Genes and environment in celiac disease

L. M. Sollid, S. N. McAdam, Ø. Molberg, H. Quarsten, H. Arentz-Hansen, A. S. Louka and K. E. A. Lundin

Institute of Immunology and Medical Department, Rikshospitalet, University of Oslo, Oslo, Norway

Sollid LM, McAdam SN. Molberg Ø, Quarsten H, Arentz-Hansen H, Louka AS, Lundin KEA. Genes and environment in celiac disease. Acta Odontol Scand 2001;59:183–186. Oslo. ISSN 0001-6357.

Celiac disease is an intestinal disorder that develops as a result of interplay between genetic and environmental factors. HLA genes along with non-HLA genes predispose to the disease. Linkage studies have failed to identify chromosomal regions other than the HLA region which have major effects, indicating the existence of multiple non-HLA predisposing genes with modest effects. Association studies have shown that CTLA4 or a closely located gene is one of these genes. The primary HLA association in the majority of celiac disease patients is with DQ2 (DQA1*05/DQB1*02) and in the minority of patients with DQ8 (DQA1*0301/DQB1*0302). Gluten reactive CD4+ T cells can be isolated from small intestinal biopsies of celiac patients but not from controls. DQ2 or DQ8, but not other HLA molecules carried by patients, present peptides to these T cells. A number of distinct T cell gluten epitopes exist, most of them posttranslationally modified by deamidation. DQ2 and DQ8 bind the epitopes such that the glutamic acid residues created by deamidation are accommodated in pockets that have a preference for negatively charged side chains. There is evidence that deamidation in vivo is mediated by the enzyme tissue transglutaminase (tTG). Overall, the results point to control of the immune response to gluten by intestinal T cells restricted by the DQ2 or DQ8 molecules. This is likely to be a critical checkpoint for the development of celiac disease and could explain the dominant genetic role of HLA in this disorder. The products of the other predisposing genes may participate in pathway(s) that lead(s) to lesion formation. The minor genetic effects of the non-HLA genes could indicate a lack of critical checkpoints along these pathways, or that there are several pathways leading to the lesion formation. □ *Celiac disease; genetics; HLA, polygenic; T cells*

Ludvig M. Sollid, Institute of Immunology, University of Oslo, Rikshospitalet, NO-0027 Oslo, Norway. Tel: +47 23073811, fax: +47 23073822, e-mail: l.m.sollid@labmed.uio.no

Celiac disease, rheumatoid arthritis, type 1 diabetes, multiple sclerosis, and myasthenia gravis are examples of chronic inflammatory diseases with multifactorial etiologies where genetic and environmental components are involved. The diseases demonstrate associations with particular HLA alleles indicating that HLA genes contribute to the genetic susceptibility. It has been estimated, however, that the HLA genes in the majority of the HLA associated disease contribute less than half of the heritable susceptibility (1). This knowledge has compelled intensive searches to identify non-HLA susceptibility genes and, up to now, linkage methods have been the prevailing approach. In general, these studies have failed to demonstrate any genetic region having a major effect except the HLA, but have indicated multiple regions harboring loci that may elicit weak effects. Hence the emerging general picture for the HLA associated, chronic inflammatory diseases is that HLA genes, along with a number of genes with weak (susceptibility or protective) effects, constitute the genetic predisposition. Moreover, there are probably modifying effects between the different genetic factors (epistasis) and interactions between the genetic and environmental factors. It may also be that in different patients with identical phenotype, only some of the genetic/environmental factors are shared and only some of the pathways leading to disease development are identical. If such heterogeneity exists, there will obviously be serious implications for the ease with which the hitherto unknown disease factors can be identified.

The concept of multiple predisposing/protective genes that interact with each other and/or the environment probably applies to other chronic inflammatory diseases with less clear HLA associations, such as periodontitis, but here the HLA genes seem to play less dominant roles. It is noteworthy that the ease with which a susceptibility locus can be identified is likely to be correlated to the genetic impact of the locus. Hence the task of identifying multiple weak loci with epistatic and environmental interactions represents a formidable challenge. Without the help of functional models to identify candidate genes and to test effects of gene mutations, this would probably be a futile enterprise.

Celiac disease is a common condition in which ingested wheat gluten (consisting of the subcomponents gliadin and glutenin) or related proteins from rye and barley are not tolerated. The patient suffers from chronic inflammation in the small intestine, resulting in villous atrophy and crypt hyperplasia. People are diagnosed at all ages, but there is a peak in incidence at early childhood. The disease is efficiently treated with a gluten exclusion diet. Poor diet compliance and undiagnosed disease are associated with complications, including increased risk of anemia, infertility, osteoporosis, and intestinal lymphoma. Among the human multifactorial chronic inflammatory disorders,

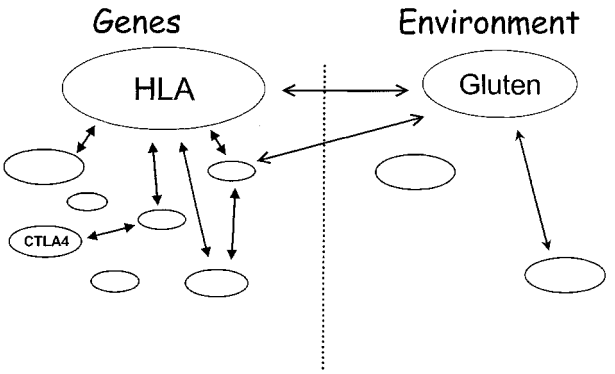

Fig. 1. Celiac disease, as many other chronic inflammatory diseases, is a multifactorial disorder that develops as a result of interplay between genetic and environmental factors. HLA and CTLA4 have been identified as genetic factors and gluten is a well-known environmental factor in this disease. Several additional genetic and probably also other environmental factors have yet to be identified. This figure illustrates in a simplified way how the genes and environmental factors may interact and modulate the effect of each other.

celiac disease stands out as a particularly good model due to a unique combination of three factors. First, an environmental factor (gluten) that precipitates disease is known, and withdrawal or re-introduction of gluten is frequently done to establish the diagnosis. Second, the HLA molecules that confer predisposition to the disease have been identified (HLA-DQ2 for the majority of patients and HLA-DQ8 for the minority). Third, access to the affected organ is simple; detailed in situ studies can be performed and disease-relevant cell populations can be isolated. This paper briefly covers some of the recent advances in this model disease.

Genetic factors in celiac disease

A high prevalence (10%) among first-degree relatives of celiac disease patients indicates that susceptibility to develop celiac disease is strongly influenced by inherited (genetic) factors (2). Familial clustering is often expressed as λ_s (the ratio of the prevalence in sibs over the population prevalence) (1). If this ratio is close to 1, there is no evidence for genetic factors in susceptibility. In contrast, the λ_s value for celiac disease is estimated to be 30–60 (1, 3), which is high compared to multifactorial disorders like rheumatoid arthritis, type 1 diabetes, and multiple sclerosis. The strong genetic influence in celiac disease is further supported by a high concordance rate of 70–100% in monozygotic twins (4). The sibship aggregation attributable to HLA ($\lambda_{s\ HLA}$) is estimated to be 2.3–5 (1, 3). Assuming a multiplicative model of disease-predisposing genes, these estimates suggest that the overall importance of non-HLA genes is greater than that of HLA genes (1, 3). However, these figures should be interpreted with caution, as increased sharing of environmental factors by the sibs (e.g. living in the same household) would tend to cause overestimation of the role of the non-HLA genes.

HLA genes

The majority of celiac disease patients carry the DRB1*0301-DQA1*0501-DQB1*0201 haplotype (the DR3, DQ2 haplotype) or are DRB1*11/12-DQA1*0505-DQB1*0301/DRB1*07-DQA1*0201-DQB1*0202 heterozygotes (carry the DR5-DQ7/DR7-DQ2 haplotypes) (5–9). The DQα chains encoded by DQA1*0501 and DQA1*0505 differ by one residue in the leader peptide, and the DQβ chains encoded by DQB1*0201 and DQB1*0202 differ by one residue in the membrane proximal domain. These substitutions are highly unlikely to have any functional consequences. Hence celiac disease patients with these DR-DQ combinations share the same functional DQ molecule on the cell surface, encoded by genes carried in *cis* (e.g. DQA1*05 and DQB1*02 carried on the same haplotype) or *trans* position (e.g. DQA1*05 carried on a different haplotype to DQB1*02) (10). Recombination (crossing over; rearrangement of genes on the same chromosome) seems to be an important mechanism for the generation of HLA haplotypes (11). Accumulating evidence suggests that the DR3-DQ2, the DR7-DQ2, and the DR5-DQ7 haplotypes have a close evolutionary relationship. Based on analysis of microsatellites, fragments of DNA flanking the DQA1 gene of the DR3-DQ2 haplotype have been identified on the DR5-DQ7 haplotype, and fragments of DNA flanking the DQB1 gene of the DR3-DQ2 haplotype have been identified on the DR7-DQ2 haplotype (12, 13). The genetic information in the DQ subregion of the DR3-DQ2 haplotype is thus re-established in DR5-DQ7/DR7-DQ2 heterozygotes, although the sequence information is split between two chromosomes. Susceptibility to celiac disease therefore probably depends on an interaction between at least two genes on the DR3-DQ2 haplotype that are reunited in DR5-DQ7/DR7-DQ2 heterozygous individuals. Theoretically, this gene interaction could involve any HLA-linked genes in the DQ region. However, complete sequencing of an 86 kb genomic fragment spanning the DQ subregion of the DR3-DQ2 haplotype failed to identify genes other than the DQA1 and the DQB1 genes in this region (14). Furthermore, the DQA1 and DQB1 genes are very good candidates in their own right, since their products interact to form an HLA class II heterodimer and since they are situated close to a putative recombination site. This evolutionary consideration along with the fact that most celiac patients share a particular pair of DQA1 and DQB1 genes either located in *cis* or in *trans* are strong arguments that the DQA1*05 and DQB1*02 alleles jointly confer susceptibility to celiac disease by coding for the DQ2 heterodimer.

Almost all the celiac disease patients who are DQA1*05 and DQB1*02 negative bear the DRB1*04, DQA1*0301, DQB1*0302 haplotype (i.e. DR4-DQ8 haplotype), and it

is likely that these patients have an HLA association which is different from those who are DQ2 positive. Although it is less clear what the primary disease susceptibility determinant of the DR4-DQ8 haplotype is, most data favor DQ8 (15).

CTLA4/CD28 genes

Genome-wide linkage studies in celiac disease have indicated numerous susceptibility regions with weak genetic effects, and the indications are strongest for susceptibility genes located at 5qter and 11qter (16, 17). However, no disease associations have been established so far for any genes of these regions. The only gene for which there are relatively consistent reports on disease association is the CTLA4 gene on chromosome 2q33. This gene has a single nucleotide polymorphism (A/G) at position 49 in exon 1, and the A allele has been found to be associated with celiac disease in both French and Swedish/Norwegian populations (18, 19). In the Swedish/Norwegian material there was also a suggestive linkage for several markers and in a Finnish study linkage was found for the marker D2S116, located immediately adjacent to CTLA4 (20). A study of Italian and Tunisian patients failed to confirm the association to the +49 polymorphism or linkage in the CTLA4 region (21). CTLA4 is involved in down-regulating T-cell responses, and the G allele of the +49 dimorphism was recently demonstrated to be associated with reduced control of T cell proliferation (22). Thus the +49 dimorphism is a prime suspect for the observed genetic effect, although other polymorphisms in linkage disequilibrium cannot be ruled out. In this respect, it is worth noting that the CD28 and ICOS genes whose gene products are central players in T and B cell activation are located very close to the CTLA4 gene.

Environmental factors

Celiac patients go into complete remission when they are put on a gluten-free diet and they relapse when gluten is reintroduced to the diet. These observations establish gluten as a critical environmental factor. Whether other environmental factors are also involved is still an open question. Gut infections may well be involved. Adenovirus 12 has been in the forefront among the candidate microorganisms. This virus was originally proposed as a candidate because of partial linear homology over 12 amino acids in a virus protein and an α-gliadin (23). Based on our current understanding of the pathogenesis of celiac disease, the rationale for the candidacy of this virus is weak, and there are in fact very few epidemiological data supporting a role for this virus (24).

Interaction between HLA and gluten

Recent research has demonstrated that celiac disease predisposing HLA molecules directly interact with the environmental factor gluten. T cells of the celiac lesion predominantly recognize gluten peptides presented by the DQ2 or DQ8 molecules, but not by other HLA molecules that the patients express (25, 26). Interestingly, the gluten peptides that are recognized by the T cells are generally modified by deamidation (i.e. by conversion of glutamine residues to glutamic acid) (27, 28). There are a number of distinct T cell epitopes within gluten (29), most of them binding to DQ2 and DQ8 such that the glutamic acid residues created by deamidation are accommodated in pockets that have a preference for negatively charged side chains (15). Recent evidence indicates that deamidation in vivo is mediated by the enzyme tissue transglutaminase (tTG), which is expressed just beneath the epithelium in the gut (30–32). Notably, tTG can also cross-link glutamine residues of peptides to lysine residues in other proteins, including tTG itself. It is possible that complexes between gluten and tTG permit gluten-reactive T cells to provide help to tTG-specific B cells by a mechanism of intramolecular help, thereby explaining the occurrence of gluten-dependent tissue transglutaminase autoantibodies in celiac disease (33).

Conclusions

The emerging picture for the genetics of chronic inflammatory diseases is one of enormous complexity. Even with a complete map of the human genome, it will be a formidable task to map all the susceptibility genes and to come to an understanding of how they interact with each other and with environmental factors. Functional studies based on knowledge of the predisposing effect of HLA genes have contributed significantly to our understanding of the molecular mechanism of celiac disease. This has proved that combined studies of genetics and biology are fruitful and suggests that a combined approach may be the most rewarding also in the future.

Acknowledgements.—Studies in the authors' laboratory are funded by grants from the Research Council of Norway, the European Commission (BMH4-CT98-3087, QLRT-1999-00037), Medinnova, the Jahre Foundation, and EXTRA funds from the Norwegian Foundation for Health and Rehabilitation.

References

1. Risch N. Assessing the role of HLA-linked and unlinked determinants of disease. Am J Hum Genet 1987;40:1–14.
2. Ellis A. Coeliac disease: previous family studies. In: McConnell RB, editor. The genetics of coeliac disease. Lancaster: MTP Press; 1981. p. 197–9.
3. Petronzelli F, Bonamico M, Ferrante P, Grillo R, Mora B, Mariani P, et al. Genetic contribution of the HLA region to the familial clustering of coeliac disease. Ann Hum Genet 1997;61: 307–7.
4. Polanco I, Biemond I, van Leeuwen A, Schreuder I, Meera Khan P, Guerrero J, et al. Gluten sensitive enteropathy in Spain: genetic and environmental factors. In: McConnell RB, editor.

The genetics of coeliac disease. Lancaster: MTP Press; 1981. p. 211–31.
5. Keuning JJ, Pena AS, van Leeuwen A, van Hooff JP, van Rood JJ. HLA-DW3 associated with coeliac disease. Lancet 1976;1: 506–8.
6. 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.
7. Mearin ML, Biemond I, Pena AS, Polanco I, Vazquez C, Schreuder GT, et al. HLA-DR phenotypes in Spanish coeliac children: their contribution to the understanding of the genetics of the disease. Gut 1983;24:532–7.
8. Trabace S, Giunta A, Rosso M, Marzorati D, Cascino I, Tettamanti A, et al. HLA-ABC and DR antigens in celiac disease. A study in a pediatric Italian population. Vox Sang 1984;46:102–6.
9. Tosi R, Vismara D, Tanigaki N, Ferrara GB, Cicimarra F, Buffolano W, et al. Evidence that celiac disease is primarily associated with a DC locus allelic specificity. Clin Immunol Immunopathol 1983;28:395–404.
10. Sollid LM, Markussen G, Ek J, Gjerde H, Vartdal F, Thorsby E. Evidence for a primary association of celiac disease to a particular HLA-DQ α/β heterodimer. J Exp Med 1989;169: 345–50.
11. Carrington M. Recombination within the human MHC. Immunol Rev 1999;167:245–56.
12. Lin L, Jin L, Kimura A, Carrington M, Mignot E. DQ microsatellite association studies in three ethnic groups. Tissue Antigens 1997;50:507–20.
13. Lin L, Jin L, Lin X, Voros A, Underhill P, Mignot E. Microsatellite single nucleotide polymorphisms in the HLA-DQ region. Tissue Antigens 1998;52:9–18.
14. Ellis MC, Hetisimer AH, Ruddy DA, Hansen SL, Kronmal GS, McClelland E, et al. HLA class II haplotype and sequence analysis support a role for DQ in narcolepsy. Immunogenetics 1997;46:410–7.
15. Sollid LM. The molecular basis of celiac disease. Annu Rev Immunol 2000;18:53–81.
16. Greco L, Corazza G, Babron MC, Clot F, Fulchignoni-Lataud MC, Percopo S, et al. Genome search in celiac disease. Am J Hum Genet 1998;62:669–75.
17. Greco L, Babron MC, Corazza GR, Percopo S, Sica R, Clot F, et al. Existence of a genetic risk factor on chromosome 5q in Italian coeliac disease families. Ann Hum Genet. In press.
18. Djilali-Saiah I, Schmitz J, Harfouch-Hammoud E, Mougenot JF, Bach JF, Caillat-Zucman S. CTLA-4 gene polymorphism is associated with predisposition to coeliac disease. Gut 1998;43: 187–9.
19. Naluai AT, Nilsson S, Samuelsson L, Gudjonsdottir AH, Ascher H, Ek J, et al. The CTLA4/CD28 gene region on chromosome 2q33 confers susceptibility to celiac disease in a way possibly distinct from that of type 1 diabetes and other chronic inflammatory disorders. Tissue Antigens 2000;56:350–5.
20. Holopainen P, Arvas M, Sistonen P, Mustalahti K, Collin P, Maki M, et al. CD28/CTLA4 gene region on chromosome 2q33 confers genetic susceptibility to celiac disease. A linkage and family-based association study. Tissue Antigens 1999;53:470–5.
21. Clot F, Fulchignoni-Lataud MC, Renoux C, Percopo S, Bouguerra F, Babron MC, et al. Linkage and association study of the CTLA-4 region in coeliac disease for Italian and Tunisian populations. Tissue Antigens 1999;54:527–30.
22. Kouki T, Sawai Y, Gardine CA, Fisfalen ME, Alegre ML, DeGroot LJ. CTLA-4 gene polymorphism at position 49 in exon 1 reduces the inhibitory function of CTLA-4 and contributes to the pathogenesis of Graves' disease. J Immunol 2000;165:6606–11.
23. Kagnoff MF, Austin RK, Hubert JJ, Bernardin JE, Kasarda DD. Possible role for a human adenovirus in the pathogenesis of celiac disease. J Exp Med 1984;160:1544–57.
24. 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.
25. Lundin KEA, Scott H, Hansen T, Paulsen G, Halstensen TS, Fausa O, et al. Gliadin-specific, HLA-DQ(α 1*0501,β 1*0201) restricted T cells isolated from the small intestinal mucosa of celiac disease patients. J Exp Med 1993;178:187–96.
26. Lundin KEA, Scott H, Fausa O, Thorsby E, Sollid LM. T cells from the small intestinal mucosa of a DR4, DQ7/DR4, DQ8 celiac disease patient preferentially recognize gliadin when presented by DQ8. Hum Immunol 1994;41:285–91.
27. Sjöström H, Lundin KEA, Molberg Ø, Körner R, McAdam SN, Anthonsen D, et al. Identification of a gliadin T-cell epitope in coeliac disease: general importance of gliadin deamidation for intestinal T-cell recognition. Scand J Immunol 1998;48:111–5.
28. Arentz-Hansen H, Körner R, Molberg Ø, Quarsten H, Vader W, Kooy YM, et al. The intestinal T cell response to α-gliadin in adult celiac disease is focused on a single deamidated glutamine targeted by tissue transglutaminase. J Exp Med 2000;191:603–12.
29. Lundin KEA, Sollid LM, Norén O, Anthonsen D, Molberg Ø, Thorsby E, et al. Heterogenous reactivity patterns of HLA-DQ-restricted small intestinal T-cell clones from patients with celiac disease. Gastroenterology 1997;112:752–9.
30. Molberg Ø, McAdam SN, Körner R, Quarsten H, Kristiansen C, Madsen L, et al. Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells. Nat Med 1998;4:713–7.
31. van de Wal Y, Kooy Y, van Veelen P, Peña S, Mearin L, Papadopoulos G, et al. Selective deamidation by tissue transglutaminase strongly enhances gliadin-specific T cell reactivity. J Immunol 1998;161:1585–8.
32. Molberg Ø, McAdam S, Lundin KEA, Kristiansen C, Arentz-Hansen H, Kett K, et al. T cells from celiac disease lesions recognize gliadin epitopes deamidated in situ by endogenous tissue transglutaminase. Eur J Immunol. In press.
33. Sollid LM, Molberg Ø, McAdam S, Lundin KEA. Autoantibodies in coeliac disease: tissue transglutaminase—guilt by association? Gut 1997;41:851–2.