Coeliac disease in children and adults is characterised by small intestinal mucosal injury and malabsorption. For almost two decades, it has been recognised that coeliac disease occurs much more frequently in individuals having certain human leucocyte antigens.

Human leucocyte antigens in man are encoded by genes in the major histocompatibility complex on the short arm of chromosome 6. The major histocompatibility complex contains genes that encode a broad array of proteins. These genes can be divided into three major groups; the class I, II, and III genes (Figure).

Class I and class II genes code for specialised and highly polymorphic cell surface glycoproteins which play a key role in antigen recognition by T cells. Class III genes code for components of the complement system – for example, C4A, C4B, Bf, and C2. The genes encoding steroid 21 hydroxylase also map to that region. The Class I genes code for molecules which are present on the surface of all nucleated cells. Human leucocyte antigen (HLA) class I molecules present antigen to T cells which express the CD8 protein on their surface. Thus, CD8 T cells are said to be class I restricted.

Class II genes are encoded in the HLA-D region at the centromeric end of the major histocompatibility complex. The HLA-D region can be divided into three major subregions termed -DP, -DQ, and -DR (Figure). Each class II molecule on the cell surface occurs as a heterodimer containing an alpha chain encoded by an A gene and a beta chain encoded by a B gene. HLA class II molecules are present on the surface of B cells, macrophages, dendritic reticular cells, and activated T cells. They can also be expressed, under certain circumstances, however, on other cell types including intestinal epithelial cells. Moreover, their expression on cells can be up-regulated by T cell lymphokines such as gamma interferon. HLA class II molecules present antigen to T cells which have the CD4 molecule. Thus, CD4 T cells are said to be class II restricted.

The HLA-D region contains multiple genes in addition to those described above (Figure). These include pseudogenes within DP (DPB2, DPA2) and DR (DRB2), DQB2 and DQA2 genes (also known as DX) which are not transcribed, and a DN A and DO B gene. Because protein products of these genes are not expressed on the cell surface, they do not appear to have a role in antigen presentation to T cells.

In addition to HLA class I and II genes, and complement genes, the major histocompatibility complex contains genes that encode tumour necrosis factor alpha and beta, and heat shock proteins – 70. Recent studies have also described five genes (termed BAT genes), whose function thus far is unknown. Those genes map between C2 and HLA B – that is, in the region that also contains the tumour necrosis factor and heat shock proteins – 70 genes.

Initially, coeliac disease was associated with the HLA class I B locus marker -B8, as defined by serology. Later, it was recognised, also by serology, that coeliac disease had a still stronger association with the class II D region marker, HLA-DR3. The association with HLA-B8 simply appeared to reflect strong linkage disequilibrium between the allele that encodes HLA-B8 and the allele that encodes HLA-DR3.

Polymorphism is an important characteristic of HLA class I and II major histocompatibility complex genes and their protein products. For example, there are 50 or more different alleles encoded by the class I HLA B locus and 18 or more different alleles at a single class II DR locus – that is, DRB1. Despite this extensive polymorphism and the theoretical possibility of enormous numbers of different HLA haplotypes within a given race, a relatively limited number of different haplotypes are present in most members of the population. Further, coeliac disease patients, like those with several other HLA-associated diseases – for example, insulin dependent diabetes mellitus and rheumatoid arthritis, carry a specific constellation of major histocompatibility antigen alleles, many of which are inherited together on the disease associated haplotype. The HLA B8, DR3 haplotype which is associated with coeliac disease is common among northern European caucasians.

With the advent of convenient methods for cloning and sequencing of the polymorphic regions of the HLA class II genes, it became clear that their protein products were substantially more polymorphic than had been recognised by serology. Thus, there are at least eight different alleles of the DQA1 gene, and 19 different alleles of the DPB1 gene. In addition, there are multiple DRB gene loci – that is, DRB1, DRB3, DRB4 that code for expressed DR beta chains, although some haplotypes lack the DRB3 and/or DRB4 locus. The DRB1 locus, which encodes the -DR3 serologic marker, is particularly polymorphic. In contrast, the DPA1 gene is relatively non-polymorphic (two alleles) and the DR A gene is non-polymorphic.

HLA-DR3 (DRw17) and -DQw2 are present in 90% or more of northern European caucasians with coeliac disease. In caucasians HLA-DQw2 is strongly linked to -DR3 (DRw17). The same markers, however, can be found in as many as 25–30% of individuals, within the same population, who do not have coeliac disease. Further, less than 0.1% of individuals with the serologic markers HLA-DR3 and -DQw2 develop coeliac disease. This limits the value of the DR3 and DQw2 antigens as serologic markers of disease, and led us to examine the HLA class II genes associated with coeliac disease at the molecular level.

To define the genetic markers for coeliac disease on the DR3, DQw2 haplotype more precisely, we used restriction
fragment length polymorphism analysis. Our studies revealed that approximately 95% of -DR3, DQw2 coeliac patients had a polymorphic 4-0 kb Rsa I DNA fragment, derived from a HLA class II B gene, that was present in less than 30% of healthy -DR3, DQw2 matched controls. This restriction fragment length polymorphism helped to distinguish the -DR3, DQw2 haplotype associated with coeliac disease from the DR3, DQw2 haplotype present in most controls. In further studies, the HLA class II B gene associated with this restriction fragment length polymorphism was isolated from a genomic library and shown to encode a HLA -DP beta chain.

Studies indicate that susceptibility genes for coeliac disease may be encoded as far centromeric in the class II D region as the HLA-DP subregion. Additional evidence for the importance of genes in or linked to HLA-DP in determining coeliac disease susceptibility was provided by the finding that the 4-0 Kb DP B gene restriction fragment length polymorphism was as prevalent in individuals who lacked -DR3, DQw2 as in healthy -DR3, DQw2 subjects. Thus, in non-coeliacs, the restriction fragment length polymorphism did not preferentially associate with -DR3, DQw2.

Our studies suggest that the HLA-associated susceptibility to coeliac disease is multigenic, with genes in the DQ/DR subregion and genes linked to, or in the DP subregion, determining disease susceptibility. Generally there is a high degree of recombination, estimated to be 2-5%, between the HLA DP and the HLA DR/DQ subregions and therefore only weak linkage disequilibrium between genes in those subregions. Coeliac disease may select for an extended HLA haplotype that includes the genes coding for -DR3, -DQw2 and a particular DP allele marked by the 4-0 Kb DP B gene restriction fragment length polymorphism. Alternatively, it could be that the putative DP-linked susceptibility gene is present on the other haplotype.

This work poses two further questions. First, is there a unique DR, DQ, or DP structural variant encoded within the HLA class II region in patients with coeliac disease? Second, which specific HLA-DP B genes are associated with coeliac disease? The greatest polymorphism in HLA class II genes occurs within the second exon sequences that encode the portion of the HLA class II molecule that is involved in binding antigen and in interacting with the T cell receptor for antigen. To address the first question, we sequenced the second exons of DR, DQ and DP genes (DRB1, B3, B4; DQA1, B1; DPB1) from heterozygous coeliac disease patients having the serologic markers DR3, DQw2, and the 4-0 Kb Rsa I DPB gene restriction fragment length polymorphism. The relevant gene segments were amplified from genomic DNA using the polymerase chain reaction (PCR), and then cloned and sequenced. In each case, the sequences from coeliac disease patients were identical to sequences that are also present, although at a significantly lower frequency, in unaffected individuals.

This finding is compatible with a model in which specific HLA class II genes are necessary, but not sufficient, for the phenotypic expression of coeliac disease. In support of that model, only 25-30% of siblings who appear to share one or both HLA haplotypes (as determined by DR serology and, by inference DQ), are concordant for coeliac disease. Moreover, our findings in coeliac disease parallel those in other HLA class II associated diseases. Thus, specific alleles of HLA-DR and -DQ on various HLA DR4 haplotypes are over-represented in patients with insulin dependent diabetes mellitus, rheumatoid arthritis and pemphigus vulgaris, compared with controls. However, these patients do not have class II gene sequences that are unique to the disease.

To analyse the HLA class II DPB genes associated with coeliac disease, sequence specific oligonucleotides were used to probe PCR amplified second exons of DP B genes from coeliac disease patients. Those studies revealed a significant over-representation in coeliac disease patients of the relatively rare DPB alleles, DPB1 and DPB3. Furthermore, the increased frequency of those alleles in coeliac disease accounted for the increased frequency of the 4-0 Kb Rsa I DPB gene restriction fragment length polymorphism in that disease. Our patients were mostly of northern European caucasian ancestry. Specific DP alleles, however, also appear to determine coeliac disease susceptibility among Italians. In that population, DPB4-2 and DPB3, but not DPB1, were most important. Italian coeliac disease patients also differ from northern European main, Italy). Coeliac disease patients in the distribution of DR markers as determined by serology that is, -DR3 (DRw17) less common; DR5/7 heterozygotes more common.

Several conclusions can be drawn from the available data regarding which HLA class II genes are most important in determining coeliac disease susceptibility. Within the DR/DQ subregions, the DQB2 allele is important for coeliac disease susceptibility. The DQB2 allele encodes the DQw2 serologic specificity and is present in virtually all caucasian coeliacs with -DR3 (DRw17). In addition, the DQB2 allele is present in most individuals with the -DR7 serologic specificity, which is often associated with coeliac disease in southern Europeans (Spain, Italy). Coeliac disease patients with -DR7, however, are usually heterozygous for -DR5 or -DR3 on the second chromosome. In this regard, a specific DQ A allele (DQA4-1) that is common to DR5 and DR3 haplotypes also has been associated with coeliac disease. The finding of increased susceptibility to coeliac disease in individuals with -DR7 that is, having the DQB2 allele but lacking the DQA4-1 allele, associated with heterozygosity for -DR5 that is, lacking the DQB2 allele but having the DQA4-1 allele, suggests that a DQ molecule encoded by both DQB2 and DQA4-1 alleles contributes to coeliac disease susceptibility. This DQ molecule, which is recognised as DQw2 by serology, would be encoded in cis – that is on the same chromosome, on –DR3 (DRw17) haplotypes and in trans on DR5/7 haplotypes – that is by the DQA4-1 allele from the chromosome having DR5 and the DQB2 allele from the chromosome having DR7. In further support of the importance of DQw2, coeliac disease is unusual among black populations in the United States. In this regard, the -DR3 haplotype present in many blacks has undergone recombination between DQ and DR such that those -DR3 (DRw18) individuals often lack DQw2. A final conclusion relates to the role of DP genes in determining coeliac disease susceptibility. There is a striking increase in coeliac disease among individuals with DQw2 who also have specific DP B alleles – that is, the DPB1 or DPB3 allele in Northern European caucasiens, DP4-2 or DPw17 Italian. The association of those DP B alleles with coeliac disease may reflect a direct effect of those genes, or an effect of genes linked to the DPB locus. In either case, more than one gene within the major histocompatibility complex appears to determine coeliac disease susceptibility.

What unique feature of those HLA class II alleles results in increased susceptibility to coeliac disease? To address this question, one can compare the DQ and DP alleles associated with coeliac disease susceptibility to those not associated with disease. The goal of such an approach is directed towards defining putative epitopes on the susceptibility alleles that differ from those on non-susceptibility alleles. Such an analysis suggests distinct amino acid residues on the class II molecules associated with coeliac disease that may be important in determining disease susceptibility. Moreover, those residues lie within the region of the class II molecules thought to be important in peptide binding.
Understanding the molecular basis of coeliac disease

Several issues have not been answered. For example, the mechanism by which HLA class II D region alleles result in coeliac disease susceptibility remains enigmatic. Further, which other genes, in or outside the major disease preparation, interact with the class II D region genes to increase disease susceptibility is not known, although concordance data suggest additional susceptibility genes exist. In addition, it is not known whether coeliac disease is associated with specific HLA class II alleles because of events that take place in the thymus during development of the T cell repertoire, or because of events that take place in the periphery after T cell differentiation. Finally, the specific antigens and the determinants on those antigens that interact with HLA class II molecules and T cells as a key element in the pathogenesis of coeliac disease are poorly understood. However, our studies suggest that environmental factors other than dietary grains – for example, viral proteins, also play a role in determining coeliac disease susceptibility. 18

M F KAGNOFF

University of California, San Diego La Jolla, CA 92037, USA

Supported by NIH Grant DK35108. I would like to thank Ms V Shu for preparation of the manuscript.