age at onset</td>
<td>1 year</td>
<td>Asymptomatic</td>
<td>1.5 years</td>
<td>50 years</td>
</tr>
<tr>
<td>Age at diagnosis</td>
<td>2 years</td>
<td>6 years</td>
<td>2 years</td>
<td>50 years</td>
</tr>
<tr>
<td>Family history</td>
<td>none</td>
<td>Sibling with CD, family screening</td>
<td>Sibling with CD</td>
<td>Sibling with CD</td>
</tr>
<tr>
<td>Symptoms</td>
<td>Diarrhoea, failure to thrive, vomiting</td>
<td>Asymptomatic</td>
<td>Diarrhoea, vomiting, failure to thrive, abdominal pain</td>
<td>Diarrhoea</td>
</tr>
</tbody>
</table>

CD = coeliac disease
No new predisposing risk factors within HLA was revealed by this European collaboration. No association was seen with HLA class I alleles, HLA-A or -B, in the patients with solely the DQA1*05 or DQB1*02 part of DQ2. Only HLA-B*44 and perhaps B*13 were enriched but this may be due to linkage disequilibrium with DR7. No indication of previously described association of DPB1*0101 (207;228) or DPB1*0301 (229) alleles were noted either. Also the DRB4* and DRB1*04- alleles among patients were apparently equally distributed, each population however having its own common alleles.

This study of DQ2 negative patients revealed no new information about other possible risk factors located within the HLA region, but confirmed the reported results of the strong importance of DQ2 and the milder effect of DQ8. We have to accept that DQ2 and DQ8 are the major predisposing molecules for coeliac disease. Also, even either one of the alleles, DQA1*05 or DQB1*02, encoding for the DQ2 molecule may be sufficient alone. As patients without these alleles are extremely rare, it is possible to use the HLA-DQ typing to exclude CD in risk families. The use of DQ typing in routine diagnostics has been tested by our group (230) and has given encouraging results. This kind of genetic testing in risk groups may become more justified in the future if the high prevalence figures, many associated disorders and increased lymphoma risk in undiagnosed individuals prove to be correct.
CONCLUDING REMARKS

In the present thesis it was formally demonstrated by the twin study that the gluten sensitivity has a strong genetic component. The concordance rate was very high, 0.91. This figure is extremely similar to a recently published rate based on a population study. We also confirmed that intestinal and dermatological coeliac disease, dermatitis herpetiformis, appear to have the same genetic background, as the monozygotic twins could develop either manifestation with the same set of genes. In addition, we studied sib pairs who were discordant for the manifestation type and they too shared the predisposing DQ alleles and gave no evidence that the HLA haplotypes would have an effect on variation in clinical manifestation. These results strongly indicate that from a genetic point of view, intestinal coeliac disease and dermatitis herpetiformis should be regarded as different manifestations of one disease entity.

In spite of intensive studies, only a very limited number of predisposing genes for coeliac disease are known. The only firmly established locus is still the HLA-DQ. The European Genetics Cluster on Coeliac Disease collected 1008 families for genetic analysis and we were able to show that 90% of the index cases had genes encoding the HLA-DQ2 molecules, 5% had genes for DQ8 and most of the rest the DQA1*05 or DQB1*02 allele i.e. either allele for the DQ2 molecule. Only four genetically unique patients were left. Hence, the HLA typing can be used for genetic diagnostics. However the predisposing effect of HLA genes has been estimated to be only about 40% of the total genetic risk. Hence, there must be numerous unknown genetic factors with minor predisposing effects elsewhere in the genome, maybe also within or near the HLA region, hidden behind the strong linkage disequilibrium as actually was suggested by the present segregation study of DR3;DQ2 haplotypes. However the putative novel HLA-linked risk locus remains to be mapped.

The incidence of coeliac disease, including its silent and non-typical forms, may be surprisingly high in European populations. The manifestations may substantially vary from classical enteropathy to osteopenia or even to autoimmunity. To avoid these diseases, an effective combination of genetical, immunological and clinical diagnostics should become available. The present thesis provides further scientific evidence that due to the extremely strong association with certain alleles, the HLA-DQ determination might be one useful part of the diagnostics. It remains to be seen when another predisposing gene with a milder effect will be located.
ACKNOWLEDGEMENTS

This work was carried out at the Finnish Red Cross Blood Transfusion Service, in the Tissue Typing Laboratory. I'm grateful to the former and present Directors of the Institute, Professor Juhani Leikola and Docent Jukka Rautonen, and all the other superiors, Docent Jarmo Laine, Docent Tom Krusius, Docent Jukka Partanen, and Professor Saija Koskimies, for providing the excellent research facilities and possibilities to carry out the work.

Most sincere thanks to my supervisor Docent Jukka Partanen, whose ideas, knowledge, and skills and even their combination, the know-how, kept this study going. His wild theories and deep enthusiasm for research as well as wide comprehension and many times cheerful character have surprised again and again in numerous official-and unofficial-conversations.

The text of this book owes sincere gratitude to reviewers Docent Jorma Ilonen and Docent Erkki Savilahti for their constructive criticism in improving the text. Warm thanks also to Professor Hannu Saarilahti, Head of the Division of Genetics, Department of Biosciences, University of Helsinki, for helping with the bureaucracy and for his encouragement.

The studies would not have been possible without the excellent and devoted clinicians from Tampere, Markku Mäki, Kirsi Mustalahti, Pekka Collin, and Katri Kaukinen and the rest of the ‘ bile hile’ Coeliac Disease Study Group, without cooperation with physicians from Helsinki ‘DH guru’ Timo Reunala and Kaisa Hervonen and abroad collaborators the efficient Ilma Korponay-Szabo and Zsuzsa Szalai from Hungary and the joyful European Coeliac researchers from France, Italy, Norway, Sweden and the UK, especially Andrew Louka and Simon Moodie.

The experienced personnel of the Tissue Typing Laboratory have always offered a helping hand in preparing samples, HLA typing or detailed information of HLA alleles, haplotypes and distribution quite often combined even with lively chatting. I really appreciate it and would like particularly to thank people from the DNA-laboratory and magical-Marjaana, for the tiptop results even from the impossible samples.

Sincere acknowledgements of revision of the language of this thesis to James Woolley and for editing and lay-out work to Mikka Haimila. The grateful thanks also to the library ladies for their library services as well as to the secretaries for making the work so much easier.

I’m grateful to the ever growing group of senior and junior colleagues, their scientific knowledge as well as psychological help keeping up the appearance. The work has been many times only fun and the days, months, years, have gone faster than could be believed as there has been such a wonderful working team waiting. I’ll be missing you, my dear room mates Ani, Jyrki, Katri, Liisa, Niina, Noora, and Taina.
like I long for the previous colleagues and friends from other rooms. Special thanks to Anne, Antti, Nina, and Päivi for sharing their information and enabling articles of this thesis to be published and Katri, Niina, Päivi, and Taina for the deep discussions concerning every aspect of coeliac disease, for the help in organising dissertation and over all company.

I’m filled with gratitude towards my non-geneticists friends who have brought me up from the deep helixes of DNA, provided me with new thoughts concerning actual world news, weddings or growing children, introduced me to new interesting people and hobbies and opened my eyes for luxurious wines, food and sweets – is chocolate after all a proper food?

The warmest thanks belong to my mother and father for their mental as well as material support and for my darling sister, the style counsellor and tolerant travel mate. How I have relied on your ever lasting down to earth attitude and never-failing encouragement!

And the patients: we do our best for you; thank you for making these studies possible by participating. We all work hard keeping in our mind that in the end the goal is always the cure for the disease.

The studies of this theses have been supported by a grant from the Commission of the European communities, specific RTD programme “Quality of life and management of living resources”: QLKT-1999-00037, “Evaluation of the prevalence of coeliac disease and its genetic components in the European population”, by Sigrid Juselius Foundation, Medical Research Funds of Finnish Red Cross Blood Transfusion Service, University of Helsinki, Emil Aaltonen Foundation, Maud Kuistila Memorial Foundation, Päivikki and Sakari Sohlberg Foundation, Foundation of the Friends of the University Children’s Hospitals in Finland, Academy of Finland Research Council for Health, funding decision number 73489 and Medical Research Fund of Tampere University Hospital.

Helsinki, February 2003

Kati Karell
REFERENCES

5. Jarry A, Cerf-Bensussan N, Brousse N, Selz F & Guy-Grand D: Subsets of CD3+ (T cell receptor alpha/beta or gamma/delta) and CD3- lymphocytes isolated from normal human gut epithelium display phenotypical features different from their counterparts in peripheral blood. Eur J Immunol 1990; 20: 1097-1103.
REFERENCES


47. Park YS, She JX, Noble JA, Erlich HA & Eisenbarth GS: Transracial evidence for the influence of the homologous HLA DR-DQ haplotype on transmission of the HLA DR4 haplotypes to diabetic children. Tissue Antigens 2001; 57: 185-191.


REFERENCES


154. Spurkland A, Ingvarsson G, Falk ES, Knutsen I, Sollid LM & Thorsby E: Dermatitis herpetiformis and celiac disease are both primarily associated with the HLA-DQ (alpha 1*0501, beta 1*02) or the HLA-DQ (alpha 1*03, beta 1*0302) heterodimers. Tissue Antigens 1997; 49: 29-34.


REFERENCES


REFERENCES

REFERENCES


