HYPOTHESIS

Human cartilage proteoglycans as T cell autoantigens

J A Goodacre, J P Pearson

Many theories on the mechanisms of rheumatoid arthritis (RA) invoke models whereby pathogenic T lymphocytes are activated by antigens found in joints. The HLA-DR genotype associations of RA suggest strongly that DR molecules are involved in this process, perhaps through their role in antigen presentation in the periphery or in selection of the T cell antigen receptor repertoire in the thymus. This concept is supported by the fact that DR molecules which confer susceptibility to RA in different groups of subjects (DR4Dw4, DR4Dw14,^{1 2} DRw10,³ and DR6Dw16⁴) possess amino acid sequence homology in their third hypervariable regions, a site likely to be intimately involved in the formation of tri-molecular complexes between the DR molecule, peptide antigen and T cell receptor. Interestingly only a minority of subjects with RA susceptible DR genotypes develop RA indicating that other variables (presumably including exposure to the relevant antigens) must also be important. As yet the identity of the antigens which determine RA pathogenesis is unknown and, from the state of current knowledge, there is no reason to presume that only one antigen plays a part. Candidate autoantigens in RA include stress proteins, type II collagen and cartilage proteoglycan aggregates. In this paper the relation between the structure of human cartilage proteoglycan and its antigenicity for human T lymphocytes is considered in the light of recent findings and the potential importance of this relationship for the pathogenesis of RA is discussed.

Proteoglycan aggregates are important for the organisation of tissue matrices and the physical properties of cartilage.^{5 6} With respect to their main structural features aggregates are formed extracellularly by the binding of monomers to hyaluronic acid. These ionic interactions are stabilised by link proteins. Monomers are formed by post-translational processing when many chondroitin sulphate and keratan sulphate side chains as well as N- and O-linked oligosaccharides are attached to the core protein. Chondroitin sulphate chains attach at sites of paired serine-glycine residues. Core proteins consist of several distinct domains.⁷ From the N-terminus, two globular domains (G1 and G2) adjoin a major extended region bearing the keratan sulphate and chondroitin sulphate side chains. The chondroitin sulphate attachment region is subdivided into CS1 and CS2 domains. The core protein C terminal end consists of the G3 globular domain. The G1 domain plays an

important part in forming aggregates but the functions of G2 and G3 domains are unknown.

Evidence that immune recognition of proteoglycans has an important role in arthritis has been obtained in a mouse model in which a destructive polyarthritis develops in BALB/c mice but not other strains of mice following immunisation with chondroitinase ABC treated human fetal cartilage proteoglycans.⁸⁻¹⁰ In this model female mice are most susceptible and the source of proteoglycans is critical such that immunisation with bovine, rather than human, pig, or canine proteoglycans, is not arthritogenic. It has been proposed that arthritogenicity is determined by regions of the proteoglycans not conserved between species, though immunisation with human, pig, and canine proteoglycans primed spleen T cells for responses to mouse proteoglycans, showing that autoreactive T cells were activated in these mice.¹⁰ Transfer experiments using lymphocyte subpopulations suggested that both T and B lymphocytes play a part.¹¹ In humans T cell responses to purified human articular cartilage proteoglycans have been found in some patients with RA,^{12 13} ankylosing spondylitis,^{14 15} and relapsing polychondritis¹⁶ as well as in healthy subjects.^{12 13 17} These studies show that when whole purified proteoglycans are used as an autoantigen there is no simple relation between high levels of T cell responsiveness and disease status. This is also the case for other candidate antigens in RA. For example, although T cells responsive to purified type II collagen have been cloned from arthritic joints,¹⁸ T cell responses to type II collagen among patients with RA are generally low.¹⁹ Similarly *Mycobacterium bovis* hsp65 is not necessarily an immunodominant T cell antigen in RA²⁰ and cells from joints of some adult patients with RA do not respond to human hsp60.²¹ One possible explanation for these findings is that the pathogenicity of some self antigens in joints may be altered by exposure to extracellular proteases and that analysis of T cell responses to enzyme degraded, rather than whole, antigen might be more appropriate. Studies of the properties of defined enzyme cleaved fragments of proteoglycans²² lend support to this view. Furthermore analysis of the amino acid sequence of proteoglycan core protein (aggrecan)²³ in the light of recent information characterising the structural features which determine the antigenicity of other proteins enables regions of proteoglycans particularly likely to contain relevant T cell recognition sites to be identified.

Rheumatology
Laboratory,
Catherine Cookson
Building,
The Medical School,
Framlington Place,
Newcastle upon Tyne
NE2 4HH,
United Kingdom
J A Goodacre
J P Pearson

Correspondence to:
Dr Goodacre.

Accepted for publication
10 March 1992

The concept that enzyme degraded cartilage matrix components have the capacity to stimulate the release of proteases from phagocytic cells, leading to perpetuating cycles of tissue destruction, has been attractive for many years.²⁴ However, it has now been shown that extracellular enzyme degradation also influences the way in which proteoglycan aggregates are taken up and processed by antigen presenting cells.²² Fragments produced from iodine-125 labelled proteoglycan aggregates following degradation with trypsin or stromelysin were purified then tested for uptake by human and murine cells. There were significant differences in the profile of fractions taken up and retained intracellularly from enzyme degraded compared with whole aggregates from proteoglycans. In particular an N-terminal peptide cleaved from link protein was taken up rapidly by human monocytic and B cell lines (but not mouse fibroblasts or peritoneal macrophages), following which it appeared to localise to the cytoplasm rather than to lysosomes. These findings raise the interesting possibility that extracellular degradation of aggregates of proteoglycans by proteases may influence the identity of the peptide fragments made available ultimately for binding to major histocompatibility complex (MHC) molecules in antigen presenting cells and subsequent recognition by T lymphocytes. It is possible that lysosomal proteases would access and degrade fragments differently compared with intact molecules, a factor particularly important for T cell recognition sites dependent on nearby enzyme cleavage sites as found in sperm whale myoglobin.²⁵ This may also determine the extent to which potential T cell recognition sites in aggregates of proteoglycans are destroyed rather than exposed by lysosomal proteases as shown for ovalbumin.²⁶ Furthermore fragments produced by extracellular degradation of proteoglycans might be taken up by antigen specific B lymphocytes with a different range of fine specificities compared with intact proteoglycans, and, as has been shown for tetanus toxin,²⁷ B cell antigen specificity can influence the outcome of antigen processing. In addition to these examples in which processing mechanisms clearly have an important role there are now several examples in which the configuration or 'molecular context' of T cell epitopes has important effects on their antigenicity. In particular, the orientation of epitopes in relation to adjacent sequences²⁸⁻³⁰ and the characteristics of the amino acid flanking sequences on either side of the epitope³¹ have been shown to be important. Although these effects might be determined by processing mechanisms, other possibilities include changes in binding in the trimolecular complex. Each of the observations described here provides insights into how the production of degraded fragments of proteoglycans by extracellular proteases could lead either to epitope destruction or the revelation of otherwise 'hidden' epitopes in proteoglycans, leading overall to altered potential for pathogenicity.

As HLA-DR molecules are implicated strongly in the pathogenesis of RA (and other autoimmune diseases) and are presumably also

important in the induction of protective immune responses against microbial antigens, much effort is being devoted to characterising epitopes presented by DR molecules. It is envisaged that this approach may help elucidate autoimmune disease mechanisms and also help design defined epitope vaccines to protect against infectious diseases. To this end in vitro binding assays have been used to compare the efficacy of many peptides from several antigens for binding to DR molecules. It is important to acknowledge that the extent to which the conditions under which these assays are performed influence the binding between peptides and DR molecules is unclear. A more convincing approach has been used to identify peptides bound to purified mouse class II molecules by eluting and sequencing peptides from the grooves of these molecules.³² Nevertheless DR binding assays appear sufficiently reliable to identify residues in DR β 1 chains involved in binding peptides³³ and to identify a putative DR binding motif.³⁴ From these and other studies the propensity of different regions of human proteoglycans for containing DR restricted T cell recognition sites can be assessed because the link and core protein (aggrecan) genes have been sequenced.^{23 35}

The human aggrecan gene cloned from a chondrocyte cDNA library prepared from a child's costal cartilage consisted of 7137 nucleotides encoding a protein 2316 amino acids long. The G1 domain consists of a sequence homologous with immunoglobulin variable region (1-A) and two consecutive repeat sequences (1-B, 1-B') homologous with each other and with sequences 2-B and 2-B' which form the G2 domain. The G3 domain contains two exons which are alternatively spliced resulting in the less abundant expression of these sequences in transcripts from fetal and juvenile cartilage. Comparison with the human link protein gene sequence,³⁴ cloned from articular cartilage chondrocytes, reveals a strong homology with aggrecan G1 domain. The link protein gene encodes a 354 residue protein consisting of an immunoglobulin variable region fold-like domain and two consecutive tandem repeat sequences homologous with the B sequences in aggrecan G1 and G2 domains.

Apart from the homologous regions in aggrecan and link protein, another prominent feature revealed by these studies is the presence of multiple repeat sequences in the aggrecan keratan sulphate and chondroitin sulphate attachment regions. In the keratan sulphate attachment region, residues 772-837 consist of 11 repeats of a six residue sequence EEP(S or F)PS. Even more strikingly, residues 979-1339 near the beginning of the chondroitin sulphate attachment region consist of a 19 residue sequence repeated virtually identically (EVLET(A or T)APGVEDISGLPSG) 19 times. It has been suggested that the size of keratan sulphate and chondroitin sulphate regions may be related to a greater capacity for load bearing offered by a higher degree of glycosaminoglycan attachment.²³ However, with respect to antigenicity of proteoglycans the overall effect of numerous homologous repeat sequences is to

reduce the potential number of different T cell epitopes in relation to the size of these molecules. That this might influence T cell responsiveness to proteoglycans is suggested by our findings that among congenic strains of mice the magnitude of primed popliteal lymph node responses to purified human cartilage aggrecan was determined by MHC haplotype, despite the size of the antigen.³⁶ The antigenic properties of repeat sequences in other proteins have already attracted much interest, particularly in epitope mapping studies of malarial antigens. Protective immunity to malaria sporozoites can be mediated by antibodies to a major surface antigen, the circumsporozoite protein. A six residue repeat sequence in *Plasmodium yoelii* circumsporozoite protein induced antibody responses in H-2^b but not H-2^d mice, apparently due to activation by this peptide of different helper T cell subsets in these strains.³⁷ Tandem repeat sequences in *Plasmodium falciparum* circumsporozoite protein contain epitopes for B cells (NANP) and T cells (DPNANP-NVDPNANPNV) in mice and humans.³⁸ In mice responses to these epitopes are genetically controlled and dependent on the configuration of the epitopes in the antigen constructs used. The antigenicity of tandem repeat sequences has also been studied using experimental antigens. Lowenadler *et al.*³⁹ made recombinant fusion proteins containing *Escherichia coli* heat stable enterotoxin, which is non-immunogenic, and various numbers of the sequence for peptide 323–339 from ovalbumin, a well characterised helper T cell epitope. Levels of antibodies to enterotoxin were increased using constructs containing multiple tandem repeat 323–339 epitopes, indicating that the presence of multiple repeat 323–339 sequences in the protein increased its capacity for activating helper T cells. This may have been related simply to increasing the number of available T cell epitopes contained in the construct, or to the effects of epitope configuration just described, or to the increased propensity of tandem repeat epitopes for binding to class II molecules.⁴⁰ The extent to which these findings are relevant generally to the antigenicity of other proteins remains to be determined, and in the case of proteoglycans it is not yet known whether T cell repertoires focus on only a limited number of immunodominant regions. Nevertheless these factors suggest that homologous repeat regions in proteoglycans may be particularly important in T cell responses to these molecules.

In native proteoglycans the keratan sulphate and chondroitin sulphate repeat regions are heavily glycosylated. Little is yet known about the part played by carbohydrate side chains in antigen presentation and recognition, but glycosaminoglycan side chains might limit the access of proteases to core protein or the incorporation of T cell recognition sites into trimolecular complexes, or both. The glycosylation status of proteoglycans changes, however, with age and disease,⁴¹ leading perhaps to important changes in proteoglycan antigenicity. In support of this chondroitinase ABC treated proteoglycans are more potent immunogens than whole proteoglycans for inducing the mouse polyarthritis

model.^{8,9} Furthermore some antibodies in synovial fluids from patients with RA recognised the core protein chondroitin sulphate region only after the removal of glycosaminoglycan side chains.⁴⁰ Interestingly several strategies for identifying T cell recognition sites in antigens,^{42–44} including that proposed for identifying DR restricted epitopes³⁴ predict the presence of T cell epitopes in the CS1 repeat region.

To determine the relevance of these observations to the pathogenesis of RA it will be necessary to characterise DR restricted presentation and recognition of both whole proteoglycans and defined regions of proteoglycans looking for RA related characteristics. One aim might be to identify epitopes in proteoglycans recognised by T cells from patients or subjects genotypically susceptible to RA but not by T cells from healthy subjects. As described here, however, there is evidence that high levels of T cell responsiveness may not be the key factor in determining pathogenesis, a view supported by results from animal models in which pre-immunisation with 'pathogenic' antigens can protect against the induction of arthritis^{45–48} One alternative possibility would be that disease is determined primarily by high concentrations of 'self' peptide antigens binding to DR molecules thereby inhibiting normal protective DR restricted responses in susceptible subjects. Whatever mechanisms are involved it is likely that the propensity of certain epitopes for binding to DR molecules plays an important part together with the capacity of the T cell repertoire for responding to these epitopes. The issues discussed here illustrate that the properties of human proteoglycans as T cell auto-antigens and the potential part played by these molecules in the pathogenesis of RA are now amenable to detailed investigation at a molecular level.

We acknowledge gratefully the support of the Arthritis and Rheumatism Council.

- Wordsworth B P, Lanchbury J S S, Sakkas L I, Welsh K I, Panayi G S, Bell J I. HLA DR4 subtype frequencies in rheumatoid arthritis indicate that DRB1 is the major susceptibility locus within the HLA class II region. *Proc Natl Acad Sci USA* 1989; **86**: 10049–53.
- Gao X, Olsen N J, Pincus T, Stastny P. HLA DR alleles with naturally occurring amino acid substitutions and risk for development of rheumatoid arthritis. *Arthritis Rheum* 1990; **33**: 939–46.
- Ollier W E R, Stephens C, Awad J, *et al.* Is rheumatoid arthritis in Indians associated with HLA antigens sharing a DRβ1 epitope? *Ann Rheum Dis* 1991; **50**: 295–7.
- Willkens R F, Nepom G T, Marks C R, Nettles J W, Nepom B S. Association of HLA Dw16 with rheumatoid arthritis in Yakima Indians. *Arthritis Rheum* 1991; **34**: 43–7.
- Ruoslahti E. Structure and biology of proteoglycans. *Ann Rev Cell Biol* 1988; **4**: 229–55.
- Muir H. The coming of age of proteoglycans. *Biochem Soc Trans* 1990; **18**: 787–9.
- Paulsson M, Morgelin M, Wiedemann H, *et al.* Extended and globular protein domains in cartilage proteoglycans. *Biochem J* 1987; **245**: 763–72.
- Glant T T, Mikecz K, Arzoumanian A, Robin Poole A. Proteoglycan-induced arthritis in BALB/c mice. *Arthritis Rheum* 1987; **30**: 201–12.
- Mikecz K, Glant T T, Robin Poole A. Immunity to cartilage proteoglycans in BALB/c mice with progressive polyarthritis and ankylosing spondylitis induced by injection of human cartilage proteoglycan. *Arthritis Rheum* 1987; **30**: 306–17.
- Glant T T, Fulop C, Mikecz K, Buzas E, Molnar G, Erhardt P. Proteoglycan-specific autoreactive antibodies and T lymphocytes in experimental arthritis and human rheumatoid joint diseases. *Biochem Soc Trans* 1990; **18**: 796–9.

- 11 Mikecz K, Glant T T, Buzas E, Robin Poole A. Proteoglycan-induced polyarthritis and spondylitis adoptively transferred to naive (nonimmunized) BALB/c mice. *Arthritis Rheum* 1990; 33: 866-76.
- 12 Glant T, Csongor J, Szucs T. Immunopathologic role of proteoglycan antigens in rheumatoid joint disease. *Scand J Immunol* 1980; 11: 247-52.
- 13 Sigal L H, Johnston S L, Phillips P E. Cellular immune responses to cartilage components in rheumatoid arthritis and osteoarthritis: a review and report of a study. *Clin Exp Rheumatol* 1988; 6: 59-66.
- 14 Mikecz K, Glant T T, Baron M, Robin Poole A. Isolation of proteoglycan-specific T lymphocytes from patients with ankylosing spondylitis. *Cell Immunol* 1986; 112: 55-63.
- 15 Golds E E, Stephen I B M, Esdaile J M, Strawczynski H, Poole A R. Lymphocyte transformation to connective tissue antigens in adult and juvenile rheumatoid arthritis, osteoarthritis, ankylosing spondylitis, systemic lupus erythematosus and a nonarthritic control population. *Cell Immunol* 1983; 82: 196-209.
- 16 Herman J H, Dennis M V. Immunopathogenic studies in relapsing polychondritis. *J Clin Invest* 1973; 52: 549-88.
- 17 Goodacre J A, Middleton S, Patterson A, Ferreira P, Lessan K, Pearson J. Activation of human and murine T lymphocytes by cartilage proteoglycans. *Biochem Soc Trans* 1990; 18: 677-8.
- 18 Londei M, Savill C M, Verhoef A, et al. Persistence of collagen type II-specific T cell clones in the synovial membrane of a patient with rheumatoid arthritis. *Proc Natl Acad Sci USA* 1989; 86: 636-40.
- 19 Desai B V, Dixit S, Pope R M. Limited proliferative response to type II collagen in rheumatoid arthritis. *J Rheumatol* 1989; 16: 1310-4.
- 20 Res P C M, Orsini D L M, Van Laar J M, Janson A A M, Abou-Zeid C, De Vries R R P. Diversity in antigen recognition by *Mycobacterium tuberculosis*-reactive T cell clones from the synovial fluid of rheumatoid arthritis patients. *Eur J Immunol* 1990; 21: 1297-1302.
- 21 De Graeff-Meeder E R, Van der See R, Rijkers G T, et al. Recognition of human 60 kD heat shock protein by mononuclear cells from patients with juvenile chronic arthritis. *Lancet* 1991; 337: 1368-72.
- 22 Martin H, Dean M. A proteolytic fragment from human link protein is taken up and processed by monocytes and B cells. *Biochem J* 1991; 280: 679-86.
- 23 Doege K J, Sasaki M, Kimura T, Yamada T. Complete coding sequence and deduced primary structure of the human cartilage large aggregating proteoglycan, aggrecan. *J Biol Chem* 1991; 266: 894-902.
- 24 Golds E E, Poole A R. Connective tissue antigens stimulate collagenase production in arthritic diseases. *Cell Immunol* 1984; 86: 190-205.
- 25 Van Noort J M, Boon J, Van der Drift A C M, Wagenaar J F A, Boots A M H, Boog C J P. Antigen processing by endosomal proteases determines which sites of sperm-whale myoglobin are eventually recognized by T cells. *Eur J Immunol* 1991; 21: 1989-96.
- 26 Vidard I, Rock K L, Benacerraf B. The generation of immunogenic peptides can be selectively increased or decreased by proteolytic enzyme inhibitors. *J Immunol* 1991; 147: 1786-91.
- 27 Davidson H W, Watts C. Epitope-directed processing of specific antigen by B lymphocytes. *J Cell Biol* 1989; 109: 85-92.
- 28 Perkins D L, Berriz G, Kamradt T, Smith J A, Gefter M. Immunodominance: intramolecular competition between T cell epitopes. *J Immunol* 1991; 146: 2137-44.
- 29 Cox J H, Ivanyi J, Young D B, Lamb J R, Syred A D, Francis M J. Orientation of epitopes influences the immunogenicity of synthetic peptide dimers. *Eur J Immunol* 1988; 18: 2015-9.
- 30 Golvano J, Lasarte J-J, Sarobe P, Gullon A, Prieto J, Borrascueta F. Polarity of immunogens: implications for vaccine design. *Eur J Immunol* 1990; 20: 2363-6.
- 31 Del Val M, Schlicht H-J, Ruppert T, Reddehase M J, Koszinowski U H. Efficient processing of an antigenic sequence for presentation by MHC class I molecules depends upon its neighbouring residues in the protein. *Cell* 1991; 56: 1145-53.
- 32 Rudensky A Y, Preston-Hurlbert P, Hong S-C, Barlow A, Janeway C A Jr. Sequence analysis of peptides bound to MHC class II molecules. *Nature* 1991; 353: 622-7.
- 33 Busch R, Hill C M, Hayball J D, Lamb J R, Rothbard J B. Effect of natural polymorphism at residue 86 of the HLA DR β chain on peptide binding. *J Immunol* 1991; 147: 1292-8.
- 34 O'Sullivan D, Arrhenius T, Sidney J, et al. On the interaction of promiscuous antigenic peptides with different DR alleles. *J Immunol* 1991; 147: 2663-9.
- 35 Dudhia J, Hardingham T E. The primary structure of human cartilage link protein. *Nucl Acid Res* 1990; 18: 1292.
- 36 Goodacre J A, Ferreira P, Lynn S, Middleton S L, Patterson A, Pearson J. Comparative analysis of murine T cell responses to cartilage proteoglycans. *J Autoimmun* 1991; 4: 743-53.
- 37 Grillot D, Martine M, Muller M, et al. Immune responses to defined epitopes of the circumsporozoite protein of the murine malaria parasite, *Plasmodium yoelii*. *Eur J Immunol* 1990; 20: 1215-22.
- 38 Munasinghe D Y, Clavijo P, Calle M C, Nussenzweig R S, Nardin E. Immunogenicity of multiple antigen peptides (MAP) containing T and B cell epitopes of the repeat region of the *P. falciparum* circumsporozoite protein. *Eur J Immunol* 1991; 21: 3015-20.
- 39 Lowenadler B, Svennerholm A-M, Gidlund M, et al. Enhanced immunogenicity of recombinant peptide fusions containing multiple copies of a heterologous T helper epitope. *Eur J Immunol* 1990; 20: 1541-5.
- 40 Sette A, Sidney J, Albertson M, et al. A novel approach to the generation of high affinity class II-binding peptides. *J Immunol* 1990; 145: 1809-13.
- 41 Bayliss M T. Proteoglycan structure and metabolism during maturation and ageing of human articular cartilage. *Biochem Soc Trans* 1990; 18: 799-802.
- 42 DeLisi C, Berzofsky J A. T cell antigenic sites tend to be amphipathic structures. *Proc Natl Acad Sci USA* 1985; 82: 7048.
- 43 Rothbard J B, Taylor W R. A sequence pattern common to T cell epitopes. *EMBO J* 1988; 7: 93-100.
- 44 Sette A, Buus S, Colon S, Miles C, Grey H M. I-A^d binding peptides derived from unrelated protein antigens share a common structural motif. *J Immunol* 1988; 141: 45-8.
- 45 Myers L K, Stuart J M, Seyer J M, Kang A H. Identification of an immunosuppressive epitope of type II collagen that confers protection against collagen-induced arthritis. *J Exp Med* 1989; 170: 1999-2010.
- 46 Billingham M E J, Carney S, Butler R, Colston M J. A mycobacterial 65 kD heat shock protein induces antigen-specific suppression against adjuvant arthritis, but is not itself arthritogenic. *J Exp Med* 1990; 171: 339-44.
- 47 Van den Broek M F, Hogervorst E J M, Van Bruggen M C J, Van Eden W, Van der Zee R, Van den Berg W E. Protection against streptococcal cell wall-induced arthritis by pretreatment with the 65 kD mycobacterial heat shock protein. *J Exp Med* 1989; 170: 449-66.
- 48 Thompson S J, Rook G A W, Brealey R J, Van der Zee R, Elson C J. Autoimmune reactions to heat-shock proteins in pristane-induced arthritis. *Eur J Immunol* 1990; 20: 2479-84.