A PRIMER OF CLINICAL IMMUNOLOGY
(BASED ON A HISTORIC OVERVIEW OF DIAGNOSTIC TESTS)

Professor Robert Clancy AM
Robert Clancy is a Clinical Immunologist who had the Foundation Chair of Pathology in the Newcastle Medical School, and who developed the Hunter (HAPS) Immunology Unit as a regional service in the Hunter. These developments included the introduction of a hospital-serviced allergy programme, a clinical service focussed on autoimmune and related diseases, and the first nurse practitioner – based patient service group (“Scleroderma and Lupus Support Group”). There was always a focus on providing clinical and diagnostic services to regional centres. Of particular importance was establishing quality training programmes in both clinical and laboratory immunology for doctors and scientists alike. He was Chairman of the examining body in Clinical Immunology for a number of years and part of the founding group of ASCIA, the professional body for clinical immunology. A particular interest was career development in immunology for doctors and scientists.

Within the University he established the Hunter Mucosal Immunology Group, which as one project developed an oral vaccine for COPD. Many of the clinical and laboratory staff within the Hunter Immunology Unit became involved in research programmes leading to MSc and PhD awards with valuable career paths in basic and clinical immunology.

The writing of this “Primer” was aimed at making sense of often complex aspects of clinical immunology, through an historic overview of its development. Many of the ideas and developments in this speciality have occurred through the time span of his working life. The “Primer” is produced for the Hunter Immunology Unit as a way of saying thanks to those clinical and laboratory colleagues who has supported and taught him over the last 30 years.
CONTENTS

1. Introduction

2. Immunoglobulins: Back To The Future

3. Antinuclear Antibody: The Key To Systemic Lupus Erythematosus And Autoimmunity

4. Anti-Transglutaminase Antibody Defines Coeliac Disease – But Is It An Autoimmune Disease?

5. Innate Immunity: The Alpha And Omega Of Clinical Immunology

HUNTER AREA PATHOLOGY SERVICE (HAPS)

HAPS was established in 1991 with the amalgamation of pathology services from the Royal Newcastle, Mater Misericordiae and Wallsend hospitals and Lower Hunter and Northumberland regions.

In 2009, HAPS joined with Mid-North Coast Pathology Service (MNCPS), Northern Rivers Pathology Service (NRPS), Pacific Laboratory Medicine Services (PaLMS) and Pathology New England (PNE) to form Pathology North.

Pathology North provides comprehensive public pathology services across general and tertiary referral hospitals. Service provision includes pathology for public patients, private patients, and private hospitals and patients within the community. In addition, Pathology North delivers clinical consultation and clinical research testing.
1. INTRODUCTION

Clinical Immunology in the 21st century is enshrined within the structure of medical practise – it is represented by professional bodies at national and international levels. It was not always so! The current practise of clinical immunology evolved in a staggered fashion, often in conflict with existing organ-based specialities. Historically, clinical immunology has been dated from vaccination against small pox by Jenner in 1796, or immunisation against rabies and anthrax by Pasteur a century later. Indeed use of active and passive immunisation, the discovery of antibody-based hypersensitivity reactions and the development of immunotherapy for allergy by Noon and Freeman in 1911, together with the beginning of blood-typing and diagnostic serology, highlight an emergence of a practise in clinical immunology, but these were little more than a use of ‘immunological methods’ by microbiologists and clinicians to enable management of particular diseases. Modern comprehensive clinical immunology began in the 1960’s. There was a burst of discoveries of autoantibodies and the use of corticosteroids in the 1940’s, but more particularly, there emerged a body of immunological knowledge relevant to autoimmune disease beginning with Burnet’s clonal selection theory and the discovery by Weir of the antinuclear antibody as a simple and useful diagnostic assay, in the late 1950’s. In the 1960’s academic departments appeared with a research focus on the ‘new’ immunology, and many of these had a clinical involvement. In 1976 the Lancet published a report of a meeting by an international consortium, representing the interests of those working with immune disorders. What is important about this group is their link with basic immunological research and the variation of interest amongst this ‘founding father’ group e.g. Roitt from the UK was a PhD scientist with an interest limited to diagnostic immunopathology; Pruzansky from Canada supported a comprehensive clinical strategy ranging from allergy to transplantation immunology; Mackay from Australia focussed on teaching hospital needs in terms of autoimmune disease within a strong laboratory and research context. These founding fathers of the discipline would have a profound impact on the subsequent development of clinical immunology in their respective countries. In most countries clinical immunology would develop from the 1970’s, influenced by the experiences of the founding fathers, the ease of ‘fit’ with existing programmes, and the training requirements of the funding bodies. The two areas of difficulty were the accommodation of allergy with its long and independent history as a medical speciality, and laboratory-based diagnostic immunology with its increasingly complex technology and affinity for non-medical scientists. Post 2000 the increase in knowledge and relevance of immunological mechanisms to a widening spectrum of disease, and most importantly, availability of new therapies such as monoclonal antibody directed against critical pathways in the causation of disease, has encouraged immunological ‘specialisation’ within traditional disciplines.

While the ‘users’ of clinical immunology services have been quick to adopt many of the tests and therapies of the discipline, the ‘body corporate’ of the knowledge base underpinning these values, remains for many, confusing. For this reason I have written a set of ‘primers’, focussed on common assays requested for our immunology services, each constructed from an historic perspective. The reason for this is that the mass of new information – especially within my working lifetime – has come at such a pace and in such a manner, that the evolution of the ideas that shape clinical immunology practise are at best blurred. It is through understanding historic perspective that the ideas, models and terminology of the contemporary clinical immunology, can best be appreciated.

In 2011 the Nobel Prize for Physiology and Medicine was awarded to Ralph Steinman for discovery of the dendritic cell and its role in adaptive immunity, and Bruce Beutler and Jules Hoffman for discoveries concerning the activation of innate immunity following recognition of conserved surface patterns on pathogens. These discoveries relating to innate protection turn traditional ideas on their head. They come with new ideas about disease pathogenesis, and underpin the essential values of the clinical immunologist in interpreting these changes.
to the benefit of the patient, but always in the perspective of an historic context, much of which has happened within one generation.
2. IMMUNOGLOBULINS: BACK TO THE FUTURE

In 1972, Gerald Edelman and Rodney Porter were awarded the Nobel Prize for determining the structure and amino acid sequence of gammaglobulin (IgG). This followed their discoveries through the 1960's:

(i) of light chains and their identity with Bence Jones protein originally described by Henry Bence Jones in 1845 in a patient with myeloma; and
(ii) that antibodies are made of disulphide bond – linked heavy and light chains; and
(iii) that this structure comprised antigen – binding (Fab), and antibody tail (Fc), regions.

These latter discoveries were facilitated when Porter used enzymes to split antibody into regions which could be purified through the capacity of (rabbit) Fc component to crystallise.

These discoveries were a watershed period in immunology, culminating a period of 80 years of dominant immunochemistry, transitioning into a decade marked by Burnett’s concept that cell receptors were attached antibody, an idea that switched focus in immunology from molecules to cells. However, antibodies and their structure would remain a cornerstone of immunology research, a participant broadly in immunopathology, the most recognisable component of diagnostic immunology, and a basic unit of immunotherapy.

The clinical relevance of antibodies was recognised in 1885 when Nuttall demonstrated antibacterial activity in immune serum, and 1890 with the description of antitoxin activity following immunisation with diphtheria and tetanus toxins by Kitasato Shibasaburo. The term antibody was coined in 1891 by Paul Ehrlich when he recognised specificity. Ehrlich subsequently developed his side – chain theory for antibody – antigen interaction in 1897, when he postulated ‘side-chains’ on the surface of cells which specifically bound toxins (antigens) in a ‘lock – key’ fashion. This binding initiated secretion of antibodies. Thus began the era of immunochemistry that would dominate the emerging science of immunology for the next 60 years! It was clear from the beginning that antibodies could neutralise toxins. In 1896 Bordet at the Pasteur Institute demonstrated that a ‘principle’ in serum previously shown to kill bacteria had two components, one heat resistant and the other heat sensitive. The former was responsible for specific (antibody) activity whereas the heat – labile component was responsible for the non-specific antimicrobial activity conferred by all normal serum. The term ‘complement’ was first used by Ehrlich in the late 1890’s to describe this heat - labile non-specific mechanism of protection as it ‘complemented’ specific immune activity, working in conjunction with antibody. In 1904 Wright expanded the idea of specific antibody interacting with “innate” non-specific amplification effector mechanisms, by demonstrating that antibody bound to bacteria to facilitate phagocytosis. This capacity of specific antibody to control and direct both humoral and cellular ‘innate’ pathways – themselves blinded to specificity – would evolve into the cornerstone of protective immunity and (when the coupling of specific and innate mechanisms was deficient) hypersensitivity disease (see section 5).

In the 1920’s Heidelberger and Avery showed that antibodies were protein, and new protein chemistry technologies advanced knowledge of structure. In 1930 Tiselius invented a method of separating protein fractions within a buffer according to charge. Human plasma was separated into α, β and γ globulin fractions. In 1939, Kabat showed most antibody molecules were in the γ fraction. Svedburg had developed an analytic ultracentrifuge that enabled molecular weights of protein molecules to be calculated. Using Svedburg’s scales, the main immunoglobin migrating in the γ globulin region was 7S. Longworth in 1939 reported a spike in the γ globulin region with a height: width ratio of >4 using serum from a patient with myeloma. At that time diagnostic value was limited as the equipment for moving boundary electrophoresis was awkward with a single run taking a full day and interpretation remained difficult. Solid support media were introduced in 1951 with filter paper, followed by cellulose acetate. Separated protein bands were identified using dyes which also allowed for
scanning and quantification. Currently most diagnostic laboratories perform electrophoresis on agarose gel or by capillary electrophoresis, but the principles are the same. Grabor introduced immunoelectrophoresis in 1953 which used precipitation in gels of electrophoresed proteins with specific antisera to identify particular immunoglobulins and their components. Most now use immunofixation for this purpose.

In the 1960’s Mancini developed the process of immunodiffusion which involved diffusion of immunoglobulin through a semisolid medium such as agar or agarose containing antibody against the immunoglobulin to be quantitated. The area of precipitin reaction measured against a standard curve gave the concentration in the test sample. Single radical immunodiffusion has been replaced by automatic methods such as nephelometry in diagnostic laboratories.

In 1944 Waldenstrom discovered a second immunoglobulin in a patient with ‘incipient myelomatosis’ – electrophoresis demonstrated a ‘spike’ in the B-globulin region which had a molecular weight of about one million and a Svedberg value of 19S. Several years later, Waldenstrom reported his ideas on ‘paraproteinaemia’. He postulated that narrow electrophoretic ‘spikes’ had a clonal origin, an observation that would have immense importance to future studies on the structure of immunoglobins, and indeed protein biology in general because of the quantitative amount of clonal protein available for study. Despite this focus on myeloma and monoclonal protein, it was not until 1948 that Fagreous identified plasma cells as the source of antibodies.

In 1965 the WHO published a nomenclature for immunoglobulins, naming the 7S and 19S gamma globulins, respectively, as IgG and IgM. Prior to this, in 1959, Heremans identified a third immunoglobulin as a unique immunoreactive region in the beta and gamma fractions, which he called B2A globulin. In the new WHO nomenclature this molecule was IgA. Tomasi recognised that IgA was the dominant antibody species in mucosal secretions and Cebra showed that Peyer’s patches contained an enriched source of plasma cell precursors that could ‘home’ to the gut mucosa where they secreted IgA for transport into the gut, while Bienenstock recognised IgA as a ‘marker’ of a compartmentalised mucosal system based on selective cell traffic. Unusual paraproteins that did not react in gels with known antisera led to recognition of the structural identity of two additional immunoglobulins: Fahey in 1965 (IgD) and Johannson in 1967 (IgE). The previous year the Ishizaka’s had used protein chemistry on allergic serum to identify the physiochemical properties of ‘reaginic antibody’ as a novel antibody (IgE).

Clinical Significance By 1950 the hard work had been done to provide a framework for the use of immunoglobulin in diagnosis and management of disease – antibodies were understood to protect against infection and these molecules migrated on electrophoresis in the B-globulin region. Indeed “immune serum” had been used to treat diphtheria (1890), tetanus (1892), and other diseases caused by bacterial toxins. By 1910 antisera specific for each of the then recognised four serotypes of pneumococcus were available to treat pneumococcal pneumonia, requiring a shift in diagnosis from the clinic to the laboratory, in order to define which serotype was responsible for the infection (and therefore which specific antiserum to use for therapy). In one sense, this was the beginning of the clinical / laboratory duality that is the keystone of modern clinical immunology. The pressures to provide a stable blood substitute provided by war in the early 1940’s, led the eminent immunochemist Edwin Cohn to develop a plasma fractionation method based on physical characteristics and solubility in alcohol, aimed at the production of albumen. Cohn fractions II and III, though by-products, contained a high concentration of gammaglobulin. As pooled plasma (eventually sourced from 1,000 to 15,000 donors) was used, this fraction contained a wide repertoire of antibody activities, and was immediately found to be valuable in the treatment / prevention of
measles, poliomyelitis and hepatitis A. The corollary was that when patients were identified in the early 1950’s with recurrent bacterial infections and a low level of gammaglobulin, a logical and effective therapy was available as monthly intramuscular injections of gammaglobulin containing wide ranging antibiotic activity. In the 1970’s sugars were eliminated and salt content adjusted, to give a safe and stable product appropriate for intravenous administration to those with a low IgG and recurrent bacterial infection, often with bronchiectasis and chronic lung damage. The current guidelines suggest treatment every 3 – 4 weeks with a dose of about 200mg/kgm, aimed at maintaining an infection – free status with an IgG level greater than 400mg / 100ml. Guidelines are just that, as each individual should be assessed and in some instances a 6 month therapeutic trial can be of value, particularly in those with recurrent infections and IgG subclass deficiency. Little IgA is present in the concentrates, but sufficient to react with anti – IgA antibody sometimes found in those with isolated IgA deficiency and recurrent infection, causing significant clinical adverse effects.

In 1981 in a patient with an agammaglobulinaemia and immune thrombocytopenia treated with replacement therapy, the platelet count increased to normal. This would initiate a new era for the role of gammaglobulin in health and disease (including treatment of Guillain-Barré syndrome, myasthenia gravis and multifocal motor neuropathy autoimmune diseases of the skin such as pemphigus and bullous pemphigoid, cytopenia’s, and Kawasaki’s disease). There is no common feature of this group of disorders and multiple mechanisms may operate. These could involve both the Fab end of IgG (to neutralise cytokines, chemokines and receptors) and the Fc end (through attachment and modulation of receptors especially Fc8R expression). Recent interest has been on a small sialated fraction of IgG required for Fc8R binding (1 – 3% of IgG in commercial preparations) which may account for the high doses required for benefit when treating chronic inflammatory disease (about 2Gm/kg/month). Currently more than 75% of gammglobulin usage is for active therapy, imposing great pressure on an expensive and limited resource. The need to bioengineer an active molecule to replace current ‘blind’ therapy is a high priority. By 1950, the other main area of clinical interest in immunochemistry, that of paraproteins, was established. Paraproteins were recognised as homogenous proteins secreted from monoclonal proliferation of plasma cells. The clinical question became ‘Does this paraprotein reflect a malignant process?’

(i) Immune Deficiency

Over a period of 5 years in the early 1950’s, the three major immune deficiency disorders were described. The genetic and immunological lessons learnt from study of such conditions, led Robert Good to introduce the term ‘experiments of nature’ to emphasise the value of such discoveries to the study of immune physiology. In 1952 Colonel Ogden Bruton described a boy with recurrent pyogenic infections, with an absent 8 globulin band, who would derive clinical benefit from replacement therapy. This condition became known as sex - linked agammaglobulinaemia. The functional defect was a failure of bone marrow pre- B cells to develop into antibody secreting cells. Pre B cells are confined to the bone marrow, so there are no circulating B cells with surface immunoglobulin staining. In 1993 the genetic defect was identified as an absence of Bruton’s tyrosine kinase, an enzyme essential for signalling antigen binding. In 1953 Janeway described what would be named ‘common variable immune deficiency ‘of which over 150 genetic variants are now recognised. Yet only 10 – 20% have an identified genetic defect. Here
the functional defect affects late stage maturation of B lymphocytes and immunoglobulin switching, and control of follicular hyperplasia. The best recognised defect is of TAC1, a transmembrane activator that promotes antibody synthesis. These subjects thus show ‘downstream’ defects with reduced IgG and IgA but do have, surface-immunoglobulin positive B cells in blood. Some subjects have defective T cell function, suggesting impaired ‘help’ to B cells; it is not surprising that about one third of subjects have an autoimmune disease while others developed cancer (especially colonic) or lymphoma. The older idea of Fudenberg that there was a ‘B cell defect → T cell defect → autoimmune / neoplasia ’ transition but with patients presenting at any stage, is too superficial and overriding an idea, but it remains important to remember a risk profile with review including regular colonoscopy. The third of the more common forms of primary immunodeficiency is selective IgA deficiency which was described by Gierdon in 1957, initially as a deficiency in B2A. Selective IgA deficiency is the most common immune deficiency – about 1:500 in the community having this disorder. Only one third of this group are infection – prone, and many of these have an associated IgG subclass deficiency. There is a clinical link with allergy and autoimmune disease, and a high incidence of dental caries and periodontal disease. The potential danger of severe hypersensitivity reactions if given gammaglobulin has been discussed. The pathogenesis of IgA deficiency involves mechanisms similar to those involved in common variable immune deficiency with late development abnormalities involving immunoglobulin class switching and secretion and similar genetic abnormalities such as TAC1 deficiency. Some subjects appear to have an environmental trigger such as a virus infection.

(ii) Paraproteins

Defined as a “spike” migrating between the inter α and post β regions after serum electrophoresis, the monoclonality and link with neoplasia and / or dysfunctional plasma cells has been recognised since Waldenstom’s studies in the 1940’s. 3% over 50 have a paraprotein, a number that increases with age. About 1% these become malignant each year, and given the catastrophic bone damage that can occur if diagnosis is delayed, patients with ‘benign’ paraproteins should be followed at regular intervals. The laboratory characteristics indicating a malignant transformation or a ‘smouldering’ myeloma, include a progressive increase in paraprotein, a reciprocal depression of non ‘M- protein’ immunoglobulins (though common variable immune deficiency with a paraprotein can be confusing) and the appearance of free light chains in the urine. Myeloma is associated with a range of other clinical and laboratory features, though in those subjects with uncertainty, a normal ESR is rarely seen in malignant disease but is usual with benign paraproteins. Primary amyloidosis is a systemic plasma cell dyscrasia with an excess of free monoclonal kappa or lambda light chain deposited in tissues in a beta pleated sheet formation with characteristic staining patterns. It was not till 1971 that Glenner showed that some amyloid fibrils were derived from the variable portion of light chains. Diagnosis combines staining characteristics with identification of free monoclonal light chain in urine and serum usually by immunofixation. Progressive amyloid deposition and tissue destruction in amyloidosis due to light chain disease can be modulated with chemotherapy.
(iii) Monoclonal Antibody Therapeutics

In 1973 Schwaber described human / mouse hybrid cell clones secreting immunoglobulins. Two years later Kohler and Milstein used ‘continuous cultures of fused cells to secrete antibody of predefined specificity’ for which they would receive a Nobel Prize. The enormous discovery of a way to produce ‘designer label’ monoclonal antibody, making available an almost limitless array of diagnostic and therapeutic tools relevant to most disciplines, was a sea change in medical therapeutics. The early hurdle of immunogenicity caused by mouse sequences, was soon overcome with the development of chimeric, humanised, and human monoclonal antibodies. Technical improvements enhanced efficiency and prolonged the half-life of monoclonal antibodies. The two major problems that remain are the capacity to screen safety and efficacy before phase 1 studies in man, and how to ensure focussed delivery to optimise therapy while limiting systemic complications. The danger in man was exemplified in 2006 when a CD28 – specific monoclonal antibody caused a life – threatening “cytokine storm” in healthy volunteers, when the antibody acted as a super agonist in a fashion not predicted by pre-clinical safety testing.

The success of monoclonal immunotherapy is reflected in the licensing of about 30 monoclonal antibody preparations over the last 30 years. The first was OKT3 directed against CD3 on T cells, for acute renal graft rejection in 1986, followed by Basiliximab (1998) that reacts with IL-2R, again to treat acute rejection. In the 1990’s, three important monoclonal antibody activities became available to down-regulate inflammation: infliximab (1998) which is anti-TNF that acts to inhibit cytokine activity and to induce apoptosis in TNF-secreting lymphocytes, had clinical efficacy in Crohn’s disease and rheumatoid arthritis; abciximab (1994) directed against platelet glycoprotein IIb / IIIa, the final common path for platelet aggregation, reduced ‘failed’ angioplasty outcomes by 35%; and rituximab (1997) directed against CD20 on the surface of B cells, induced remission in non-Hodgkin’s lymphoma and suppressed a range of immune disorders dependent on secretion of pathogenic antibodies. A third group of interfering antibodies were developed to control tumour growth – they do this by inhibiting growth promoting factors (such as EGF) or their receptors, or by interfering with tumour – sustaining vascularisation . The best recognised monoclonal used in cancer is trastuzuma which targets an EGF receptor HER2 which is over expressed in about 25% of breast cancer. Immunotherapy increases survival by 25% in late stage disease. Major advances can be expected as the biology and critical pathways in cancer and inflammation become recognised with receptors, enzymes at critical points, and non-redundant mediators, becoming defined.

1. Conclusion

Immunoglobulins have been the cornerstone of clinical and experimental immunology since the Germ Theory in the 1880’s. Between 1890 and 1960 ‘antibody activity and gammaglobulin centred a period dominated by immunochemistry – around the time that their structure was discovered, Burnett postulated that antibody attached to cells as ‘antigen – receptors’, initiating an era of ‘cellular immunology’. However, immunoglobulins remained of importance to all aspects of theoretical and clinical immunology including therapeutic modalities. Perhaps most surprising has been the discovery of an ‘active’ benefit for immunoglobulin therapy in patients with chronic inflammatory disease, beginning a new phase of discovery for an old molecule, as the search begins to find and synthesise novel therapeutics: it is ‘back to the future’ for immunoglobulins. As knowledge of network defects in disease grows exponentially, the use of monoclonal antibody therapy will be tested with equal enthusiasm. This must always, however, proceed with the caution required by a recognition of the dynamic interplay of immune networks, with the consequence of modulation at one point in main not necessarily being capable of predicting, even when ‘safety’ appears documented in animal models. Some unwanted sequelae may only appear after considerable delay.
(a) The Discovery of Antinuclear Antibody:

In 1957 Donald Weir working in the laboratory of Dr E.J. Holborow, published his finding that the antinuclear factor responsible for the LE cell phenomenon, then in use as a cumbersome and insensitive in vitro diagnostic assay for systemic lupus erythematosus (SLE), could be detected in more than 98% of those with this diagnosis using Coons indirect fluorescent antibody technique with either blood smear leucocytes or thyroid tissue sections, as the test substrate. That year similar conclusions were made by Friou and Holman, but it was Weir who would detect the immune nature of 7S antibody combining with nucleosomal antigen (including DNA – at the time not considered to be antigenic) and develop the diagnostic platform that would become the foundation stone of modern clinical immunology. Weir acknowledges as prelude to his studies, refinements in methods for detecting antibody and the demonstration by White of the role of antibody in producing Russell bodies (a tissue equivalent of LE cells) in the rabbit spleen. Review of Weir’s thesis gives an insight into why this period was so important to our understanding of autoimmune disease, as well as the central role that his discovery would be in the development of clinical immunology. First, in establishing SLE as the cornerstone autoimmune disease, ANA serology transformed SLE from being a rare, usually fatal disease of uncertain origin, to a common, often mild, autoimmune disease. By the late 1950’s, a sequence of historic ‘causes’ involving infection, endocrine abnormalities, and hypersensitivity reactions (this based on analogy with vascular lesions found in experimental chronic serum sickness) gave way to the idea of autoimmunity. This was due to the frequent presence of autoimmune cytopenias. However, and as with the earlier theories of pathogenesis, autoimmunity lacked a persuasive ‘big principal’, until Weir’s recognition that high titre ANA defined the disease. He recognised the likelihood that antibody-antigen complexes would induce inflammatory lesions around small blood vessels by analogy with serum sickness which had been extensively studied over the preceding decade, and by a reduction in complement in active disease. SLE almost immediately became the model for study of autoimmune disease in both man and mouse. Waksman described autoimmune thyroiditis as ‘delayed hypersensitivity autoimmunity’ at the same time that Miller was describing thymus-derived (T) lymphocytes and their role in mediating delayed hypersensitivity reactions. Subsequently ‘clusters’ of (cell-mediated) organ-specific and (antibody-mediated) non-organ-specific, autoimmune diseases became recognised.

Second, Weir gave a context to his discoveries by reviewing the evolution of the idea of autoimmunity to that time. The germ theory proposed by Louis Pasteur in the 1880’s with its focus on the specificity of both the cause of disease and of the host response, changed thinking in medicine forever. Serum factors (termed antibody by Ehrlich) were specific markers for both diagnosis and therapy in the laboratories of Pasteur, Kock, and Ehrlich. Ehrlich adapted ‘antibodies’ into his ‘side-chain theory’ and considered ‘auto reactions’ and their control by ‘internal regulating devices’ as physiological, and that a breakdown of such a system would give rise to ‘horror autotoxins’. A widespread misunderstanding of Ehrlich’s ideas, aided by his enduring influence, led to the view that autoimmunity was incompatible with life, holding back acceptance of autoimmune disease despite numerous observations in support of the idea. Syphilis was an infectious disease of great concern in the early part of the 20th century, and in this context, identification of two auto-antibodies proved particularly confusing – cold haemagglutinins in congenital syphilis and cardiolipin antibody as the basis of the main diagnostic tests. The latter was discovered by Wasserman searching for a diagnostic test urgently needed as the spirochaete proved impossible to grow in vitro. He used liver tissue from an infected foetus. It was later found that the antigen was lipid rich and present in normal liver and indeed in many other tissues, leading to the use of the term ‘reagin’ so as not to engage in the ‘difficulties’ of autoimmunity. It was much later that ‘false
biological' positive WR tests were noted, especially in SLE, and correlated by Hughes with thrombotic disease and circulating anticoagulant activity. In the meantime, William Osler, had collated clinical features of 'classical' SLE, to provide a framework for clinical diagnosis. One delayed outcome of Pasteur's human experiments with a rabies vaccine which caused a demyelinating disease in some subjects, was the development in 1933 of a rabbit model of demyelination by injecting animals with spinal cord in an adjuvant, a model that would become of value as experimental allergic encephalomyelitis (EAE) in elucidating mechanisms in multiple sclerosis. In the 1940s observations were made in man that would provide a platform for diagnosis and therapy of autoimmune disease – the LE cell was described by Hargreaves as was the rheumatoid factor and Coombs test – though only the Coombs test was thought at that time to detect an autoantibody – while Hench used corticosteroids to suppress inflammation in subjects with rheumatoid arthritis. As momentum for the idea of autoimmunity developed in the late 1940s the main theory of antibody production involved antigen 'instructing' antibody by either acting as a template or a scaffold for antibody formation, which was an awkward 'fit' with autoantibodies. Burnet suggested that 'self-markers' existed, which in some way inactivated auto-reacting antibody-secreting cells. In the period immediately before Weir's description of ANF, three important observations were made:

(i) The LE cell became of great interest and was defining the clinical spectrum of SLE – it was also detected in subjects taking hydralazine and in others with different diagnoses especially rheumatoid arthritis and chronic liver disease; heteroantibody could react with damaged nuclei to form LE cells in vitro and the LE cell ‘factor’ could cross the placenta.

(ii) Witebsky described an animal model of autoimmune thyroiditis by injecting thyroid tissue with an adjuvant.

(iii) A sea-change came in understanding antibody formation in 1955 when Jerne postulated that antigen ‘selected’ antibody from an extensive repertoire. This ‘selection’ would in some way induce cells to secrete specific antibody.

Third, Weir provided a human model of autoimmune disease based on objective diagnostic criteria, with a database involving assays on nearly 1000 subjects. Importantly he linked the pathology of SLE as an aberration of a physiological process. This contribution has largely been forgotten, accounting for the difficulty in understanding the emphasis in SLE on autoimmunity directed against exposed determinants in the nucleosome. He noted that serum from disease other than SLE (and in some normal serum) low affinity 19S (IgM) ANA could often be detected. These antibodies were studied in a model of carbon tetrachloride-induced liver damage, and the hypothesis developed that low affinity IgM antibodies were produced in response to cell death, to facilitate clearance of necrotic tissue. The concept that SLE represents a broad-based production of autoantibody directed against exposed determinants of the nucleosome fits with the idea that SLE represents largely a generic reduction in tolerance with respect to immune-based ‘debris-clearance mechanisms’ with a focus on nuclear material. This represents a T cell-induced switch from low affinity IgM ANA to a high titre IgG high affinity ANA that binds complement to form pathogenic immune complexes. The popular view on autoimmunity by 1960 was Burnett's development of his selection theory encompassing clonal selection. He postulated that 'antibody' was attached to antibody-secreting cells as a specific antigen receptor. This would dramatically change immunology from a humoral to a cell-based research discipline. He postulated that 'self-reacting' immunocytes contacting antigen before birth were eliminated/inactivated, and that autoimmunity occurred when 'forbidden clones' arose due to an accumulation of random events (or mutations) the number of which could be predicted from the slope of age-specific prevalence curves. Collectively this cluster of ideas and observations stimulated clinicians such as Damaskek, and Dubois to promote the idea and document large patient groups,
followed by Witebsky and Mackay, taking a collective message and listing ‘markers’ for
diseases thought to have an autoimmune basis.

(b) The Next Fifty Years:

(i) Clinical Disease:

In 1960 SLE was a rare disease with LE cells and with an eighty percent five year mortality.
The routine use of ANA as a screening diagnostic test re-defined SLE as a common mild
disease with over 90% surviving ten years. Fifty years ago most deaths were due to renal
disease or the complications of high dose corticosteroid therapy. Because of the great range
of potential clinical features, disease descriptions reflected the interests of the physicians
involved, until the appearance of clinical immunology with autoimmune disease as a central
plank of this specialty. For example renal physicians instruct that 30-50% of all patients have
clinically relevant nephritis when the real incidence is less than ten percent while
rheumatologists developed diagnostic formulae useful for research but restrictive and of little
value to the clinician. When significant nephritis occurs it usually does so early in the course,
shaping the disease and its management thereafter. The evolution of management
strategies for renal lupus is a blueprint for more severe disease in general (see Table 1).
Patients present with asymptomatic microscopic haematuria and/or proteinuria, or acute
renal failure or nephritic syndrome, require a renal biopsy to define pathology. This in turn
influences management decisions. In 1961 Pollack and colleagues showed high dose
corticosteroid therapy dramatically improved renal function and survival, though many
continued to progress to end stage renal failure. In the 1970s and 1980s cytotoxic drugs – in
particular azathioprine and cyclophosphamide – were introduced as adjunct therapy for
diffuse proliferative glomerunephritis. The often small, poorly controlled, variable and short
term studies were difficult to analyse, but later meta-analyses showed clear benefit. Isolated
studies followed comparing azathioprine and cyclophosphamide, then oral versus pulse
intravenous cyclophosphomide with subsequent dose modification to pulse
corticosteroid/cyclophosphamide to improve the safety margin. Most recently mycophenylate
mafetil has been shown to provide a safety/efficiency advantage over pulse
cyclophosphamide in inducing remission. As about one third of subjects relapse after
achieving remission, maintenance therapy with azathioprine became routine to reduce
relapse rate – current evidence indicates that again mycophenylate has a better
safety/efficacy record for maintenance therapy than does azathioprine.

Those with mild-moderate disease can have any mix of a myriad of features with the most
common listed in Table 1. The symptom of concern to most is fatigue, while cognitive
defects are often neglected in favour of organ-specific symptoms such as joint pain, skin
rashes and pleurisy. Few diseases require better integration between the family physician
and specialist clinic than does SLE. The role of the latter is to monitor and intervene as
necessary in a chronic relapsing and remitting disease, while providing an educational and
support resource geared to patient’s needs. There is evidence that such support by itself
reduces exacerbation rate while improving the quality of daily living. Assessment of level of
disease activity (as per table 1) provides useful guidelines for treatment. Those with mild to
moderate disease often benefit with a 2 month trial of hydroxychloroquine (plus or minus
symptomatic therapy including NSAIDS) monitored with patient-administered visual
analogue scales. High titre ANA is the central plank for diagnosis, with most laboratories
using commercially available cell lines for quality control and to determine the ‘pattern’ of
fluorescence. Patterns of fluorescent staining reflect the presence of particular anti-
nucleosomal antibodies. Anti-DNA antibody gives a pattern of homogenous nuclear staining
and was developed as a specific test using the method of Farr in 1959 and is more specific
than ANA and associated with more severe disease. In 1971 Sharp reported an RNA-
containing antigen (mix) that could be extracted in saline (hence ‘extractable nuclear antigen’
or ENA) that correlated with ‘mixed connective disease’ where features of SLE combined with those characteristics of limited scleroderma and polymyositis. A small proportion of this group will develop pulmonary hypertension which can be missed as a cause of breathlessness. As many improve with endothelin receptor blockade and phosphodiesterase type 5 inhibitors, it becomes important to detect pulmonary hypertension by echocardiography. Today ‘ENA’ results identify a family of antibodies. Statistical linkage of these different antibodies with clinical patterns has limited value; perhaps the strongest association is the presence SSA/antiRo and SSB/antiLa antibodies with Sjogren’s syndrome. The most important additional antibody activity is cardiolipin antibody (CLA) which promotes thrombotic episodes. Historically cardiolipin antibody was detected by the Wasserman Reaction and known as ‘reagin’ (above) then recognised as directed against lipid-containing (mitochondrial) auto-antigen. As syphilis became less common ‘false’ biologically positive CLA tests became more frequent, especially in SLE. In 1975 Hughes noted a link with thrombotic events (including late miscarriages). Antibodies most likely to cause thrombosis also bind B2 glycoprotein1; some of these antibodies are responsible for the in vitro finding of a ‘circulating anticoagulant’. Those with a history of thrombosis require protracted anticoagulation, while those with a history of miscarriage benefit from subcutaneous heparin through subsequent pregnancies.
### Table 1.

<table>
<thead>
<tr>
<th>Disease Variants</th>
<th>Type 1 disease (40%) (mild)</th>
<th>Type 2 disease (50%) (moderate)</th>
<th>Type 3 disease (10%) (severe)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical Features</td>
<td>- Fatigue</td>
<td>as for type 1, but failed therapy trial (Hydroxychloroquine for 2 months)</td>
<td>Significant vital organ involvement, e.g. renal, CNS Failed steroid/ cylotoxic drug therapy</td>
</tr>
<tr>
<td>- Cognitive defects</td>
<td>- Organ-based symptoms e.g. joints, skin, etc Raynaud’s, cytopenia</td>
<td></td>
<td>- Cardiolipin antibody syndrome (CLA)</td>
</tr>
<tr>
<td>- Urine (sediment and protein)</td>
<td></td>
<td></td>
<td>- Mixed connective tissue disease (MCTD)</td>
</tr>
<tr>
<td>Diagnostic Tests</td>
<td>DNA antibody</td>
<td>DNA antibody</td>
<td>DNA antibody</td>
</tr>
<tr>
<td>- ANA (≥1:320)</td>
<td></td>
<td>Renal biopsy</td>
<td>- CLA and β2GPI antibody and circulating anticoagulant</td>
</tr>
<tr>
<td>- C3; C4↑</td>
<td></td>
<td></td>
<td>- ENA, CPK</td>
</tr>
<tr>
<td>- γ globulin↑</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Therapy</td>
<td>Support clinic</td>
<td>Corticosteroids (plus steroid sparing drugs: e.g. Azathioprine, Methotrexate, Mycophenolate)</td>
<td>Pulse corticosteroids/ cylotoxic (Cyclophosphamide or Mycophenolate)</td>
</tr>
<tr>
<td>- Hydroxychloroquine</td>
<td></td>
<td>Consider belimumab if Prednisone 8mg/day</td>
<td>Follow with maintenance: Azathioprine OR Mycophenolate</td>
</tr>
<tr>
<td>- NSAIDs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comment</td>
<td>- Most subjects have stable relapsing disease with ‘type’ established in first 12m</td>
<td>Main aim is to control with minimum steroid usage, and to monitor steroid complications (e.g.) keep at prednisone 7mg/day</td>
<td>Failed therapy: Consider cytotoxic ablation therapy to ‘re-start immunosstat’ with autologous stem cell replacement</td>
</tr>
<tr>
<td>- Support environment critical</td>
<td></td>
<td></td>
<td>Combination monoclonals</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Add tacrolimus</td>
</tr>
</tbody>
</table>

(ii) **Pathogenesis:**

Changes in ideas on pathogenesis of autoimmune disease in general and SLE in particular have evolved through five overlapping phases, themselves mainline to the development of immunology. They are: the thymus, T lymphocytes, cytokines, innate immunity and immunogenetics. Burnett in 1960 had set the bar for this period with his clonal selection theory and ‘forbidden clone’ concept of autoimmunity, and through his concept of cell-associated antigen receptors, kick-started the era of cellular immunology which would dominate immunology theory until the limitations of in vitro experiment were exposed by the in vivo realities of the knock-out mouse.
The thymus as a site for the generation of Burnet’s ‘forbidden clones’ of antibody-secreting cells was favoured in the early 1960s due to evidence of B cell activity such as germinal centre formation in the thymus of subjects with autoimmune disease and of mice with SLE. A flurry of interest in thymectomy in both subjects and mice with SLE gave disappointing results, possibly due to timing (i.e. thymectomy may need to be neonatal to be effective). The ‘thymic’ phase morphed into the ‘Tcell’ phase when Jacques Miller showed in 1961 that neonatal thymectomy in mice led to a failure to develop ‘cellular’ immune responses of the delayed-hypersensitivity type. Miller extended the role of Tcells to a regulatory function in 1968, with the pivotal observation that they provided ‘help’ to B lymphocytes secreting antibody. When two years later Gershon described T cells that could ‘suppress’ immune responses, a new framework for control of immunity (including the prevention of autoimmunity), had been established. There followed a series of major discoveries related to control mechanisms and the maintenance of self-tolerance, that are critical to a modern understanding of the pathogenesis of autoimmunity. These include thymic education of T cells and negative selection as a mechanism of central tolerance, and the dual specificity of T cells for structures coded by the major histocompatability gene complex (MHC) as well as for antigens. In man, these ‘structures’ are the HLA class 2 molecules on antigen presenting cells (APCs) involved in regulation of the development of a Tcell immune response, and HLA class 1 molecules on all nucleated cells, involved in control through ‘self-recognition’ of targets by cytotoxic Tcells. More recently, the Gershon concept of Tcell ‘suppression’ has evolved with the discovery of Treg cells, developed within the thymus but acting peripherally to down regulate autoimmunity, as well as other feedback loops of uncertain importance. Cytotoxic drugs such as azathioprine and cyclophosphamide were introduced to suppress autoimmunity in the 1960s and 1970s by inhibiting cell division and function. Their value in steroid-sparing and in disease control is in a sense remarkable given first, the complexity of the defects in autoimmunity with only vague connections to cell proliferation, and second the lack of specificity inherent in the cytotoxic drug effects.

Perhaps more surprising has been the growing evidence that innate immune mechanisms have an important part to play in both central and peripheral tolerance, and in the development of autoimmunity. A major mechanism involves dendritic cells – the most potent APC, first identified by Steinman in 1972. These cells influence negative selection within the thymus as well as tolerance to auto-antigens in peripheral tissues. Particular interest in SLE has been the role of toll-like receptors (TLRs) on the surface of dendritic cells (especially TLR-9) in activating a response to DNA. Hydroxychloroquine long used to down regulate SLE disease activity interferes with TLR-9 binding to DNA as well as having a more general effect on antigen-presentation by raising the pH within lysosomes of APCs.

Study of cells contributing to autoimmune disease became focussed on the soluble proteins secreted by these cells to direct activities and outcomes. These molecules are generically known as cytokines. Those of particular interest in autoimmunity are the subset secreted from lymphocytes known as lymphokines. The first to be identified was IL-2 in 1965, promoting growth, differentiation and survival of Tcells, but it was not until 1981 when the IL-2 receptor was recognised that immune regulatory networks began to take shape. IL-2 has a particular role in preventing autoimmunity by promoting generation of Treg cells within the thymus. Cytokines entered a new phase in relation to both understanding mechanisms of autoimmunity and in maintaining self-tolerance in 1975 when Kohler and Millstein developed the technology to produce monoclonal antibodies that could define specific pathways and initiate a new era of specific immunotherapy. The first targeted biological treatment for SLE was BLYS-specific inhibitor, a humanised monoclonal antibody that enables auto-reacting B-lymphocytes to undergo apoptosis, thus preventing escape of autoimmune B-cell clones, with the consequence of reducing DNA antibodies and hyper- gammaglobulinaemia. This leads to an increase in complement levels reflecting a reduction in formation of pathogenic immune complexes. However, clinical benefit was only moderate. This represents the first new treatment for SLE recognised by the FDA in fifty years, and it can be expected that
some of the many monoclonal antibody preparations currently being investigated targeting points of defective regulation in SLE will have roles in management. Considerable redundancy exists within these cytokine networks, and great care needs to be taken in trialling agents based on in vitro study and theoretical concepts as ‘logical’ intervention can give no benefit (e.g. reduction of CD20 B cells in SLE with retuximab) or create unexpected and inappropriate responses known as a ‘cytokine storm’ (e.g. TGN1412 – a monoclonal antibody that binds CD28, a co-receptor for the Tcell receptor). Key molecules acting within the complex immune-regulatory pathways currently of interest in SLE, include IL-6 (which promotes Th17 activation in SLE ‘flares’), anti C5 (to reduce complement activation), IL-17 and IL-23 (Th17-derived cytokines mediating nephritis) and even inhibitors of certain kinases over-expressed within Tcells of patients with SLE have caused down-regulation of disease activity in mice with SLE.

Genetic factors are important in the development of SLE. Rare patients have single gene defects (e.g. the complement components C1q and C4), while the overwhelming majority of patients depend on the combined effect of variants of a larger number of genes. Those single-gene defects that have been defined target critical pathways such as the generation of negative selection within the thymus (C4 deficiency) and the elimination of necrotic material (C1q deficiency) whereas most of the more common single-nucleotide polymorphisms (SNPs) detected in SLE involve the non-coding DNA regions of genes related to control of the immune response, in particular at the level of MHC but also downstream involving Tcell function at the level of immune signalling pathways affecting Tcell and Bcell receptors, co-stimulating molecules and receptors, pattern-recognition receptors on dendritic cells such as TLRs and other innate immune molecules. This profusion of gene defects discovered by genome-wide association studies now identifies about 200 gene loci in autoimmune diseases with about 8-9 consistently found in SLE. Cumulative effects of a combination of genetic defects has been explained in terms of the idea of quantitative thresholds for immune-cell signalling, where multiple genetic defects of small individual effect can affect susceptibility to SLE. These newer studies also focus on the contribution of epigenetic change induced by the environment in somatic cell DNA, again a view consistent with clinical observations. Despite the extraordinary profusion of genetic data now available, only about 15% of the heritability of SLE has been identified by the loci studied.

(c) Conclusion:

The last 50 years has been a remarkable period for SLE. At a clinical level the discovery of ANA by Donald Weir in 1957 changed thinking in SLE and provided the tool to define lupus as a common mild disease while providing a focus for research that would springboard from the scientific framework of the day. Somewhat surprisingly, management has changed little in terms of the drugs used, but there is significant improvement in outcome as clinicians have learnt to use these drugs more effectively and with less toxicity, while modifications of these drugs have continued to improve safety margins. The period of focus on cellular immunology began to clarify the ‘black box’ of autoimmunity, but it has been recognition of the molecular basis of immune control involving cytokines, that has enabled production of specific immune therapies. Currently the first of these drugs to be accepted by a regulatory body is the monoclonal antibody anti BLyS, which targets Bcells – while of only moderate clinical benefit, it marks the pathway for the future. The very profusion of defects, the individual variations, the new genetic concepts of ‘quantitative thresholds’ of immune signalling as well as redundancies within these systems ensure that the pathway will be rocky and that new lights will be needed to maximise therapy for SLE in the future. One such ‘light’ promises to be combining the disciplines of genetics, molecular chemistry, cell biology with classical clinical documentation, to perhaps individualise therapy and restore functioning self-tolerance. Autosensitisation in SLE can precede clinical disease by decades, and a combination of screening (with ANA) and intervention to prevent the Tcell-dependent ‘march’
to high affinity ANA and pathogenic complexes, maybe one window for the future.

Environmental influences effecting genetic change in somatic DNA (epigenetics) will be a major area of investigation, seeking to identify and neutralise factors that can accelerate this ‘march’. An important concept initiated by Weir was that SLE represented a breakdown of physiological autoimmune process of ‘debris removal’ involving low affinity anti-nucleosomal antibody, brings focus onto the important role dendritic cells and ‘innate’ immunity has on the development of autoimmunity. In brief, the pathology of SLE is the outcome of two linked and genetically influenced events each reflecting a breakdown in the physiology of debris clearance: a net loss of tolerance to nucleosomal determinants, and a level of pathogenic immune complexes reflecting balance of production and clearance in excess of a critical threshold (Fig 1). The therapeutic challenge for the future is whether prevention as reversal of defective tolerance and/or reduction of pathogenic complexes in excess of ‘disease threshold’, will come from targeting one or more of the genetically determined defects that underpin abnormal pathways, or through the development of unique and specific immunotherapeutics that override dysregulation. For the present, management requires a holistic approach based on support and tailored intervention which is monitored, based on an informed family practitioner with the backing of units geared to the needs of patients with autoimmune disease and capable of filtering new information and therapies. The availability of specialist nurse practitioners to provide ongoing understanding and support, does much to facilitate patient involvement in decision-making leading to less active disease.
Figure 1.

Antinuclear Antibody: the Keystone of SLE
(a) Introduction

Coeliac disease (CD) and systemic lupus erythematous (SLE) are both defined by an autoantibody, but there similarities fade. In SLE antinuclear antibody (ANA) is the cornerstone of diagnosis because of its high sensitivity and clear role in pathogenesis. Because SLE (and ANA) is an outcome of the physiological process of nuclear debris clearance, diagnosis requires a high titer of antibody and additional laboratory and clinical data. By contrast, the autoantibody IgA anti-tissue transglutaminase (aTtG) with a sensitivity and specificity for CD of over 90%, is essentially diagnostic of CD. This test (and its precursor anti-endomysial antibody) has transformed CD from an uncommon disease of children presenting with malabsorption, abdominal distension and failure to thrive, to a common disease with a prevalence of around 1% but with wide regional variations, presenting at any age often with subtle malabsorption or even systemic disease (though only about 20% are clinically recognised). Surprisingly there is little evidence that aTtG or autoimmunity play a significant role in pathogenesis or indeed has any identifiable role. While ANA was described at a critical and early stage in our understanding of the pathogenesis of SLE and was always a robust component of the sequence of events that underpinned clinical disease, aTtG was first described only 14 years ago and then within a well understood clinical and pathogenic framework. These considerations lead to the question of whether autoimmunity is cause or consequence of CD. What is not in dispute is that CD is a chronic inflammatory disease of the mucosa of the upper small bowel, dependent on exposure to gluten. Histologically the gut lesion is characterised by a lymphocytic infiltrate of the epithelium and lamina propria, crypt, hyperplasia, and a loss of villus structure. This pathology was recognised by the early 1950’s, though the appearance of these changes following gluten challenge and an identity of lesions between childhood CD and adult ‘non-tropical’ sprue was subsequently established with the use of swallowed biopsy capsules. Observations on cell kinetics of normal epithelium show cells persist for several days. In CD this period is shortened to a few hours due to gluten-dependant damage. The normal epithelium includes intra-epithelial lymphocytes at between 6 – 40 per 100 epithelial cells. In 1971 this number was shown to be increased in CD, an increase that was gluten – dependant and was the first morphological change following gluten challenge. Several observations underpin our current understanding of the pathogenesis of CD:

![Autoimmunity Diagram]

**Autoimmunity**

Gluten is a toxin (1930’s) → Genetic predisposition (1970’s) → Mucosal T cell activation (1970’s) → Failed immune homeostasis → Disease
1. **Gluten as a Toxin**

From the time coeliac disease was described by Samuel Gee in 1888, diet of one form or another has been the prescribed therapy. An important study of eight subjects in 1924 by Haas showed remission following a banana diet, with the fruit and vegetable diet of Fanconi also popular. However, the breakthrough came from Dutch paediatrician Willem – Karel Dicke working in the Hague, who began a series of dietary experiments between 1934 and 1936 following comment from parents linking symptoms to ingestion of bread and rusks. He reported his finding in 1941, a time of privation due to World War II that exacerbated the benefit of gluten deprivation, of clinical improvement on a simple diet avoiding wheat. He showed a close correlation between retarded growth curves and high faecal fat excretion, and the presence of wheat or rye flour in the diet. Subsequently the ‘toxic’ component of wheat flour was shown by clinical challenge to be gluten, the main storage protein of wheat, rye and barley. Further, most toxic peptides were proline–glutamine–rich, and found in the alcohol – soluble fraction of gluten known as gliadin. These peptides are resistant to complete proteolysis by gastric, pancreatic, and brush–border – associated enzymes. An important ‘residual’ peptide is a 33 mer polypeptide derived from gliadin. The incompletely digested glutamine – rich peptides are absorbed across the epithelium. tTG is found beneath the brush border, where it binds the peptides and catalyses deamination of glutamine residues generating negatively charged glutamate as well as cross-linking gluten peptides with matrix proteins, which helps to retain ‘toxic’ peptides within the mucosa. A gluten threshold for mucosal damage of 10 – 50 mg per day is usual, with a slice of bread containing 105 gm of gluten. In 1970, the pathogenesis of CD was considered likely to involve mucosal damage due to exposure to toxic gluten –derived peptides. However, in the early 1970’s two observations would change thinking of the pathogenesis of CD from a biochemical defect to one of immune dysregulation. They were (i) susceptibility associated with HLA genes within the MHC, and (ii) mediation of mucosal damage by gluten-specific T cells.

2. **Genetic – Predisposition**

Genetic factors in CD had long been recognised, with 10% of asymptomatic first degree relatives having CD. In 1972 groups both sides of the Atlantic described an association of CD with the histo-compatibility antigen HL-A8, thought at the time to be a marker of an ‘immune response gene’ through linkage disequilibrium, as was being described for microbes, food antigens and allergens.

Subsequent studies on the genetic region coding for Class II antigens, have shown that for practical purposes CD does not occur in the absence of alleles that encode for HLA – DQ2 or HLA – DQ8 proteins. These two HLA genes are relevant to the generation of an immune response confirming the earlier idea of linkage. It is now recognised that these Class II molecules are present on the surface of antigen – presenting cells within the mucosa where they bind and present specific gluten peptides such as the 33 mer gliadin peptide in association with tTG. While this genotype is required for CD, it is not sufficient as 30 – 40% of the population have these alleles. Twin and sibling studies have also shown that HLA genes only provide 30 – 40% of the genetic contribution to CD. One unexpected finding was that instead of the CD HLA genotype being subject to negative selection as is found with lactose intolerance, the opposite was observed. This suggests a balanced polymorphism with an inherent benefit provided by the CD genotype to the population, perhaps through enhanced protection against pathogens. Recent genome–wide association studies are identifying non-HLA genetic loci but to date only about 5% of the additional genetic risk has been found. Many of these genetic loci are also risk factors for autoimmune diseases especially type 1 diabetes.
3. **Gluten–sensitive T cells secrete enteropathic factors**

Between 1972 and 1976 a series of studies in animal models and in CD by Anne Ferguson in Edinburgh, led her to state with respect to CD and food allergy “the clinical, pathological and immunological features of the two diseases are so similar that the conclusion that both are due to hypersensitivity to foods is inescapable!” Animal studies showed that activated T cells had the capacity to induce mucosal changes characteristic of CD in the mouse, following both intestinal allografts and induction of graft–versus–host reactions due to injected lymphoid cells, with lymphocytic infiltration of the lamina propria and epithelium, followed by an increase in length of the crypts of Lieberkühn. Addition of gliadin to cultured small bowel biopsies from CD stimulated secretion of the cytokine MIF and a cytotoxic factor – neither was observed in supernatants of similarly treated control biopsies. These findings were interpreted as evidence that gluten–associated epithelial damage was a consequence of antigen (gluten)–induced T cell activation, rather than any direct toxic effect of gluten peptides. These core observations were extended to identify gluten–specific CD4 + T cells secreting predominantly INF-γ, with an IL - 21 dependant loop. An inappropriate activation of these cells would explain the early presence of neutrophils within the mucosae. The relevant gluten epitopes were better recognised by mucosal T cells following deamination by TG. Finally retained gliadin peptides activate innate immune pathways within the epithelium secreting IL-15 which further activates the expanded intraepithelial lymphocyte population to secrete the toxic factor NR-15.

4. **Autoimmunity or hypersensitivity**

Whether CD is an autoimmune disease with gluten as a significant environmental trigger, or a hypersensitivity disease (as envisaged by Ferguson) due to an excessive and inappropriate pathogenic response to environmental antigen, or a mix of both mechanisms, remains a central and largely unanswered question. In 1961, Truelove's group in the UK described anti-gluten antibody in CD, and later claimed that anti–gliadin antibody was specific. However, the presence of antibody to other foods, and detection of anti-gluten antibody in other gut disorders, clouded clinical value for gluten / gliadin antibody as a diagnostic tool.

Similar findings for T cell responses to gluten and other food antigens were noted, and T cell responses to gliadin had a greater specificity. More recent data of IgA antibody to immunodominant deaminated gliadin peptides claim diagnostic specificity. Taken together these observations are consistent with a regular synthesis of IgG class antibody to gluten and other food antigens following any mucosal damage, but the gut mucosal IgA response to selected deaminated gliadin peptides in CD reflects focussed binding to HLA – DQ2 / DQ8 molecules on mucosal antigen presenting cells.

In 1971, two observations began an era when CD would be considered, at least in part, to be an autoimmune disease. First, a clinical and serological link with ‘classical’ autoimmune disease was noted. Second, anti-reticulin antibody reacting with connective tissue was found in 70% of untreated CD (and 30% of those on a gluten – free diet). The timing of these observations was of interest, as it coincided with the argument against specificity of anti-gluten antibodies.

The question became: was anti-reticulin antibody cross-reactive with food antigens or was it a true auto–antibody? Doniach attempting to resolve this question described five immunofluorescent patterns reacting as ‘antireticulinantibody’, one of which (ARR-R1) would later be found to be a true auto-antibody (anti-ITG). The question remained unresolved.

The game-changer came in 1983 when Chorzelski described a new antibody directed against connective tissue surrounding smooth muscle, using primate oesophagus as substrate. The value of this antibody (called endomysial antibody) was confirmed in 1992
when Ferreira showed its diagnostic value with a sensitivity of 90%. The antigen reacting with endomysial antibody was found to be αtTG in 1997, which technically enabled development of automated assays using synthetic antigens. Thus serology had evolved with αtTG becoming a defining diagnostic test, and an unquestioned autoantibody. The question now is, ‘does αtTG antibody contribute to tissue damage in CD?’ The answer is unclear as evidence exists suggesting that this autoantibody can promote damage, but also that it may protect. Perhaps the most likely answer is that αtTG selectively reflects enhanced presentation of tTG together with its deaminated substrate to T cell receptors in association with HLA – DQ2 / DQ8 proteins, and that links with more defined autoimmune diseases such as type 1 diabetes, reflect shared genetic backgrounds.

What has not been adequately explored is CD in the context of current studies of self-tolerance, is the issue of defective mucosal control. This may be a consequence of impaired negative selection and the generation of natural regulatory T cells (nTreg) resulting from interaction of T cell receptors with their cognate antigen in the thymus, or defective regional mechanisms of oral tolerance relevant to environmental antigens. The latter could involve mucosal dendritic cells and the microenvironment of mesenteric lymph nodes, generating regulatory T cells which expand with the lamina propria. This specific regulatory mechanism is critical for the in maintenance of mucosal homeostasis and the limitation of inflammation. Any defective function in CD remains to be identified. In a sense both CD and autoimmunisation to tTG reflect a failure of mucosal immune regulation. Variable mucosal T reg function may therefore be a critical determinant of who with HLA DQ 2/8 gets CD (as most do not), who after prolonged gluten restriction can successfully revert to a normal diet (and whether or not this state of gluten tolerance is dependent on constant gluten exposure), and those who develop enhanced sensitivity following gluten deprivation. Some children at genetic risk can have variable levels of anti tTG while taking gluten, and some adults which coeliac disease in childhood appear to do well on a normal diet. With our current knowledge, such individuals should continue to be monitored with serology as intestinal damage can vary over time.

A contributing role for autoimmune mechanisms in CD may account for the common observations of protracted gut inflammation and mucosal damage despite strict adherence to a gluten – free diet, and the less common appearance of refractory sprue. The uncommon disease autoimmune enteropathy (AIE) maybe more closely linked with CD than recognised. Classically AIE presents with intractable diarrhoea unresponsive to dietary restriction, with villius atrophy but few intraepithelial T cells, and antibodies to epithelial and goblet cells. Recently mixed lesions (i.e. CD and AIE) have been described and study of patients with persistent mucosal inflammation despite gluten exclusion in CD may provide clarity regarding the significance of autoimmunity in CD.

5. **Clinical**

   **(a) Diagnosis**

The development of an automated serological assay which in controlled studies has a sensitivity and specificity approaching 100% (αtTG) resulting in CD being regarded as a common mild disease, caused a re-thinking of diagnosis. Clinical experience has modified the reliability of IgAαtTG with sensitivities and specificities closer to 90%, due to reduced antibody titre in those with minimal mucosal damage due to self-imposed diets or low level disease, and a high incidence of IgA deficiency (2%) giving false negative results. Small bowel biopsy with at least 4 endoscopic – biopsy specimens remains the gold standard for diagnosis (reflecting the focal nature of the disease). IgA αtTG remains the first up diagnostic test, now in a wider clinical settling. Less than 50% of those with CD diagnosed today have diarrhoea as the main presenting
In recent times, there has been a forty-fold increase in subjects diagnosed with CD. These subjects present with a range of gut symptoms including constipation and abdominal pain, or extra gut manifestations such as iron or folate deficiency, or osteoporosis. Occasional patients present with dermatitis, herpetiformis or gluten ataxia. For these reasons anti-tTG has become a screening assay in a wide range of clinical circumstances, much like antinuclear antibody in the diagnosis of systemic lupus erythematosus. The traditional ‘anti-gluten’ antibody assays have no clinical value, except for the IgA anti – deamnated gliaden peptide assay which has a sensitivity and specificity of greater than 90% and can be particularly useful in young children. The use of ‘point-of-care’ anti-tTG assays using finger–tip blood may have value in both diagnosis and in monitoring diet compliance, once reliable sensitivities and specificities have been established. In clinical practice, most diagnoses are clear-cut, but in about 10% there remain discrepancies between clinical, serologic and histologic data. In this patient group, determination of the absence of HLA – DQ2 (present in over 90% patients) and HLA – DQ8 (present in most the remainder) has valuable negative predictive value. Positive diagnostic value is limited by the fact that these alleles are present in one third of the population. Monitoring outcome of gluten challenge or the impact of a long term gluten-free diet (eliminating wheat, barley, rye and oats) can assist clinical decision-making. Diagnostic challenge can occur not only at presentation, but in the 10 – 30% who fail to respond (clinically and histologically) to a gluten – free diet and in those with complications. ‘Complicated’ CD usually presents in long established cases, with recurrent symptoms despite maintenance of a previously successful diet. Complications are uncommon and include adenocarcinoma of the small bowel, enteropathy – associated T-Lymphocyte lymphoma and refractory sprue.

Table 1

<table>
<thead>
<tr>
<th>Phase I</th>
<th>Phase II</th>
<th>Phase III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confirm Diagnosis</td>
<td>Confirm</td>
<td>Confirm Complication</td>
</tr>
<tr>
<td>• Review / repeat biopsy</td>
<td>• Diet adherence</td>
<td>• Adenocarcinoma of small bowel</td>
</tr>
<tr>
<td>• Phenotype: HLA – DQ2 / DQ8 alleles</td>
<td>• 2° diagnoses (e.g. pancreatic deficiency; bacteria overgrowth; IBS)</td>
<td>• Enteropathy – associated Tcell lymphoma</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Refractory disease</td>
</tr>
</tbody>
</table>

Table 1 summarises an approach to management of patients failing to respond to a gluten – free diet. About a quarter fail to have clinical and histological remissions. Persistent antiTG antibody in subjects on a gluten – free diet for 12 months commonly predicts poor adherence. Most with a primary failure can be effectively managed as per Phase I/II (above), with the majority due to a failure to adhere to diet. Poor compliance is particularly common in those diagnosed as an adolescent or adult, especially in those with minimal symptoms. About 5% have refractory disease despite stringent attention to diet, leading to the use of the term ‘refractory sprue’ to reflect uncertainty of diagnosis. Patients with refractory CD can present after a long period of stable control, with a recurrence of diarrhoea and weight loss, but also abdominal pain and mucosal haemorrhage due to ulcerative jejunitis.

A minority of those with refractory CD have a pre-malignant lesion with a clonal expansion of aberrant intraepithelial lymphocytes which can be detected in formalin – fixed biopsies with immune histochemical phenotyping. This group is prone to develop enteropathy associated T cell lymphoma, which is often multifocal and has a poor
prognosis. In uncomplicated CD, intraepithelial lymphocytes express the surface markers CD3, CD8, and T cell Receptors αβ and αδ, while in those with pre-malignant refractory CD there is a loss of these surface markers, with CD3 expressed only in the cytoplasm. The prognostic value of identifying clonality is that those with a normal phenotype usually respond to corticosteroids. If focal lesions are considered possible, a range of procedures including CT enteroclysis, video capsule endoscopy, positive-emission tomographic scanning, extended upper endoscopy and even laparoscopy can define lesions for biopsy and phenotypic analysis.

(b) Management

Carefully monitored gluten-free diet for life remains the cornerstone of therapy, with regular review of level of αtTG, and nutrient deficiencies and their consequences such as osteoporosis. This important review can be neglected especially in those without complaint. Many tolerate oats, but those not responding should also eliminate oats. The incidence of malignancy is double that of those without CD, including both T and B cell lymphomas, and carcinoma at any level of the gut. It is unclear whether malignancy is less in those adhering to a gluten–free diet. A range of therapies aimed at reducing dietary restrictions in all those with CD and in fine-tuning elimination of gluten with those with exquisite sensitivity, are attractive but not yet available or validated. The most likely supplement would be re-combinant proteolytic enzymes to digest ‘toxic’ gliadin peptides, though drugs blocking binding of deaminated gliadin peptides, or otherwise interfere with the immune response to gluten, are being considered.

6. Summary

Coeliac Disease (CD) was described 125 years ago as a rare disease of children. Effective therapy became available in the 1930’s, with pathogenesis considered to involve toxic gluten peptides. In the 1970’s a shift in thinking led to CD becoming recognised as a common mild disease akin to food allergy, caused by a hypersensitivity reaction to food protein. The breakthrough observations were first, a recognition of an association of gene alleles linked to ‘immune response genes’, and second, that gluten – reactive T cell mediated mucosal damage. The three lynch pins of pathogenesis (toxic gluten peptides, genetic susceptibility, and T cell – mediated hypersensitivity) converge to give a robust framework of pathogenesis, with glutamine – rich peptides binding to HLA – DQ 2 / 8 on antigen presenting cells where they activate specific T cells within the gut mucosa, damaging epithelium. An enzyme, tissue transglutaminase (ITG) binds to these peptides, and by doing so initiates an antibody response to ITG: this autoantibody though recognised for only 15 years has become the basis of both diagnosis and a primary argument supporting autoimmunity as a contributing cause of disease. These late surfacing ideas about a disease once considered ‘sorted’, have changed the clinical face of CD and raised important questions in relation to gluten – independent inflammation, refractory sprue and neoplastic complications and the role of mucosa – specific immune regulation in control of mucosal inflammation.
5. HLA – B27: Immunogenetics and the Cellular Immune Response

HLA-B27 is a class I surface antigen encoded by the B locus in the major histocompatibility complex (MHC) on the short arm of chromosome 6; its physiological role is to present antigenic determinants to cytotoxic (CD8) T cells. It is the most common class I antigen requested of a clinical immunology laboratory because of its strong association with the spondyloarthropathies. The association with ankylosing spondylitis (AS) was described in 1973, with over 90% AS positive compared with 8% of the population. There is also a significant association with the reactive arthropathies (formerly Reiter’s Syndrome). Though these disorders are often considered ‘autoimmune’, they are neither characterised by an autoantibody of diagnostic value (as with coeliac disease) nor does a pathogenic autoantibody dictate the pattern of disease (as with systemic lupus erythematosiis). The characteristics of AS are (i) a close association with a product of the gene cluster known as the MHC, HLA-B27, which is present on the surface of nucleated cells, and (ii) tissue damage associated with an infiltration of immunologically active cells. Here the evolution of the ideas of ‘immunogenetics’ and ‘cellular immunity’ are discussed to prove a framework for understanding the immunopathy of AS in particular, and more widely chronic inflammatory disease with genetic links.

1. Immunogenetics

Gregor Mendel in 1866 described his principles of segregation and independent assortment based on analysis of cross-pollinated pea plants, which replaced the ideas of inherited traits being ‘blended’, or variations of Lamarck’s idea of ‘inheritance of acquired characteristics’ prevalent at that time. In the early 1900’s, Tyzzer studied susceptibility to allogeneic tumour transplantation in mice and concluded the susceptibility was genetically determined.

However, the starting point of modern immunogenetics would come with the work of Gorer and Snell. In 1936 Peter Gorer identified a specific antigen as a ‘resistance’ factor to tumour growth when present in the donor, but not the recipient. These observations preceded those made by Peter Medawar in 1944, that rejection of allogeneic skin grafts in rabbits was caused by an immune response against the graft. Following World War II Gorer worked with Snell in the USA to combine their studies on tumour resistance genes (which Snell called histocompatibility or H genes), to conclude that three alleles at his H locus coded for tumour resistance. Gorer’s antiserum against his ‘antigen II’ was the first to detect an allele at the H locus. The locus became ‘histocompatibility locus 2’ or H-2. Because of the strength of the H-2 locus, it was called the ‘major’ histocompatibility locus in mice. Subsequently the H-2 locus became more and more complex with several subdivisions. In 1970 Thorsby showed the H-2 genes belonged to two segregate series encoded by two different loci – subsequently additional loci were found including those encoding the immune – response – associated (Ia) antigens. Thus the H-2 locus became the ‘major histocompatibility Complex (MHC) in the mouse.

In 1958 Dausset, Van Rood and Payne separately described antibodies in multiply transfused subjects or multiparous women that reacted with leucocytes from some but not all of those tested, detecting a polymorphic system of antigens on human leucocytes. Dausset described the first alloantigen (antigen MAC, to become HLA – A2), and noted that these leucocyte antigens ‘might become of great importance in tissue transplantation’.

In 1964 Amos organised the first of a long series of workshops to sort out the relationships of the different leucocyte (later HLA) antigens, and standardise testing. Most of the specificities identified were encoded by genes at one chromosomal region which became known as the HLA region (human leucocyte antigen). A standardised
nomenclature followed to avoid ‘local’ variations. Bodmer proposed two closely linked HLA loci in 1966, which became known as HLA-A and HLA-B, each with multiple alleles; in 1969 Dauset postulated a third locus (later called HLA – C).

Also in 1964 Bach described the mixed lymphocyte reaction, with the cell proliferative response influenced by the HLA region – that is, those with identical HLA regions did not react. The mixed lymphocyte response was not determined by HLA – A and – B antigens rather but a separate ‘MLC locus’ i.e. there were two different types of determinants: serologically defined (SD), and lymphocyte defined (LD), determinants – a concept that was also true for man. At the 6th international workshop in 1975, ‘MLC typing’ was used to define a new HLA – D region, itself containing several closely linked loci (DR, DQ and DP). In 1977 ‘blocking antisera’ were used to show that D locus antigens were on B cells but not T cells or platelets. By the 1980’s it had become clear that the HLA chromosomal region on the short arm of chromosome 6, encoded 6 different polymorphic series of determinants: A, B and C on the surface of most nucleated cells, and DR, DQ, and DP mainly on B cells, monocytes and dendritic cells. Klein in 1977 called A, B and C antigens ‘Class I’, and DQ, DP, and DR antigens ‘Class II’. By 2004 the HLA complex covered 7.6Mb, with more than 250 genes. One third of these genes had immune functions; some complement and cytokine genes also mapped with the HLA complex, indicating a genetic linkage of innate and adaptive immune responses.

Skin grafting studies by Dausset and van Rood in the early 1960’s showed that HLA antigens were strong histocompatibility antigens. The demonstration that graft survival times correlated with antibody against HLA antigens, supported this concept. Terasaki extended these findings to renal allograft survival in 1965 – the HLA complex was the MHC in man (as it was for H-2 in mice). HLA matching proved of value for selecting living related donor kidneys, but became controversial for cadaveric donors. In 1978, matching for HLA-DR antigens was shown to be important, and subsequent clinical studies considered good matches between HLA-A, HLA-B and DR antigens was important for living and cadaveric donors. As well, in patients with antibodies, a negative cross match between donor lymphocytes and recipient serum was required to avoid acute rejection. It is clear that optimal HLA matching is required for bone marrow transplantation, especially for class II antigens as here graft v host disease is a major risk. Successful transplants in immunodeficient patients and aplastic anaemia used HLA – identical sibling donors. International registries of HLA – typed bone marrow donors provides a pool of donors available for bone marrow transplantation. In 1987 Strominger used x-ray crystallography of HLA – A2 to show structural determinants that defined a groove to accommodate a ‘processed’ peptide – a similar groove was demonstrated in class II molecules 6 years later. Thus particular T cells recognise within this groove specific peptide - HLA complexes: CD8 T cells recognise peptide fragments in association with class I molecules (Townsend, 1986) and CD4 T cells recognised peptide fragments associated with class II molecules (respectively the effector and generating ‘ends’ of an immune response). This explained the MHC – linked Ir gene effects, due to different peptide – binding repertoires of MHC molecules. It also explained HLA – associated autoimmune disease, due to preferential binding of disease – associated HLA to particular immunogenic ‘self’ peptides. It is noted that the first HLA molecule discovered in man – HLA- A2 discovered by Dausset 29 years earlier – was the first to have its structure defined.

In 1967 Amiel reported that an HLA antigen (then known as 4c) was significantly more frequent in patients with Hodgkins disease, with a relative risk of about 3. This proved controversial, but initiated a search for HLA – associated disease. In 1973 Brewerton and Schlesstein independently described HLA – B27 in 88 – 96% of those with ankylosing spondylitis compared to 4 – 8% of controls. The initial concept was that either an immune response (Ir) gene in linkage disequilibrium to HLA – B27 was
involved, or there was a cross-reaction with an infective agent. Klebsiella was one popular candidate. The same year multiple sclerosis was associated with a HLA class II antigen by Gersild (now recognised as HLA-DR2). Subsequent search found many disease associations, especially autoimmune disease.

HLA class I and II antigens grew out of transplantation biology. They were strong histocompatibility antigens and quickly became determinants of transplantation success. However, their biological function was obscure. In 1963 Benaceraf showed that antibody to a polypeptide was controlled by a single gene (later called an Ir gene). He was to share the 1980 Nobel Prize with Snell and Dausset for their work on histocompatibility and immune response genes. In 1968, McDevitt showed Ir genes were linked to the H-2 complex. It was subsequently shown that the Ir genes mapped to loci in the H-2 complex corresponding to HLA-DQ and HLA-DR, i.e. to class II antigens. In 1972 Kindred showed cooperation between T and B cells required MHC identity and similarly it was shown for interaction of macrophage – associated antigen with T cells. The critical observation that MHC antigens were directly involved in T cell recognition of antigen was in 1974 by Zinkernagel and Doherty. They showed T cell recognition of antigen required identity of MHC antigens between the responder and the donor, a phenomenon termed MHC restriction i.e. in their model, the T cell receptor (TCR) recognised either MHC altered by the virus or a complex formed by the antigen and MHC antigens – a later comprehensive explanation had recognition by both cytotoxic (CD8) and helper (CD4) T cells using similar mechanisms (involving class I and class II antigens, respectively). A corollary was that this concept explained why MHC antigens were strong histocompatibility antigens, and predisposed to disease (see primer on coeliac disease). A similar situation for man was demonstrated in 1986. Doherty and Zinkernagel received the Nobel Prize in 1996. By the early 1980’s the structure of class I molecules had been defined with a glycoprotein ‘heavy’ chain non covalently bound to B-2 microglobulin, and class II molecules constructed of an alpha and beta heavy chain, with variation in the beta chain.

2. Cellular Immunity

In the 1880’s when the sway of Pasteur’s germ theory and vaccine studies suggested specific immunity was mediated by humoral factors, the idea that cells played a protective role was a surprise! Phagocytosis of bacteria by leucocytes was considered a mechanism to transport pathogens and Metchnikoff’s observations in 1882 that phagocytes killed anthrax bacillus was against opinion. Wright’s discovery in 1904 that antibody enhanced phagocytes (called opsinisation) reassured the humoralists, while in reality his discovery was a convergence point of innate and adaptive immunity. Robert Koch followed his discovery of mycobacterium tuberculosis with injection of a protein extract he called ‘old tuberlin’ in an attempt at inducing immunity. He observed a delayed (by 24 – 72 hours) indurated swelling with an influx of mononuclear cells, when ‘old tuberlin’ was injected into the skin of infected guinea pigs. Subsequently it was shown by Landsteiner and Chase in 1942 that passive transfer of this cutaneous sensitivity could only be achieved using specifically sensitised cells, and Feldman showed in 1963 the presence of transferred cells within the lesion. This delayed reaction contrasted with the more ‘immediate hypersensitivity’ skin reactions found in clinical allergy, where Prausnitz in his famous study injected serum from a subject sensitive to fish into his own skin. Prausnitz developed a red, swollen and itchy lesion at the site of injection when he ate fish. In 1903 Arthus described an ‘intermediate’ hypersensitivity (appearing in hours) which became linked with ‘serum sickness allergy’ first described in 1906 by von Pirquet and Schick in subjects treated with horse serum containing diphtheria antitoxin. Through the first half of the 20th century, it became clear that ‘delayed hypersensitivity’ reactions were not unique to tuberculin, but could be demonstrated with other microbes, and purified protein antigens particularly when linked to a hapten, or injected with an adjuvant. The term ‘allergy’ was introduced by von
Pirquet in 1906 to describe the ‘altered response’ that underpinned all three forms of hypersensitivity, but almost immediately ‘allergy’ was used in a clinical setting to describe any hypersensitivity reaction due to the ‘improper activation of the immune system’. To clarify what had become confusing, particularly to clinicians faced with an explosive appearance of information related to immunopathology, Gell and Coombs in 1963 proposed a simple classification of immune – based mechanisms of ‘hypersensitivity. In this classification delayed – type hypersensitivity was termed a Type 4 reaction. In 1956 after rigorous experiments, Billingham concluded that ‘homograft reaction against normal tissues like tuberculin sensitivity and other allergic responses of the delayed type, was mediated by cells not serum’. In 1953 Simonsen and Billingham independently described ‘graft versus host’ reactions: ‘in a previously formulated concept of transplantation immunity, immunisation occurred not only in the host, but vice versa as well’. Different animal models required: ‘the recipient to be young enough to be tolerant and the donor old enough to “form antibody”, with a genetic difference between donor and recipient. Splenomegaly in the recipient reflected injection of immunologically competent cells though it was not until 1975 that Sprent definitively showed that GvH reactions were due to T cells within the graft.

The third strand after hypersensitivity and transplantation that led to a recognition of cell – dependent immune activity through the first half of the 20th century, related to experimental organ specific autoimmune disease. The pathology of these disorders included destruction of parenchyma, infiltration by lymphocytes and replacement fibrosis as described by Hashimoto in 1912 in his description of (what would be) autoimmune thyroiditis. However, the traditional model was experimental autoimmune encephalomyelitis (EAE), dating from Pasteur’s development of a rabies vaccine in 1885 and the finding that paralysis could follow a rabies vaccination in 1888. An animal model was developed in 1925 when rabbits became paralysed following injections with human spinal cord; demyelination as the focal pathology was noted in 1933. In the 1940’s the classical model using Freunds complete adjuvant was described. In 1960 disease was transferred with lymphocytes, and the following year anti-lymphocyte serum was shown to inhibit the disease, firmly indicating a cellular immunopathology.

In the late 1940’s and early 1950’s Medewars group studying transplantation recognised that rejection of transplanted tissue was mediated by cells – their interest in mechanisms of promoting graft survival in man focussed attention on the concept of tolerance or ‘non-responsiveness’ which had been demonstrated by Owen in 1946 in non-identical twin cattle that shared placenta. This led Medawar and Burnett in the early 1950’s to focus on the importance of an immature ‘cell’ system around birth, such that interaction with antigen led to either elimination of reacting cells or a state of anergy.

By 1960, the stage was set for the ‘era of cellular immunology’. Burnett initiated the era with his concept of antibody acting as a ‘cell – bound’ receptor on lymphocytes enabling antigen to drive an immune response. In 1961 Miller showed the thymus to be essential for the development of cellular immunity – he followed this in studies between 1964 and 1968 (with Mitchell) by identifying T – and B – lymphocytes as separate cell populations, responsible respectively for cellular and humoral immune responses. They also demonstrated that T – and B – cells cooperated to facilitate antibody secretion (T cell ‘help’). Miller also showed that the thymus had a censorship role by removing self-reacting T cells, an activity that became known as negative selection or central T cell tolerance. Prior to Miller, lymphocytes were thought to be a single cell pool, that Gowans (in 1959) described as having a recirculating cell traffic. As the complex web of cell – to – cell interactions evolved, the ‘communication’ system between cells needed to be understood. In 1965 - 66 two key ‘messenger molecules’ secreted by activated lymphocytes - BIF and macrophage inhibition factor (MIF) – were identified, and between 1980 – 1983 IL – I and IL – 2 (respectively ‘macrophage derived’ and lymphocyte
derived) were described. IL-1 came out of 40 years study of fever, as a mediator, while IL-2 was the term given to ‘T cell growth’ factor (TCGF) which had been described in 1976 by Gallo. Various ‘blastogenic factors’ secreted by lymphocytes had been described in the mid 1960’s; Dumonde coined the term ‘lymphokine’ in 1969 to describe these mitogenic factors. The use of ‘growth factors’ enabled studies that eventually led to the cloning of the T cell receptor by Reinherz, Marrack and Kapler in 1983 and the identification of Th1 and Th2 - polarised T cell types by Hosmann (1986) characterised by the dominant cytokine secretion, respectively, of IL-1 and IL-4 (IL-4 induces naïve T cells to become Th2 cells, which themselves secrete IL-4, discovered by Howard in 1982. IL-4 promotes antibody secretion including IgE). In 1970 the more abstract term IL-2 was introduced to replace the ‘growth factor’ terminologies for factors promoting growth and survival of immune effector and killer cells (and later, as pivotal inducer of T regulatory cells).

In 1969, Nishizuka reported that neonatal thymectomy led to autoimmune oophoritis – the following year Gershon described a T cell population that suppressed antibody secretion, which he termed suppressor T cells. The observations that T cell populations could both ‘help’ and ‘suppress’ antibody secretion, turned traditional concepts of immune control on their head (though ‘networks’ such as idiotype – anti – idiotype networks had been described by Jerne as a theoretical concept in 1947). The importance of ‘tolerance’ had long been recognised by transplantation scientists and Miller had described ‘central tolerance’ within the thymus. Was ‘tolerance’ simply an outcome of the balance between T cell ‘help’ and ‘suppression’?

In 1992 Bloom transferred ‘tolerance’ with T cells adding support to this idea. These observations consolidated a model of immunity based around the idea that ‘self / not self’ was an outcome of T cell ‘control’, one that had dominated immunology for 30 years. The sequence of ideas relating to control of cell-mediated tissue damage began in the 1970’s with the concept of T suppressor cells, only to come to a halt when the MHC was cloned in 1982 and the expected ‘suppressor gene’ (I–J genes) could not be found. The Th1 and Th2 hypothesis of Mosmann and Coffman (1986) seemed a ready replacement with whole T cell subsets being capable of down regulating other subsets. This idea was remarkably resilient and held pride of place until newer methods such as use of Knock-Out mice showed IL-23, not INF-γ, was critical in the model autoimmune encephalitis (EAE). A new hypothesis for maintaining all – mediated tissue damage in autoimmune disease and chronic infection, had been born – the Th17 cell which was driven by IL-23. The fate of Th17 cells as mediators of chronic damage will be defined in time, but it is unlikely that a single cytokine will be responsible for chronic inflammation.

T cells, rather than being the ‘directors’ of immune outcome, were beginning to be seen to be instructed by cells of the innate system. This ‘new’ segment of immunology actually began in 1868 when Langerhans identified dendritic cells. These cells would be characterised by Steinman as a component of the lymphoid system in 1973, and then in 1998 as antigen presenting cells with direct influence on the pattern of T cell response. Prior to this discovery by Steinman two options were being discussed. The first, by Janeway was that the innate immune system was the ‘gate – keeper’, determining if T cells responded or failed to respond, to antigen. He saw the key ‘activator’ as conserved ‘ancient’ patterns common to pathogenic microbes. The second, by Matzinger in 1994, took control back a step, postulating that cell damage created chemical ‘danger signals’ that were detected by antigen presenting cells (APC’s) to influence the balance of the T cell response. She later suggested Toll-like receptors on APC’s recognised similar conserved signals from cell damage or pathogens, to unify the emerging ideas regarding control of the T cell response. These ideas themselves became challenged when the earlier concept of ‘suppressor T cell’ became more refined as T reg cells, particularly when they were characterised by markers. In particular, a defining T cell transcription
factor (Fox P3) was identified in 2001, and ‘down regulating’ cytokines IL-10 and TGFB became linked to the T reg cells. Using markers natural T reg cells were identified in the thymus, and induced T reg cells in peripheral lymphoid tissue were recognised. The cooperation of these cells in central and peripheral compartments prevented autoimmune disease. Currently dendritic cells are through to critically determine whether naïve CD$ T cells differentiate into Th or T reg cells, bringing components of the various preceding models, together.

**HLA – B27 AND ARTHROPATHY**

How does our current understanding of cellular immunology and immunogenetics help understand the pathogenesis of HLA-B27 positive spondylo–arthropathies?

Initially it was thought that disease was associated with HLA – B27 carriage of self-protein (causing autoimmunity) or microbial antigens (causing cross reactivity or molecular mimicry and a hypersensitivity reaction involving T cells). No evidence for these theories has been found. However HLA – B27 was found to undergo ‘misfolding’ in the endoplasmic reticulum (ER) and form dimers on the cell surface. Misfolded proteins can delete (e.g. haemophilia), accumulate in the extracellular space (e.g. Alzheimer’s), or accumulate within the ER and become unable to bind antigen, causing ER ‘stress’. In turn this increases the various processes within the cell needed to cope with accumulated protein – a process known as the ‘unfolded protein response’ (UPR). In immune cells, the UPR enhances the cytokine response to microbial infection causing an inflammatory response which includes an increase in bone formation (a feature of ankylosing spondylitis). Dimerisation of HLA – B27 on cell surfaces may activate NK cells and further contribute to chronic inflammation. To further enhance an inflammatory response additional genetic predisposition likely contributes gene products that augment disease. Thus the mechanism whereby the presence of HLA – B27 on the surface of nucleated cells is unlikely, as earlier thought, to cause chronic inflammation in the axial joints and entheses through carriage of antigen (with the possible exception of reactive arthropathy), but rather enhance an inflammatory response initiated by abnormal intracellular folding of HLA – B27 protein. That only 5% of those with HLA – B27 develop arthropathy reflects the current knowledge gap in understanding what translates the presence of HLA – B 27 into disease (see Fig 1):
Thus the disease process does not appear to involve either autoimmunity or hypersensitivity but rather a new mechanism. The major questions in search of answers are: (1) Is there an antigen involved? (2) What is that antigen? (self / non self); (3) what are additional genetic effects? (i.e. to separate the 5% with HLA – B27 with disease from 95% without; (4) what component of cellular immune response is activated? and (5) what localises T cell driven inflammation?

**Conclusion**

Though still limited, our current understanding of HLA-B27 associated spondyloarthropathies is a product of an evaluation in knowledge of both cellular immunity and immunogenetics, which proved to be closely intertwined. That cells could directly involve in adaptive immunity was recognised as a result of three separate lines of research: delayed – type hypersensitivity, organ specific autoimmune disease and graft rejection. It was this latter area of transplantation immunology that focussed on the genetic complex that coded for strong ‘transplantation antigens’ with their immediate relevance to clinical practise. Subsequently it was recognised that genetic loci relevant to immune recognition coded for molecules directly involved in both the generation and expression of the immune response. These genetic loci were part of the gene complex coding for the major histocompatibility antigens (the MHC)The details of these critical interactions could only understood within the context of an advanced understanding of ‘cellular immunology’ and the molecules responsible for communicating within a complex cellular framework that involved cells from both innate and adaptive limbs of immunity. Thus HLA – B27 related spondyloarthropathies can now be analysed within an integrated cellular system. Surprisingly, neither of the two classical antigen-driven causes of chronic inflammation (autoimmune disease – ‘self’ antigen; hypersensitivity disease – non self antigen) easily explain the disease. Rather, unique chemical changes in the HLA-B27 molecule within cells promoting inflammation, attracts current attention. Ideas in terms of regulation of effector T cells in chronic inflammation are consistent with current data in ankylosing spondylitis where Th17 cell activity closely correlates with clinical parameters of active disease – how this T cell is directed remains to be seen.
5. INNATE IMMUNITY: THE ALPHA AND OMEGA OF CLINICAL IMMUNOLOGY

(a) Introduction

Innate immunity is the first line of defence! It reacts quickly to injury with molecules and cells that are immediately available and which are encoded in germline DNA. With respect to the effector components of the innate immune system, the immunology laboratory conducts tests of complement to assess activation of the ‘classical’ pathway (reduced levels of ‘C3’ and C4” components), the ‘alternate’ pathway (reduced ‘C3’ but normal ‘C4’), screen for uncommon deficiencies (reduced haemolytic complement or CH50) or detect deficiency of the major fluid –phase inhibitor (C1 esterase inhibitor) or cell-bound inhibitors of C3b (CD55, CD59). Historically a number of functional assays measured neutrophil migratory, phagocytic and killing capacities, all replaced by flow cytometric assays testing the respiratory burst, phagocytosis and bactericidal killing. These tests of neutrophil function are primarily designed to detect chronic granulomatous disease. Collectively, these assays reflect the traditional view that innate immunity is first and foremost a front line mechanism of protection that is controlled and directed by the more evolved processes of the adaptive immune system. The central advantage of this integrated system is that relatively few molecules and cells of many specificities can immediately marshal non-specific but quantitative effector pathways to neutralise an expanding microbe population in a timely manner. This century old view was shaken by recent discoveries that led to the award of the 2011 Nobel Prize. The recipients were Bruce Beutler and Jules Hoffman (for describing receptors that recognise conserved patterns on bacteria that trigger a protective inflammatory response) and Ralph Steinman (for identifying dendritic cells and demonstrating their critical influence on T cell – dependant immunity). These new ideas have not only turned around our views on the control of immune function, but also impacted on clinical immunology with a new disease concept of ‘autoinflammatory disease’ and novel approaches to tumour immunotherapy using ‘antigen – pulsed’ dendritic cells for ‘designer vaccines’.

In this section we discuss an incredibly complex system of protection that evolved in metazoan organisms about 750 million years ago before the need arose about 500 million years ago in vertebrates for a more sophisticated ‘adaptive’ immune system that modified existing innate systems by adding clonally expanded specific antigen receptors. By integrating with existing innate immunity this integrated system gave a powerful process that more broadly recognised danger signals, had memory linked to an anamnestic response, and activated effector options that could be tailored to the challenge of the moment. Innate immune mechanisms are remarkably conserved – for example homologous peptides toxic for microbes (defensins) are found in plants and animals, and cellular mechanisms such as phagocytosis and the ‘killing’ of cells lacking self recognition markers, have wide phylogenetic origins. Particular attention is paid in this document to selected components of this complex system, as they relate to chronic inflammatory conditions (hypersensitivity and autoinflammatory disease). These disorders complement autoimmune disease which reflects breakdown of control in adaptive immunity.
The pivotal role of complement as a host protection mechanism has been recognised for more than 100 years – the historic evolution is important, but given the great changes in our understanding of the relationship between innate and adaptive immunity in recent years, it is useful to first look at a summary statement of a contemporary understanding of complement pathways:

The central player is C3b. C3 has an unstable internal thioester bond, which spontaneously hydrolyses to C3b which binds through a receptor to cell surfaces. Live human cells have inhibitors of C3b anchored to their surfaces ensuring no progression of activity. However, bacteria and apoptotic cells have no inhibitor (known as DAF or Decay – accelerating Factor) – this “complex” activates the “alternate pathway” of complement activation – essentially an amplification loop for the generation of C3b with its own inhibitor (control) system now called ‘properdin’. The most important function of complement is the driving of a neutrophil – based inflammatory response – this mediated by C3b due to its chemotactic
property. Thus complement is ‘poised’ to generate protective inflammation due to the continuous production of small amounts of C3b, which by attaching to bacterial or apoptotic cells lacking inhibitors, activate an amplification loop to generate more C3b and initiate neutrophil – mediated phagocytosis. Classically the complement system is a sequence of enzyme – proenzyme reactions operating as a cascade, initiated by C1q binding to antibody – antigen complexes with the terminal components (C5b, C6, C7, C8, C9) forming a Membrane Attack Complex (MAC) that creates a pore in the target cell membrane followed by osmotic lysis. Review of Figure 1 shows that complement can be activated by adaptive immunity (through C1q) and by direct activation through recognition of pathogens – first by spontaneously generated C3b attaching to bacteria and second by C1q recognising terminal mannose units in bacterial cell wall structures (the ‘mannose lectin pathway’). “Split” products of C3 (C3a & C3b) and C5 (C5a and C5b) mediate the main effector pathways of inflammation, while C3b can also continue the cascade to terminal cell lysis. Auto-control through inhibitors preventing excessive activity is common in innate immune systems. Here there are surface (e.g. DAF or CD55) and fluid – phase (e.g. C1 esterase inhibitor) inhibitors, each with clinical sequelae when defective. C1 esterase inhibitor prevents formation of the C1 complex, as well as inhibiting Kallikrien which releases bradykinin, which turn causes vasodilation and oedema. This condition is called Hereditary Angioedema. Deficiency of cell surface inhibitors of cell – bound C3b such as DAF, leads to lysis of red cells in Paroxysmal Nocturnal Haemoglobinuria.

The history of complement is, on one hand, an endorsement of protein chemistry through the first 60 – 70 years of the 20th century. On the other, it records a sequence of observations that had cell lysis at their centre; shifting focus to the control of inflammation following complement activation occurred late in the event because the experimental readouts of complement activity usually involved cell lysis. From the beginning complement was regarded as a major fluid phase amplification mechanism initiated by antigen – antibody complexes and a model for usage of ubiquitous innate immune mechanisms by the specific players of adaptive immunity. In the 1880’s interest in host protection began as the germ theory took hold primarily due to the work of Pasteur in Paris. In that decade both the cellular and humoral arms of innate immunity were discovered. In 1883 Mechnikof described phagocytosis in starfish larvae and postulated that leukocytes could similarly engulf microbes.

In 1889 Hans Buckner in Munich described a serum factor capable of killing bacteria he called ‘alexin’, which Jules Bordet would later show to have two components – one heat resistant against specific organisms, (later recognised as antibody), the other heat labile that was active against many bacteria indicating an absence of specificity. In the 1890’s, Paul Ehrlich included this non-specific antibacterial factor in his side chain theory of immunology. He postulated cell-associated receptors called amboreceptors that bound both specific antigen and non-specific ‘alexin’ (which he would re-name ‘complement’). Amboreceptors released into the circulation became ‘antibodies’. In 1901 Bordet demonstrated the critical interaction between antibody and complement via his development of complement fixation assays for antibody. In 1906, this assay procedure was used to detect ‘reagin’ as the basis of the first diagnostic of value for syphilis. In 1904 Wright identified another convergence point of innate and adaptive immunity. On this occasion for the cellular arm of innate immunity when it was noted that antibody could ‘opsonise’ antigen to enhance phagocytosis. A main theme of complement became detection of the sequence of interactions from binding with antibody – antigen complexes to cell lysis. Bordet recognised a ‘single’ complement but Ferrata in 1907 postulated a sequence of reactions involving at least two components (C1 & C2), while Coca (1914) described a heat – stable fraction (C3) that could restore lytic activity to yeast – inactivated complement. In 1926 Gordon identified ‘C4’ by inactivation with ammonia. ‘C3’ was found to include 6 components (C3a –f) which only in the 1920’s were

1 Appendix 1
recognised as proteins. It was not until the 1960’s that chemists such as Muller-Eberhard and Nelson purified components and ‘built up’ the reactive sequence. The confusing nomenclature based more on the order of discovery was rationalised by the WHO in 1968, with a terminology based on order of activation. In 1961 Mayer postulated ‘one hit’ lysis. In the 1930’s C3b was shown as a divergence point by Stanley who showed that as well as being ‘online’ for lysis, it could enhance phagocytosis. C3 (though only vaguely understood) was taking shape as a critical component following Van Dungern in 1911 who showed non-antibody activation of complement and Coca’s demonstration that yeast – depleted complement could be reversed with heat-treated serum (i.e. C3). It was however nearly 50 years (1954) before Pillemer postulated his ‘properdin’ pathway based on direct activation by bacteria, that the idea of an ‘alternate complement pathway’ was born. This was not a popular concept, and only in recent times has the central role of spontaneous generation of C3b been recognised as the critical event, with the ‘alternate’ pathway amplifying this process. A third pathway (manose-binding lectin path) was described in 1989 following investigations of a child prone to recurrent infections. This proved to be a non-antibody activation of C1q, the start point for the ‘classical’ pathway. Thus at least two direct ‘start points’ for complement activation by bacteria are now recognised – C3b and C1q. Complement is a case where the history provides insight into how its function was understood at different times, but a current view helps in the interpretation of events.

(c) Hypersensitivity Disease

Hypersensitivity disease is common and includes a range of disorders that include allergy and vasculitis amongst many acute and chronic inflammatory diseases. In general hypersensitivity disease refers to a damaging host response to otherwise innocuous extrinsic antigens. The same mechanisms can significantly contribute to damage in diseases recognised in their own right such as infections and autoimmune disease. The mechanism of hypersensitivity disease involves a disconnect between specific adaptive immunity directed against an antigen, and the non-specific innate mechanisms recruited to handle that antigen. This ‘disconnect’ leads to an inappropriate and excessive inflammatory response causing bystander damage to normal tissues. The three classical hypersensitivity disease groups were recognised by the temporal and histological characteristics of an inflammatory response that followed injection of antigen into the skin. More recently, a fourth group has been recognised, with the inflammatory response particularly at mucosal surfaces (Table 1).
Table 1. Classification of Hypersensitivity

<table>
<thead>
<tr>
<th>Type of Reaction</th>
<th>Time</th>
<th>Host Response</th>
<th>Pathology</th>
<th>Coombs Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immediate</td>
<td>Minutes</td>
<td>IgE</td>
<td>Mast cell</td>
<td>Vasodilation / oedema</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Type 1</td>
</tr>
<tr>
<td>Intermediate</td>
<td>Hours</td>
<td>IgG</td>
<td>Complement</td>
<td>Arthus reaction</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Type 3</td>
</tr>
<tr>
<td>Delayed</td>
<td>Days</td>
<td>Th1 cells</td>
<td>Cytokines</td>
<td>Granuloma</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Type 4</td>
</tr>
<tr>
<td></td>
<td>Hours</td>
<td>Th17 cells</td>
<td>Cytokines</td>
<td>Neutrophil -</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Infiltrate</td>
</tr>
</tbody>
</table>

(i) Immediate Hypersensitivity

Allergic disease caused by a host response to inhalation, ingestion or skin contact with minute amounts of antigen (here called ‘allergens’), has been recognised clinically from about 3,500BC when King Menses (Egypt) died after a wasp sting. The link with inhaled allergens was documented by Charles Blakely in 1869 who had hayfever, and who developed an immediate local inflammatory response when he applied a grass pollen to excoriated skin with dampened lint. In 1902 Richet and Portier coined the term anaphylaxis to describe unexpected life threatening reactions noted during the course of ‘routine’ repetitive immunisation studies. In 1906 von Pirquet first used the term ‘allergy’ to describe an ‘altered state’ they observed when unexplained symptoms occurred in patients with diphtheria treated with horse serum antitoxin. In 1911 a practical outcome of careful observation led Noon and Freeman to develop immunotherapy for allergic disease by injecting an increasing amount of specific allergen in patients with hayfever. In the same year, the first use of sublingual desensitisation was claimed, though the first clinical trial documenting the value of mucosal desensitisation was not reported until 1986, by Glynis Scadding in patients with house dust allergic rhinitis. The main mediators of inflammation were identified as histamine by Dale in 1920, and ‘slow reactive substance of anaphylaxis’ by Kellaway and Trehewin in 1940 (identified as Leukotriene by Samuelsson in 1979) – specific antagonists of mediators began with the development of anti-histamines by Bovet in 1937, and anti-leukotrienes in the 1990’s. The key cell involved in immediate hypersensitivity as the source of histamine was identified by Riley and West in 1953 following study of a canine mast cell tumour. In the mid 1970’s Samuelsson showed most leukotriene came from neutrophils. The ‘specific’ component of the sequence underpinning immediate hypersensitivity was recognised in 1919 when transient asthma due to horse dander allergy was observed following a blood transfusion. In 1921, Prausnitz and Kustner showed this blood factor to be ‘skin sensitising’ following the passive transfer of a positive skin test to fish antigen. Called reagin, it was considered to be a ‘labile complex’ until the work of Ishizaka in 1966 chemically defined reagin as a novel immunoglobin that bound to mast cells, with antisera to “IgE” capable of blocking reaginic activity. The same year Johansson reported a myeloma protein with the characteristics of this novel immunoglobin.

Modern allergy practise is based on these historic principles. The use of the initial terms have changed (‘allergy’ now refers to those disorders classically caused by immediate hypersensitivity; anaphylaxis refers to severe - often life threatening-systemic reactions to
antigen exposure, mediated by mechanisms of immediate hypersensitivity. Atopy refers to
the genetically determined proneness to respond to environmental allergens with a specific
IgE response). Based on these principles, diagnosis using a limited batch skin prick tests or
in vitro assays for IgE antibody, determines first, that the patient is atopic and thus could
have an allergic disease, and second, identify common potential relevant allergens. Choose
a limited selection of allergens (e.g. cat, rye grass, D. Pteronyssinus, mould mix) plus any
allergen suggested from clinical history. Some would also request a total IgE as some atopic
individuals will be missed by a limited screen, and because anti- IgE therapy may be
anticipated. Once the patient has a clinical diagnosis that could be ‘allergic’ (asthma,
hayfever, eczema and food reactions) and is shown to be atopic (by skin tests or specific IgE
assays), then the clinical imperative is to ensure a match between exposure to allergens and
the pattern of IgE antibodies. For example a positive test for rye grass pollen in a patient with
perennial disease is a non-match as rye grass exposure is seasonal, but this clinical pattern
is consistent with a positive test for D.Pteronyssinus. Having established the sequence:
clinical syndrome \(
\rightarrow \)
atopic status and IgE antibody ‘match’ a clinical therapy plan can be
developed: avoid the allergen, suppress symptoms or desensitise by injection or mucosal
routes. Recently the use of monoclonal anti-IgE antibody therapy has been added acting to
reduce the amount of specific IgE and the high affinity receptors for IgE. It has been found to
be effective therapy for severe allergic asthma and chronic urticaria and is being assessed
for other significant ‘allergic’ diseases.

(ii) Intermediate Hypersensitivity:

The ‘intermediate reaction’ to antigen injected into skin was described by Arthus in 1903. He
noted oedema followed by gangrene after repeated subcutaneous injections of horse serum
into rabbits. These reactions were classified by Gell and Coombs in 1963 as “Type 3
hypersensitivity”. They are due to small – immediate sized complexes made in antigen
excess that persist in the circulation. They do not “fix” complement while circulating and are
not removed by macrophages, but become inserted into small blood vessels and at sites of
blood filtration involving significant pressure gradients such as gomeruli and synovial
membranes. Attached to these sites the immune complexes now ‘fix’ complement with
release of anaphalotoxins (C3a and C5a) and chemataxins (C3b), the accumulation of
neutrophils and release of lysosomal enzymes causing tissue damage. Damage to vessel
walls causes thrombosis with ischaemic ‘fibrinoid necrosis’, all features of vasculitic disease.
Serum sickness became the prototypic vasculitic disease. First described by Von Pirquet
and Shick in 1905 following repeat injection of horse anti-diphtheria antitoxin, with patients
developing fever, urticaria, arthralgia and regional lymphadenopathy 1-2 weeks after
immunotherapy. Subsequent studies demonstrated persistent complexes in antigen excess,
complement fixation, with deposition of complexes in small vessels- as antigen is cleared.
Larger complexes appear in antibody excess which are cleared by macrophages to correlate
with clinical improvement: Pre-formed immune complexes where shown capable of causing
damage by Germuth in 1957. Dixon studied a range of animal models of glomerulonephritis
with a focus on immune complex deposition as the dominant cause of disease. In 1961 he
concluded that the amount of antibody was the critical determinant of both acute and chronic
glomerulonephritis.
(iv) Vasculitis

It would be convenient to view ‘vasculitis’ as an entity linked by a common mechanism, but that is not the case. Most if not all have inflammatory lesions and involve ‘immune mechanisms’ but the ‘drivers’ of inflammation and specific pathways remain uncertain for many. Given the uncertainties of cause and classification and the importance of having a therapy-based framework for assessment and management of a group of often life-threatening diseases, an historic perspective becomes important. Perhaps the best approach is to trace the evolution of vasculitic disease from the classical description of periarteritis nodosa by Kussmaul and Maier in 1866, through to the necrotising vasculitis we now recognise as ‘classic polyarteritis nodosa’ with distinguishing features that include:

- general restriction to medium vessels almost exclusively arterial
- tendency to form micro aneurisms
- an absence of pulmonary involvement
- a focal disease
- an absence of granulomatous disease and antineutrophil cytoplasmic antibodies (ANCA)

This evolution over nearly 150 years has been referred to as the ‘polyarteritis nodosa trunk’ from which five named’ vasculitic diseases with their own characteristics have been split, and four important ‘diverging’ points have emerged identifying a new ‘class’ of vasculitis with particular characteristics, increasingly linked to pathogenesis. The ‘named’ disorders based on clinical characteristics include:

- Henoch – Schönlein Purpura (1837 & 1894);
- Thromboangitis Obliterans (Buerger, 1908);
- Behcet’s Disease (1924);
- Takayasu’s Vasculitis (1952); and
- Kawasaki Disease (1967).

In most of these diseases, recent studies identify activated immune mechanisms in the context of infection, genetic susceptibility, or super antigen induced change. The ‘switch points’ from the vasculitis stem have greater clinical value and have led to a better understanding of mechanisms with more valuable classification:

- **Temporal arteritis** was recognised by Horton in 1934 as a granulomatous inflammatory disease of large vessels, which became extended to a concept of giant cell arteritis including polymyalgia rheumatica. Possibly the most common arteritis, it is usually self-limited. The cause of chronic inflammation and its localisation to large arteries in the carotid-subclavian distribution includes a genetic factor involving a common sequence motif of HLA – DR which maps to the antigen binding site, suggesting a critical role in antigen selection and presentation, though the nature of the antigen is not known. Vessel damage due to disease or sun exposure may influence localisation.

- **Wegener** recognised the classic triad of granulomatous inflammation of the respiratory tract, granulomatous necrotising vasculitis, and glomerulonephritis in 1936. Its association with anti-neutrophil cytoplasmic antibody (ANCA) in 1985, an association noted subsequently with other vasculitides (microscopic polyangiitis, Churg- Strauss syndrome and renal-limited vasculitis) has shaped a classification group based on involvement of an autoantibody of uncertain function and origin. ANCA may arise by either molecular mimicry or defective apoptosis. The former concept is based on shared regions between microbial superantigens (which can directly activate T cells) and host protein (e.g. M proteins of Streptococcus).
• Pathogens and cardiac myosin and laminin in rheumatic fever. About two thirds of those with Wegener’s granulomatosis are nasal s.aureus carriers; carriers have an eight fold increase in incidence of relapse. Apoptosis of neutrophils is a normal control to limit damage – if defective, cell components become exposed to the immune system. In models, neutrophils are activated by ANCA which then adhere to endothelial cells with degranulation causing necrotising vasculitis.

• **Hypersensitivity angitis** was separated in 1942 from polyarteritis nodosa by Rich, though microscopic and macroscopic forms of polyarteritis nodosa had been identified as early as 1923. There followed variations on this theme and varying terminology, with recognition that this common predominantly cutaneous vasculitis presenting as ‘palpable purpura’ is classically caused by immune complex deposition and a complement – mediated ‘arthus reaction’.

• **Churg Strauss Syndrome (1951)** was initially thought to be on ‘expression’ of PAN in subjects with asthma or bronchitis, eosinophilia, and granulomatous inflammation and that especially targeted the lungs. Subsequent discovery of ANCA in many led to a grouping with Wegener’s granulomatosis (above) as ‘ANCA positive vasculitis’.

This evolution of ideas is reflected in a sequence of classifications, beginning with that of Zeek in 1952 based on morphology, through various subsequent groupings aimed to give clinical guidance. In particular efforts were made to identify necrotising systemic vasculitis, limited cutaneous vasculitis and overlap syndromes. A particular focus on classification based on vessel size, pathology and the presence of ANCA came from a Consensus Conference in Chapel Hill (1993) which influenced the modern approach to management. A recent ‘classification’ by Stone (2005) has re-stated the directions of the Chapel Hill Consensus, recognising vessel size and ‘primary and secondary’ vasculitis, but attempting to group vasculitis categories in terms of pathogenesis:

• Vasculitis predominantly affecting large vessels (Takayasu arteritis, giant cell arteritis, Behcet’s disease);
• Vasculitis predominantly affecting medium vessel (Polyarteritis nodosa, Buerger disease, Kawasaki disease, cerebral angiitis);
• Vasculitis predominantly affecting small vessels caused by immune complexes (Goodpasteur’s disease; cutaneous leukocytoclastic angiitis; Henock-Schonlein syndrome; urticarial vasculitis; mixed cryoglobulinaemia);
• Vasculitis associated with ANCA. (Wegener’s granulomatous; microscopic polyangitiis Churg-Strauss syndrome, renal –limited vasculitis).

All current classifications are flawed, but the unpicking of triggers and pathways has enabled a tentative approach to a classification based on pathogenesis. The historical evaluation of understanding arteritis is particularly helpful, especially when built around the central player – Polyarteritis Nodosa. Immune complexes contribute to pathology in this life-threatening disease – about 25% are Hepatitis B positive and may benefit from intravenous gamma globulin and / or plasma exchange therapy. It is likely monoclonal antibodies directed against key molecules involved in vasculitis will prove valuable as mechanisms become better understood.

(v) **T Cell hypersensitivity**

Cell –based cytokine-mediated hypersensitivity has been discussed in relation to tuberculin and the Chase Phenomenon, when cells were shown capable of transferring the reaction. Classical delayed – type hypersensitivity (Type 4 in the Gell & Coombs classification) is responsible for contact sensitivity, drug reactions and tissue damage in many infections, classically tuberculosis. The key cell type is the Th1 CD4 cell with secretion of TNF and INF 8
and IL-12. This reaction is also responsible for certain granulomatous diseases, notably sarcidosis. In these circumstances a systemic anergy to T cell antigens occurs due to T reg-suppression of IL-2 and a sequestration of reactive T cells to sites of disease. This may predispose to infection neoplasia, and autoimmune disease. A more recently recognised form of T cell hypersensitivity has been described that involved Th17 cells which induce an inappropriate recruitment of activated neutrophils to induce purulent inflammation within airways colonised by bacteria. This same mechanism may occur in forms of vasculitis, if antigen is located in association with the vessel wall, as well as certain autoimmune diseases.

(d) Toll-like Receptors & Dendritic Cells

Phylogenically primitive organisms have circulating cells capable of both phagocytosis and ‘killing’ of foreign cells detected as lacking ‘non-self’ patterns. In mammalian systems different cell systems have evolved with these functions. Mechnikov described Phagocytosis in 1883 while; natural killer’ or NK cells were an accidental discovery in the early 1970’s when ‘background noise’ reflecting non-specific killing of target cells was noted in cultures assaying specific T cell cytotoxicity. As with complement, understanding of innate cellular immunity came to similar conclusions i.e. that innate and adaptive immune differences blur with recognition that foreign ‘patterns’ can be recognised, inhibition systems exist to prevent self-harm (here by recognition of class 1 antigens) and a level of ‘memory’ can be measured. Indeed cells of innate immunity have not only a pre-emptive role with respect to recognition of pathogens, but have a profound influence on adaptive immune function as a ‘conductor of an immune symphony’. In 1974, Ralph Steinman isolated dendritic cells which he characterised. In 1980 he showed that these cells were the critical “accessory” cell required for the induction of specific immunity acting as the “professional antigen presenting” cell. In 2001, these same cells were shown to be the gate keeper for tolerance.

The dendritic cell was also to have a parallel pathway activated by recognition of conserved surface patterns carried by bacteria, leading to secretion of pro-inflammatory cytokines that on one hand mediated protection, and on the other could profoundly modulate adaptive immune responses. Recognition of primitive bacterial patterns began with studies by Beeson in 1948 who used endotoxin (or bacterial lipopolysaccharide) to drive leucocytes to promote fever. Endogenous pyrogen (later recognised as IL-1) was described in 1940, and the genes for IL-1β and IL - 1R1 (its receptors) cloned, respectively, in 1984 and 1988. In 1989, the Toll gene was identified in the fruit fly as a determinant of the ventro-dorsal axis in embryogenesis. It was also noted in 1995 that ‘Toll’ related to protection against fungal infection. In 1997, human ‘Toll-like’ receptors (TLR’s) were described. Between 1998 and 2001 many TLR’s were described (now at least 10) and all were identified as receptors for microbial products, giving rise to the idea of Pathogen-Associated Molecular Patterns (PAMPs). Thus, the original observations with lipopolysaccharide as a ‘surface pattern’ on gram negative bacteria became a prototype (binding to TLR4), with other TLR’s binding lipotechoic acid (from gram positive bacteria) and terminal mannose residues in other microbes Several TLR’s were not on the cell surface, but rather on intracellular endosomes – these had an unique ability to bind viral nucleic acids. Signals from TLR’s were directed by ‘adaptor proteins’ along one of several pathways, giving specific response to the binding microbe. The major pathway triggered by ligation of surface TLR’s, used the adaptor protein MyD88 (discovered in 1988) induced the NF-KB transcription factor. This was discovered by Sen in 1986 as a major control point for pro-inflammatory cytokines (such as IL-1, TNF, IL-6) as well as chemokines, adhesion molecule expression, and co-stimulating molecules. These factors have a profound influence on the nature of any T cell response to antigen presented by dendritic cells in conjunction with class I molecules, to the antigen receptors on the T cell. In 2002, a second adaptor protein complex (TRIF) was described that responded to endosomal TLR’s to activate a second transcription pathway (IRF) that induced secretion of
type 1 interferons. This is appropriate given that endosomal TLR’s largely bind to viral nucleic acids.

In 2000 a second system of pattern recognising receptors was discovered. On this occasion they are within the cytosol. The most prominent of the cytosol systems became known as NOD-like receptors. By now the ‘patterns’ recognised had become known as ‘PAMPS’ and ‘DAMPs’ (Damage-associated molecular patterns – endogenous molecules released from dead or dying cells). Activation of cytosol receptors initiates a sequence known as the ‘inflammasome’ from which an enzyme (caspase) is activated to convert IL-IB precursor to the active pro-inflammatory cytokine IL-IB (the precursor is itself an outcome of the transcription factor NFkB). In summary membrane-bound and cytosol receptor systems operate in a coordinated way to recognise and respond to PAMPS and DAMPS, using selective adaptor – transcription pathways to release pro-inflammatory and anti-viral cytokine responses as well as regulatory molecules that shape the concomitant adaptive immune response through T cells. Defects in the components of these pathways, through deficient inhibitors or impaired feedback control, can lead to chronic inflammatory disease and/or impaired protection and recurrent infection. Interest in the relationship between the gut mucosa and the intestinal microbiome is emerging as a complicated but critical interaction characterised as ‘controlled inflammation’ underpinning host protective mechanisms. Genetic polymorphisms of NODs predispose to Crohn’s Disease, suggesting defective handling of gut microflora contributes to chronic intestinal inflammation. A significant difference between a commensal and a pathogen is whether the organism reacts solely with surface TLR’s, or enters the cytosol to directly activate the inflammasome. Many pathogens have mechanisms such as the capacity to form ‘pores’ in cell membranes that enable them or their toxins to enter the cell.
(e)  Autoinflammatory Disease

In the late 1990’s the term autoinflammatory disease was introduced: conditions with recurrent episodes of uncontrolled inflammation with fever, rash and synovitis, often with a risk of secondary amyloidosis. These conditions were separate from autoimmunity, lacking autoreactive antibodies and T cells. Autoimmunity was dysregulation of adaptive immunity, while this newly recognised disease cluster was an outcome of dysregulation of innate immunity. Initially attention focused on periodic fevers – a review of 2000 cases in Baltimore found mutations in one third, usually a cause of dysregulation of innate immune components. The prototype was Familial Mediterranean Fever with genetic polymorphism in the ‘pyrin domain’, a component of about 20 regulatory proteins in the innate system. This genetic variant appears to enable enhanced inflammasome activity, with an increase in IL-1 secretion that mediates fever, inflammation and acute phase proteins including SAA – the precursor protein for AA amyloidosis. Many inflammatory conditions of unknown actiology, are being found to have defective regulation: e.g. C1 esterase (hereditary angioedema). C3b persisting on red cells (paroxysmal nocturnal haemoglobinurea) expanded the concept, leading to a search for rare defects in regulators and receptor systems of innate immunity such as IL-R antagonist (chronic inflammation characterised by purpura dermatitis and periarthritis inflammation) and TNFR defects (periodic fever – TRAPS). The concept spread further when named chronic inflammatory diseases such as Behcet’s Disease fitted emerging diagnostic criteria, and in reverse, when groups with chronic inflammatory disease characterised by fever, dermatitis, and synovitis were linked to genetic cryopyrin (a critical regulatory protein within the inflammasome – when defective is associated with several inflammatory diseases including hereditary cold urticaria). Cryopyrin was also found to be engaged by uric acid and pyrophosphate crystals to activate the inflammasome, to induce uncontrolled inflammation. Dissection of the molecular machinery of innate immunity will not only continue to expand this category of inflammatory disease and will identify targets for specific immunotherapy.

(f)  Dendritic cell Dysfunction

Dendritic cells influence outcome of all immune responses through the balance of cytokines and co-stimulating factors presented to the CD4 T cell. This is apparent in the characteristics, and response to therapy, of infection. For example, the balance of the Th1-Th2 response to infection with Chlamydia pneumoniae modulates the ‘load’ of atheroma, and with Helicobacter pylori, influences response to eradication therapy. Dendritic cell dysfunction may have a genetic basis, or reflect the balance of influences acting upon these cells. This little studied area is certain to provide many answers to unresolved clinical questions, as well as defining optimal therapeutic strategies.

(g)  Regional Communication Dysfunction (or ‘Consequential Disease’)

This intriguing concept is based on a breakdown in the normal “chatter” between cells based on marker expression where the consequent interplay of cytokine / receptor interaction causes progressive disease. This idea has been explored in Alzheimer’s Disease, against the background of “brain inflammation”. Neuronal ‘stresses’ trigger expression of B amyloid precursor protein (APP) by neurons, with secretion of a soluble fragment of APP (SAPP). This in turn activates microglial cells into an autocrine activity via an IL-12 / IL-23 ‘loop’ that enhances IL-1 secretion by the microglia. IL-1 further enhances secretion of APP from neurons, which under particular conditions leads to deposition of amyloid plaques with consequent cognitive decline. This concept of ‘consequential disease’ that began by damage in a particular genetic setting that initiates a sequence of poorly controlled inflammatory events that would normally be controlled, may have wider implications.