Gut

Leading article

Molecular biology and coeliac disease

In clinical and pathological terms our descriptive understanding of coeliac disease is extensive.¹ In recent years our understanding of the immunology of coeliac disease, together with that of mucosal immunology in general, has rapidly increased.^{2,3} Several basic questions remain, however. What part of wheat gluten is involved in the pathogenesis of coeliac disease? Why are some people predisposed to develop the disease? How is the characteristic pathological lesion produced? Are other environmental factors involved? The application of molecular biological techniques may well illumine our understanding of these questions.

What part of wheat gluten is involved?

That wheat gluten is deleterious in coeliac disease has been known since 1953.⁴ Subsequent studies have usually concentrated on the ethanol soluble proteins (gliadins) rather than the water insoluble mixture of gluten. The gliadin proteins have been characterised by their relative electrophoretic mobility on starch⁵ or polyacrylamide gel electrophoresis.⁶ Four fractions were thus defined (α , β , γ , ω) with molecular weights from 32 to 58 kDa.⁷

In vitro and in vivo studies of these fractions against coeliac mucosa showed that all had a 'toxic' effect,⁸⁻¹¹ suggesting that there is a common peptide in all the gliadin proteins that is responsible for coeliac toxicity. The recognition of this common sequence has, therefore, been one goal of investigators. Molecular biological techniques have proved important in the determination of the primary structure of gliadins. In 1984 Kasarda *et al*¹² determined the complete primary structure of an α -gliadin using protein chemistry and derivation from a cloned cDNA sequence. Similar studies have permitted the determination of the amino acid sequence of other gliadins. Bartels and Thompson^{13,14} were able to produce cDNA clones from the mRNA of wheat endosperm (the flour, or storage protein component of the grain), and thus determine the primary structures of a number of α - and γ -gliadins.^{15,16}

Several exciting developments are likely following these observations. Expression vectors from cDNA could be used in which high levels of gliadins could be expressed from cloned cDNA.¹⁷ The hope would be to synthesise enough gliadin (or truncated or otherwise mutated derivatives) for in vitro testing against coeliac mucosa.⁹ Ultimately, sufficient quantities would be required for clinical challenge studies. Once a specifically toxic peptide had been recognised in this way, it might prove possible to delete the encoding DNA from the wheat germ and to develop a non-toxic strain of

wheat. Even more exciting would be to introduce DNA which encodes for high molecular weight glutenins¹⁸ into the genome of a non-toxic cereal. These viscoelastic glutenins give wheat its good baking properties. Thus, a new transgenic plant could be produced with the baking qualities of wheat, but which was non-toxic to coeliac patients.

How does gliadin produce its effect?

GENETIC STUDIES OF MHC

There are several theories of the pathogenesis of coeliac disease,¹⁹ but the balance of evidence currently favours an immunological aetiology.² Support for this hypothesis comes from analysis of the genetics of the disorder. Approximately 10% of first degree relatives of coeliac disease patients have the condition, although the precise inheritance is unknown. Coeliac disease is associated with the HLA class I B locus marker – B8^{20,21} – and there is a stronger association with the class II D region marker HLA-DR3.²² The association with B8 reflects linkage disequilibrium between the allele that encodes B8 and that which encodes DR3. An even closer association has been shown between DQW2 and coeliac disease,²³ the alleles encoding DQW2 also being in linkage disequilibrium with B8 and DR3. The importance of these strong associations between coeliac disease and the class II D locus genes is that these genes code for cell surface molecules that are critical to the immune response. HLA class II molecules occur as transmembrane glycosylated heterodimers consisting of an α chain of 34kDa encoded by an A gene and a β chain of 29 kDa encoded by a B gene. A and B genes are located in the HLA-D region of the genome, this region being divided into three major sub-regions termed DP, DQ, and DR. The class II molecules are expressed on the surface of B lymphocytes, macrophages, dendritic cells, activated T cells and also intestinal epithelial cells. These class II molecules are involved in the presentation of endogenously processed antigen to antigen specific T lymphocytes which possess the CD4 surface molecule – that is, are helper T cells. Thus, it is suggested that the particular D region haplotype associated with coeliac disease allows antigen (gliadin peptide) to be presented in such a way that T helper cells within the small intestinal mucosa are specifically stimulated to produce an immunological reaction leading to the typical pathological changes of coeliac disease.

Although 90% of coeliac patients may possess the particular HLA class II DR3, DQW2 haplotype, 25–30% of normal individuals within the same population have the same

haplotype and do not have coeliac disease. Likewise, less than 0.1% of individuals with DR3, DQW2 develop the disease. However, the associated D locus haplotypes so far described were recognised by serological typing using antibodies specific for HLA phenotypes. This serological typing depends upon antibody specificity and can lead to under- or over-recognition of particular epitopes which may not represent accurately the situation at the genetic level. Investigators have, therefore, sought to analyse the class II D genes at the molecular level to determine whether there is any specific marker or disease associated gene. Molecular biological techniques have shown a far greater complexity and polymorphism in the HLA genes than was previously recognised. For example, there are at least eight different alleles of the DQA1 gene and 15 of the DQB1 gene: similarly there are 21 different alleles of the DPB1 gene and two of the DPA1 gene.²⁴⁻²⁶

Class II genes have been studied using restriction fragment length polymorphism analysis (RFLP) and Southern blotting and also analysis of the genome using oligonucleotide allele specific probes on polymer chain reaction (PCR) amplified DNA fragments. Such studies have shown both DP α and β chain RFLPs in association with coeliac disease.^{27,28} These RFLPs were independent of the DR3, DQW2 haplotype, suggesting a multigenic susceptibility for coeliac disease.²⁸ DP locus polymorphisms associated with coeliac disease have also been described by other workers.^{29,30}

Gene probes have been used to examine the amino acid sequence of expressed class II polypeptides described serologically by the RFLP haplotypes. Kagnoff *et al.*,³¹ using PCR and DNA sequencing techniques, found no differences in the sequences between coeliac disease patients and controls, suggesting that coeliac disease patients do not have class II gene sequences that are unique to the disease and that particular HLA class II genes are necessary, but not sufficient, for the phenotypic expression of the disease. However, using oligonucleotide primers for DPB genes, a significant over-representation in coeliac disease patients of the relatively rare alleles DPB1 and DPB3 was detected.³¹ This was in patients of northern European descent, whereas DPB4.2 and DPB3 genes were more important in coeliac disease patients from southern Europe.³²

Molecular analysis of other alleles in coeliac disease has failed to show a unique disease associated sequence.³³ Nor has such analysis of large DNA fragments, using field inversion techniques, detected any disease specific association (Blair and Howdle, unpublished observations).

These findings suggest that a small genetic change may produce a variation in the antigen binding groove of the HLA class II molecule which is of critical importance to disease susceptibility.

Sollid *et al.*³⁴ showed that the DQA and B genes which code for the serologically defined DQW2 haplotype may be encoded in *cis* on DR3 haplotypes and in *trans* on DR5/7 haplotypes. Such a combination of DQA1.2 and DQB1.2 genes was found in 98.9% of coeliac patients – the strongest association yet found. This suggests that the class II receptor, coded by the DQ A and B genes and defined serologically as DQW2, presents the gliadin peptide to T cells for immune recognition. However, Mantovani *et al.*³⁵ failed to show a unique sequence in the DQA gene, stressing the fact that in coeliac disease certain genetic variations may be necessary for the disease to develop but are not sufficient in themselves.

able interest not only in the HLA cell surface molecules but also in the receptor on the T cells (TcR) which recognises the antigen in the groove formed by the HLA heterodimer.³⁵ The HLA class II molecules present antigen to CD4⁺ T lymphocytes (helper T cells). In the small intestinal mucosa in normal and coeliac tissue, the helper cells are predominantly located in the lamina propria beneath the epithelium; intra-epithelial T cells are predominantly CD4⁺, CD8⁺ (suppressor T cells) or CD4⁺CD8⁺. The CD4⁺ cells recognise the antigen-HLA class II combination via the T cell receptor (TcR), which is composed of two disulphide linked glycoprotein chains designated α and β and of molecular weights 80 kDa and 40 kDa respectively. Both these chains are clonally variable and are encoded by families of variable (V), diversity (D), joining (J), and constant (C) gene segments, which assemble by gene rearrangements during T cell differentiation (each genetic locus is thus organised in the same way as the immunoglobulin genes). Thus, the multiplicity of V, D, and J segments and their independent association creates a large potential T cell receptor repertoire. Although every individual may use a differently composed TcR for recognition of the same antigen, it is likely that there is some restricted TcR gene usage in different conditions. Evidence for this is suggested by experimental allergic encephalomyelitis in mice (a model for multiple sclerosis) where 80% of the T cells recognising the encephalitogenic peptide were found to express the same receptor, and utilised the V β 8.2 gene.^{36,37} Similarly, in diabetic mice, diabetogenic T cell clones principally expressed the V β 5 gene family.³⁸ There is very little evidence yet for such restricted gene usage in human TcRs, but it is now possible to reverse transcribe and amplify by PCR the mRNA of specific V β gene families which are being expressed. We are currently studying such V β gene usage by T cells from coeliac disease patients. Any such restriction of gene usage would have important diagnostic and therapeutic implications for coeliac disease.

Are other environmental factors involved in the pathogenesis of coeliac disease?

While it is well documented that gliadin is responsible for the small intestinal lesion in coeliac disease, it has been suggested that other environmental factors are involved. This is supported by the genetic data showing specific HLA associations with coeliac disease, but many individuals with the same associations do not have the disease. One can envisage a similar argument if certain TcR genes are shown to be associated with coeliac disease.

In 1984, Kagnoff *et al.*³⁹ suggested that adenovirus infection may be an environmental factor in coeliac disease. They showed immunological cross reactivity between an early protein (E1B-58 kDa) of adenovirus 12 (Ad12) and gliadin. Amino acid sequence analysis showed a region of homology between the two proteins, with 8 of 12 amino acids being identical. Kagnoff *et al.*³⁹ suggested that previous infection with adenovirus 12 could lead to the development of coeliac disease in susceptible individuals on exposure to gliadin, the immunological cross reactivity being the basis of the pathogenesis. In support of this hypothesis, neutralising antibody titres specifically to Ad12 but not other adenoviruses were raised in coeliac disease patients,⁴⁰ suggesting these patients had an increased prevalence of Ad12 infection. The antibody measured in these studies, however, was directed against coat proteins of the virus (probably mainly the hexon protein) and not against the E1B-58 kDa protein.

Treated coeliac disease patients have also been shown to have a cell mediated immune response in the peripheral blood to synthetic gliadin and viral peptides of the homologous sequence.^{41,42} There is no evidence yet of T cell

reactivity to the E1B-58 kDa protein itself, and we failed to show specific antibodies to this particular protein in coeliac patients.⁴³

A different approach to the hypothesis has been to seek evidence of persistent Ad12 infection in coeliac disease patients. Persisting infection would require continuing viral replication and, since the early region proteins (E1A and E1B) are involved in the initiation of replication, evidence of the early region genes in a persisting infection would be expected. We used PCR to seek evidence of persisting viral infection in small intestinal mucosa using specific oligonucleotide primers specific for the E1B-58 kDa gene.⁴⁴ DNA was isolated from small intestinal biopsy samples of coeliac disease and control patients and analysed, by PCR, for Ad12 DNA encoding the E1B-58 kDa protein. Four of 18 coeliac and 2 of 24 control patients were positive. There was thus a low prevalence of this infection in both groups of patients, but certainly no significantly increased incidence in coeliac disease. These results suggest that persistent Ad12 infection is not a major element in the pathogenesis of coeliac disease. Neither does persistent Ad12 infection seem to be involved in the development of coeliac associated lymphoma since PCR analysis of malignant tissues failed to detect Ad12 sequences (Blair and Howdle, unpublished observations).

Molecular biology is thus beginning to have an impact on several of the outstanding questions concerning coeliac disease. We envisage that the answers to some of these fundamental questions will be found in the near future.

P D HOWDLE
G E BLAIR

Departments of Clinical Medicine and
Biochemistry and Molecular Biology,
The University of Leeds,
Leeds

- Cooke WT, Holmes GKT. *Coeliac disease*. Edinburgh: Churchill Livingstone, 1984.
- Howdle PD, Losowsky MS. *The immunology of coeliac disease*. Baillière's Clinical Gastroenterology. London, Baillière, 1987; 507-29.
- Brandtzaeg P, Halstensen TS, Kett K, Krajci P, Kvale D, Rognum TO *et al*. Immunobiology and immunopathology of human gut mucosa: humoral immunity and intraepithelial lymphocytes. *Gastroenterology* 1989; **97**: 1562-84.
- Dicke WK, Weijers HA, Van de Kamer JH. The presence in wheat of a factor having a deleterious effect in cases of coeliac disease. *Acta Paediatrica* 1953; **42**: 34-42.
- Woychik JH, Boundy JA, Dimiter RJ. Starch gel electrophoresis of wheat gluten protein with concentrated urea. *Arch Biochem Biophys* 1961; **94**: 477-82.
- Bushuk W, Zillman RR. Wheat cultures identification by gliadin electrophoregrams. *Canadian Journal of Plant Science* 1978; **58**: 505-15.
- Kasarda DD. Structure and properties of α -gliadins. *Annals of Technology & Agriculture* 1981; **29**: 151-73.
- Jos J, Charbonnier L, Mosse J, Olives JP, de Tand MF, Rey J. The toxic fraction of gliadin digests in coeliac disease. Isolation by chromatography in Biogel P10. *Clinica Chimica Acta* 1982; **119**: 263-74.
- Howdle PD, Ciclitira PJ, Simpson FG, Losowsky MS. Are all gliadins toxic in coeliac disease? An in vitro study of alpha, beta, gamma and omega gliadins. *Scand J Gastroenterol* 1984; **19**: 41-7.
- Ciclitira PJ, Evans DJ, Fagg NLK, Lennox ES, Dowling RH. Clinical testing of gliadin fractions in coeliac patients. *Clin Sci* 1984; **66**: 357-64.
- Wieser H, Belitz HD, Idar D, Ashkenazi A. Coeliac activity of the gliadin peptides CT-1 and CT-2. *Z Lebensm Unters Forsch* 1986; **182**: 115-7.
- Kasarda DD, Okita T W, Bernadin JE. Nucleic acid (cDNA) and amino acid sequences of alpha-type gliadins from wheat (*Triticum aestivum*). *Proc Nat Acad Sci USA* 1984; **81**: 4712-16.
- Bartels D, Thompson RD. The characterization of cDNA clones coding for wheat storage proteins. *Nucleic Acids Res* 1983; **11**: 2961-77.
- Bartels D, Thompson RD. The endosperm of common cereals contains related poly A⁺ RNA sequences. *Theoretical Applied Genetics* 1983; **64**: 269-73.
- Bartels D, Altosaar I, Harberd NP, Barker RF, Thompson RD. Molecular analysis of γ -gliadin gene families of the complex gli-1 locus of bread wheat (*T. aestivum* L.). *Theoretical Applied Genetics* 1986; **72**: 845-53.
- Sugiyama T, Rafalshi A, Soll DG. The nucleotide sequence of a wheat γ gliadin genomic clone. *Plant Sciences* 1986; **44**: 205-9.
- Bartels D, Thompson RD, Rothstein S. Synthesis of a wheat storage protein subunit in *E. coli* using novel expression vectors. *Gene* 1985; **35**: 159-67.
- Thompson RD, Bartels D, Harberd NP, Flavell RB. Characterization of the multigene family coding for HMW glutenin subunits in wheat using cDNA clones. *Theoretical Applied Genetics* 1983; **67**: 87-96.
- Davidson AGF, Bridges MA. Coeliac disease: a critical review of the aetiology and pathogenesis. *Clinica Chimica Acta* 1987; **163**: 1-40.
- Falchuk ZM, Rogentine GN, Strober W. Predominance of histocompatibility antigen HL-A8 in patients with gluten-sensitive enteropathy. *J Clin Invest* 1972; **51**: 1602-5.
- Stokes PL, Holmes GKT, Asquith P, Mackintosh P, Cooke WT. Histocompatibility antigens associated with adult coeliac disease. *Lancet* 1972; **ii**: 162-4.
- Ek J, Albrechtsen D, Solheim BG, Thorsby E. Strong association between the HLA-Dw3-related B cell alloantigen-DRw3 and coeliac disease. *Scand J Gastroenterol* 1978; **13**: 229-33.
- Tosi R, Vismara D, Tanigaki N, Ferrara GB, Cicimoura F, Buffolano W, *et al*. Evidence that coeliac disease is primarily associated with a DC locus allelic specificity. *Clin Immunol Immunopathol* 1983; **28**: 395-404.
- Horn GT, Bugawan TL, Long CM, Erlich HA. Allelic sequence variation of the HLA-DQ loci: relationship to serology and to insulin-dependent diabetes susceptibility. *Proc Natl Acad Sci USA* 1988; **85**: 6012-6.
- Todd JA, Bell JI, McDewitt HO. HLA-DQ β gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. *Nature* 1987; **329**: 599-604.
- Bugawan TL, Horn GT, Long CM, Mickelson E, Hansen JA, Ferrara GB, *et al*. Analysis of HLA-DP allelic sequence polymorphism using the in vitro enzymatic DNA amplification of DP- α and DP- β loci. *J Immunol* 1988; **141**: 4024-30.
- Howell MD, Austin RK, Kelleher D, Nepom GT, Kagnoff MF. An HLA-D region restriction fragment length polymorphism associated with coeliac disease. *J Exp Med* 1986; **164**: 333-8.
- Howell MD, Smith JR, Austin RK, Kelleher D, Nepom GT, Volk B, *et al*. An extended HLA-D region haplotype associated with coeliac disease. *Proc Nat Acad Sci USA* 1988; **85**: 222-6.
- Caffrey C, Hitman GA, Niven MJ, Cassell PG, Kumar P, Fry L, *et al*. HLA DP and coeliac disease: family and population studies. *Gut* 1990; **31**: 668-74.
- Hall MA, Lanchbury JSS, Bolsover WJ, Welsh KI, Ciclitira PJ. Coeliac disease is associated with an extended HLA-DR3 haplotype which includes HLA-DPw1. *Hum Immunol* 1990; **27**: 220-8.
- Kagnoff MF, Harwood JJ, Bugawan TL, Erlich HA. Structural analysis of the HLA DR, DQ and DP alleles on the coeliac disease associated HLA DR3 (Dw17) haplotype. *Pro Natl Acad Sci USA* 1989; **86**: 6274-8.
- Bugawan TL, Angelini G, Larrick J, Auricchio S, Ferrara GB, Erlich HA. A combination of a particular HLA-DP beta allele and an HLA DQ heterodimer confers susceptibility to coeliac disease. *Nature* 1989; **339**: 470-3.
- Mantovani V, Corazza GR, Angelini G, Delfino L, Frisoni M, Miri P, *et al*. Molecular analysis of HLA-DQ alleles in coeliac disease: lack of a unique disease-associated sequence. *Clin Exp Immunol* 1991; **83**: 74-8.
- Sollid LM, Markussen G, Ek J, Gjerde H, Vartdal F, Thorsby E. Evidence for a primary association of coeliac disease to a particular DQ heterodimer. *J Exp Med* 1989; **169**: 345-50.
- Davis MM, Bjorkman PJ. T-cell receptor genes and T-cell recognition. *Nature* 1988; **334**: 395-402.
- Acha-Orbea H, Michell DJ, Timmerman L, Wraith DL, Tausch GS, Waldor MU, *et al*. Limited heterogeneity of T cell receptors from lymphocytes mediating autoimmune encephalomyelitis allows specific immune intervention. *Cell* 1988; **54**: 263-73.
- Acha-Orbea H, Steinman L, McDewitt HO. T cell receptors in murine autoimmune disease. *Ann Rev Immunol* 1989; **7**: 371-405.
- Reich E-P, Sherwin RS, Kanagawa O, Janeway CA. An explanation for the protective effect of the MHC Class II I-E molecule in murine diabetes. *Nature* 1989; **341**: 326-8.
- Kagnoff MF, Austin RK, Hubert JJ, Bernardin JE, Kasarda DD. Possible role for a human adenovirus in the pathogenesis of coeliac disease. *J Exp Med* 1984; **160**: 1544-57.
- Kagnoff MF, Paterson YJ, Kumar PJ, Kasarda DD, Carbone FR, Unsworth DJ, *et al*. Evidence for the role of a human intestinal adenovirus in the pathogenesis of coeliac disease. *Gut* 1987; **28**: 995-1001.
- Karagiannis JA, Priddle JD, Jewell DP. Cell-mediated immunity to a synthetic gliadin peptide resembling a sequence from adenovirus 12. *Lancet* 1987; **ii**: 884-6.
- Mantzaris GJ, Karagiannis JA, Priddle JD, Jewell DP. Cellular hypersensitivity to a synthetic dodecapeptide derived from human adenovirus 12 which resembles a sequence of A-gliadin in patients with coeliac disease. *Gut* 1990; **31**: 668-73.
- Howdle PD, Blair Zajdel ME, Smart CJ, Trejdosiewicz LK, Blair GE, Losowsky MS. Lack of a serologic response to an E1B protein of adenovirus 12 in coeliac disease. *Scand J Gastroenterol* 1989; **24**: 282-6.
- Mahon J, Blair GE, Wood GM, Scott BB, Losowsky MS, Howdle PD. Is persistent adenovirus 12 infection involved in coeliac disease? A search for viral DNA using the polymerase chain reaction. *Gut* 1991; **32**: 1114-6.