Molecular Analysis
of the Complement Component Six Gene

by
Gabrielle Jean Delbridge

A thesis submitted for the degree of
Doctor of Philosophy
of the Australian National University

July 1994

Statement

This thesis describes the results of a research project carried out under the supervision of Prof Sue Serjeantson in the Human Genetics Group, John Curtin School of Medical Research at the Australian National University, from April 1991 to July 1994.

The experiments and analyses presented in this thesis are my own work, except where otherwise acknowledged. The work described has not been submitted previously for a degree at this or any other university.

Gabrielle Delbridge

Gabrielle Delbridge

Acknowledgements

This thesis has been made possible with the help of many people. I hope that over the three years my gratitude has been obvious and that this acknowledgement is a mere formality.

I would like to acknowledge Dr RG DiScipio, Scripps Research Foundation, La Jolla, CA, USA for the donation of C6 cDNA clones. Thanks are also given to Dr Y Tokunaga, Blood Transfusion Service, University Hospital, Tokyo, Japan for protein typing the rare C6 patterns found in this study. Also Assis Prof W Hildebrand, University of Oklahoma, USA, for his help and information about the Isoelectric program.

Many thanks are due to Prof Sue Serjeantson for her guidance throughout this project and for the encouragement she gave me when problems appeared insurmountable. I truly appreciate the difficulties involved in my supervision and I believe Sue juggled her busy schedule and my impending deadline extremely well.

Dr Mike Hobart and Barbara Fernie are also greatly appreciated for their unending help and perpetual patience.

A special thankyou is due to Lynn Croft, Pippa Munro and Annette Browne whose company and help have kept me sane throughout the PhD process. I am especially grateful to Stewart Smith for his continuous technical resucitation, particularly in cloning, and for his unfaltering friendship and sense of humour.

I thank also my husband Troy for his acceptance and tolerance of me, my parents for their understanding and support and my sisters, Jennifer and Katherine just for being there.

I would like to say thankyou to Dr Simon Esteal, Dr Rohan Baker and Dr Maija Kohonnen-Corish for their help with critical reading of this thesis and Dr X Gao, Ingrid Jacobson and Dr Julleen Cavanaugh for their help with computing.

Acknowledgement is also due to the Biomolecular Resource Facility for their help with oligos and sequencing and the photography group for providing my photos.

Abstract

This study has examined complement component six (C6) protein alleles and restriction fragment length polymorphisms (RFLP) in Caucasian systemic lupus erythematosus (SLE) patients and healthy individuals. The protein studies showed that SLE patients and the healthy group had very similar protein allele distributions. Four rare-protein types in five patients were found and one in the healthy group. Both groups showed allele frequencies very similar to pooled published Caucasian data and allele frequencies did not differ from Hardy-Weinberg equilibrium values. Therefore there was no evidence that deficiency or partial deficiency of complement component six is a major contributor to inherited predisposition to SLE.

SLE patients and healthy individuals showed identical Taq I RFLP phenotypes. Allele frequencies for this polymorphic site were very similar in the SLE patients, healthy group and published Spanish data and these did not differ from Hardy-Weinberg equilibrium values. No major gene rearrangements were indicated in any of the samples examined. Three heterozygous rare-protein typed samples; BM91, BB2 and AB1 were Taq I RFLP typed and these showed identical patterns to common protein types.

The protein and Taq I RFLP alleles were not concordant, demonstrating that C6 Taq I RFLP analysis is not an adequate predictor of C6 protein type. Three methods of linkage disequilibrium estimation were utilised and all three demonstrated a significant association between the C6*A protein allele and RFLP 8.0 allele , as well as between the C6*B protein allele and RFLP 7.0 allele . However, measurements of linkage disequilibrium are of conflicting validity and a discussion of the arguments is presented.

The C6 protein has been known to be electophoretically polymorphic for more than ten years. The present study was aimed at determining the molecular basis of the common A/B protein allelism and four protein variants; A1, B2, M91 and B21. The technique of single-stranded conformation polymorphism analysis (SSCP) was chosen as a mutation screening method for the present study because of its simplicity, inexpense, sensitivity and the availability of information relating to

intron/exon boundaries. The entire C6 coding region was screened for polymorphisms using this technique and variant SSCP patterns were investigated by cloning and sequencing the variant samples.

A base difference between the two published C6 cDNA sequences was found to affect a Dde I restriction enzyme site, allowing investigation of this site using PCR and digest analysis. A perfect correlation between the common C6 protein allelism and Dde I pattern was found. Protein A alleles have an additional Dde I site, resulting from the sequence GAG at amino acid 98, a negatively charged glutamic acid residue, while B alleles lack this site and have a GCG codon, resulting in a non-polar alanine residue. The computer program, Isoelectric, was used to predict pI values for these alleles based on these sequence changes. C6*A alleles have a predicted pI of 6.64 and C6*B alleles 6.71. This is in good agreement with isoelectric focusing (IEF) results.

SSCP analysis resulted in seven variant patterns. One of these appeared to be an artifact. Five others, in exons; one, three, 10, 12 and 13 correlated with altered DNA sequences. A further variant did not correlate with a specific DNA change but four polymorphisms were sequenced from that exon. Three polymorphisms were not identified with SSCP, these related to protein alleles; A1, B21 and M91. Two variant SSCP patterns correlated with protein alleles, exon three with the common C6 A/B allelism and exon 12 with the B2 allele. The exon 12 polymorphism was identified as an A to G substitution, affecting amino acid 630. This causes an acidic, glutamic acid residue to be replaced with a polar, glycine molecule. The predicted pI for this protein allele is 6.71, assuming an A allele background, and 6.77 assuming a B allele background. These estimates are in good agreement with IEF results. Four DNA changes; three in exon 11 and one in exon 13 resulted in amino acid substitutions that altered the charge of the C6 protein. These however were not indicated with IEF, highlighting the complexities of predicting protein structure from DNA sequence data.

The level of polymorphism found in the C6 DNA sequence appears to be considerably greater than that known from C6 protein studies. A significant clustering of mutations was found in the C6 gene. Seven of nine

polymorphisms identified were located between exons 10 and 13. Exon 11 appears to be a particularly polymorphic region with four polymorphisms sequenced in this area and a further two indicated with PCR-digest analysis. This non-random distribution of polymorphisms may result from several genetic mechanisms, for example gene conversion, differential mutation rates and template mutagenesis.

Five C6 DNA polymorphisms, in exons; three 10, 11, and 13 were examined in several population groups; Aboriginal Australians, Caucasians, Micronesians and Polynesians. All of these polymorphisms were found in the groups examined, demonstrating that they predate emergence of these ethnic groups. Two of these polymorphisms were found in a small sample of chimpanzees indicating they were likely to be ancient mutational events, occurring prior to human-chimpanzee divergence around five million years ago. The occurrence and transmission of these polymorphisms across species boundaries provides further evidence for the multiple-origin and trans-species hypotheses of evolution. These theories suggest that the shared polymorphisms existed in a population of common ancestors that gave rise to the present-day species and that the polymorphisms have been maintained through selection, often up to millions of generations.

Definition of the molecular basis of C6 polymorphisms and the identification of further sequence changes enables the C6 gene to be utilised as a useful, polymorphic marker in population, evolution, disease-susceptibility and linkage studies.

Publications

Fernie BA, Delbridge G, Hobart MJ (1993) Correlation of a Glu/Ala substitution at position 98 with the complement C6 A/B phenotypes. *Human Molecular Genetics* **2**:591-592.

Fernie BA, Hobart MJ, Delbridge G, Potter PC, Orren A, Lachmann PJ, (1994) C6 Haplotypes: Association of a Dde I site polymorphism to complement deficiency genes and the Msp I RFLP. *Clinical and Experimental Immunology* **95**:351-356.

Table of Contents

List of Figures

List of Tables

Abbreviations

ATP	adenosine 5'-triphosphate
bp	base pairs
BSA	bovine serum albumin
C6, 7, 8, 9	complement components 6, 7, 8 and 9
cDNA	complementary DNA
C-test	complement test
dATP	2'-deoxyadenosine 5'-triphosphate
dCTP	2'-deoxycytosine 5'-triphosphate
dd	double distilled
ddNTP	dideoxynucleotide triphosphate
dGTP	2'-deoxyguanosine 5'-triphosphate
DIGE	direct gel electrophoresis
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
DTT	dithiothreitol
dTTP	2'-deoxythymidine 5'-triphosphate
EDTA	ethylenediamine tetra-acetic acid
EGF	epidermal growth factor
HLA	human leucocyte antigen
IAA	isoamyl alcohol
IEF	isoelectric focusing
IgG	immunoglobulin G
IPTG	isopropylthio-β-galactoside
kDa	kilodaltons
LDL	low density lipoprotein
MAC	membrane attack complex
MHC	major histocompatibility complex
NK	natural killer cells
PBL	peripheral blood leucocytes
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
RFLP	restriction fragment length polymorphism
SCR	short consensus repeat
SDS	sodium dodecyl sulphate
SLE	systemic lupus erythematosus
SSC	tri sodium citrate
SSCP	single-stranded conformation polymorphism
TAE	tris-acetate EDTA
TBE	tris-borohydrate EDTA
TCC	terminal complement components
TE	tris-EDTA
Tris	2-amino-2(hydroxymethyl)-1,3-propanediol
TSP	thrombospondin module
UV	ultraviolet
X-gal	5-bromo-4-chloro-3-indoyl-β-D-galactoside

Chapter 1

Chapter 1.

General Introduction

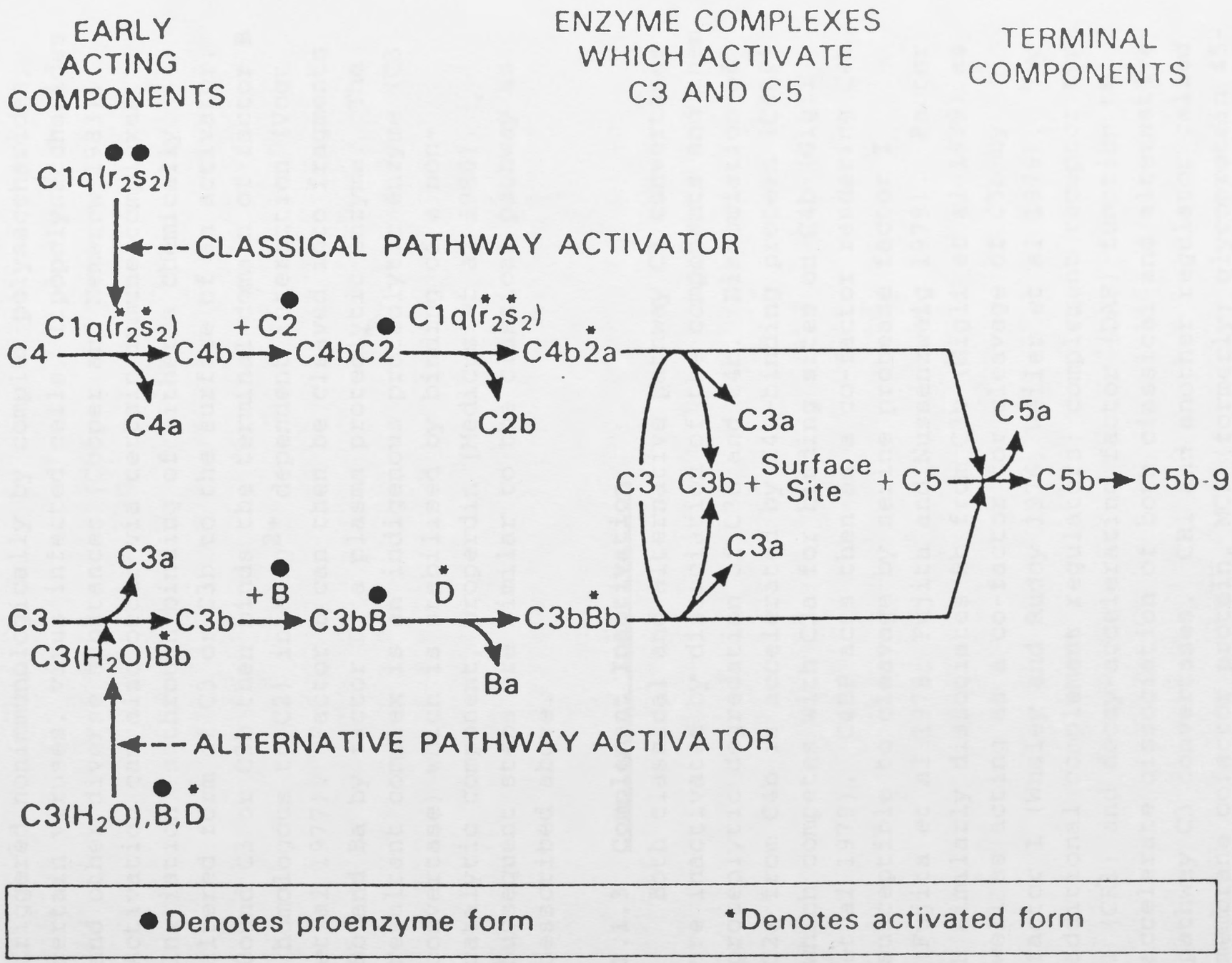
1.1 Function of the Complement System

The complement system is a major effector mechanism in humoral immunity and is also capable of lysing cellular antigens directly. Complement is activated by the presence of circulating immune complexes and also by less specific nonimmunoglobulin activators such as bacterial components, microbial particles and cell wall polysaccharides from other antigens (Bentley 1988). It consists of; 1) the classical pathway, 2) the alternative (or properdin) pathway, 3) the terminal membrane attack complex, into which 1) and 2) feed, and 4) complement regulators and receptors. These pathways are made up of approximately 30 distinct plasma proteins (Cooper *et al* 1988). Figure 1.1 shows the complement cascade diagrammatically.

1.1.1 Classical Pathway

Components of the classical pathway are C1, C2 and C4. C1 consists of three subunits C1q, C1r and C1s which are held together by calcium (Schumaker *et al* 1987). Activation of this pathway is through binding of immune complexes IgG or IgM to C1q, causing a conformational change within the C1 complex. This allows autoactivation of proenzyme C1r (Arlaud *et al* 1986) which in turn activates C1s (Valet and Cooper 1974). Down-regulation of this step is by C1 inhibitor, which binds tightly to activated C1 and its derivatives (Sim *et al* 1979, Ziccardi and Cooper 1979). The C1 complex then activates C4 to form C4b which attaches to the surface of the initiating immune complex. Utilising another site C4b interacts with the inactive form of C2. This interaction is Mg^{2+} dependent (Sitomer *et al* 1966) and probably involves the N-terminal domain of C2 (C2b) (Nagasawa and Stroud 1977, Kerr 1980). C2 is then activated by C1 to form the classical pathway C3 convertase. This molecule can then activate C3 by cleaving a polypeptide bond to yield two fragments; C3a and C3b. C3 convertases are subsequently modified by binding of further C3b so that they become C5 convertases (Medicus et al 1976, Vogt et al 1978). C5

Figure 1.1 The Complement Cascade. Major activation steps in the classical and alternative pathways are shown. Activation of the classical pathway is thought to be triggered primarily by immune complexes while the alternative pathway is initiated by a wide range of compounds and surfaces. These two pathways lead into the terminal cascade and cell lysis. Reproduced with permission, from Reid 1989, fig 1.


EARLY ACTING COMPONENTS
ENZYME COMPLEXES WHICH ACTIVATE C3 AND C5
TERMINAL COMPONENTS
C1q(r2s2)
CLASSICAL PATHWAY ACTIVATOR
C1q(r*2s*2)
C4
C4b
+C2
C4bC2
C1q(r*2s*2)
C4b2a
C4a
C2b
C3a
C5a
C3
C3b
+ Surface Site
+C5
C5b
C5b-9
C3a
C3
C3b
+B
C3bB
D
C3bBb
C3(H2O)Bb
Ba
ALTERNATIVE PATHWAY ACTIVATOR
C3(H2O),B,D
● Denotes proenzyme form
* Denotes activated form

convertases cleave C5 via proteolysis yielding C5a and C5b, thus initiating the terminal portion of complement; the membrane attack complex.

1.1.2 Alternative Pathway

The alternative complement pathway is usually triggered nonimmunologically by complex polysaccharides, certain viruses, virus infected cells, lipopolysaccharides and other diverse substances (Cooper and Nemerow 1983). Activation can also occur via certain immune complexes. Initiation is through binding of either a chemically altered form of C3 or C3b to the surface of an activator. Bound C3 or C3b then binds the terminal domain of factor B (homologous to C2) in a Mg^{2+} dependent interaction (Vogt *et al* 1977). Factor B can then be cleaved into fragments Bb and Ba by factor D, a plasma proteolytic enzyme. The resultant complex is an indigenous proteolytic enzyme (C3 convertase) which is stabilised by binding of a non-catalytic component, properdin (Medicus *et al* 1980). Subsequent steps are similar to the classical pathway as desscribed above.

1.1.3 Complement Inactivation

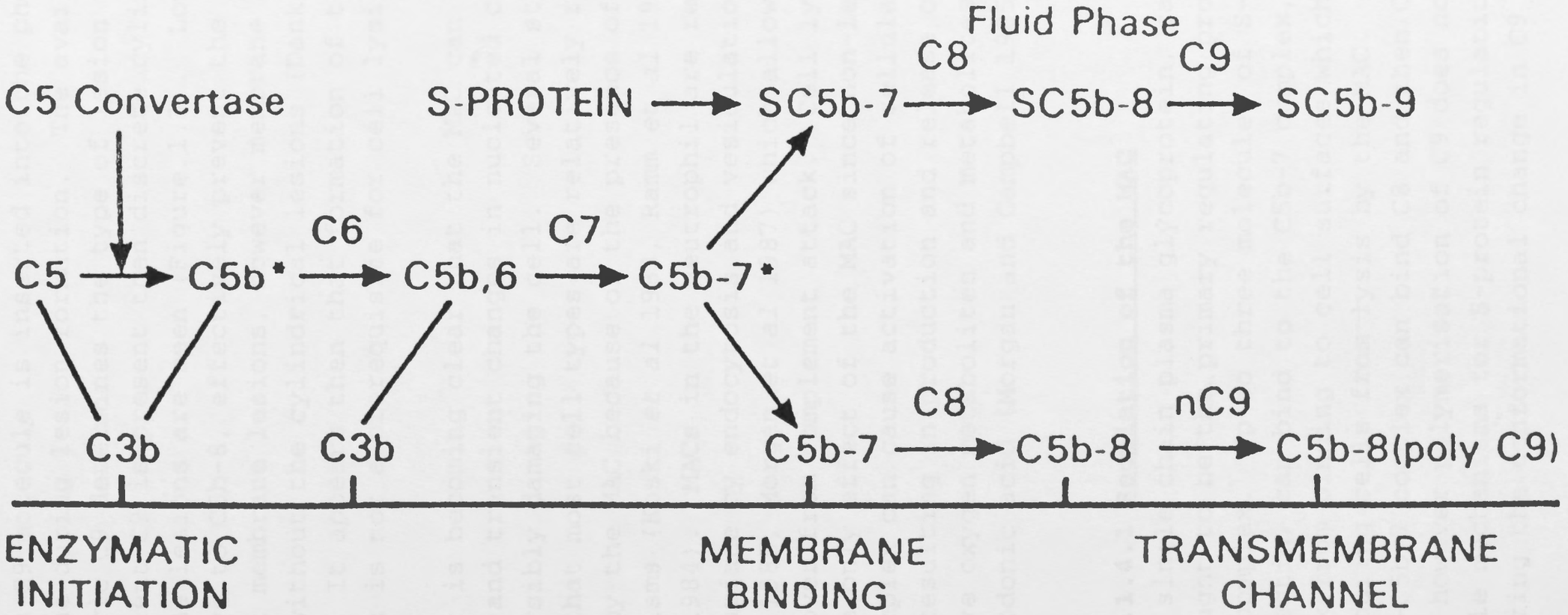
Both classical and alternative pathway C3 convertases are inactivated by dissociation of the components and then proteolytic degredation of C3b and C4b. Dissociation of C2a from C4b is accelerated by C4b-binding protein (C4BP) which competes with C2a for binding sites on C4b (Gigli *et al* 1979). C4BP acts then as a co-factor rendering C4b susceptible to cleavage by serine protease factor I (Fujita *et al* 1978, Fujita and Nussenzweig 1979). Factor H similarly dissociates Bb from C3b (Gigli *et al* 1979) as well as acting as a co-factor for cleavage of C3b by factor I (Whaley and Ruddy 1976, Weiler *et al* 1976). Two additional complement regulators; complement receptor type 1 (CR1) and decay-accelerating factor (DAF) function to accelerate dissociation of both classical and alternative pathway C3 convertases. CR1 and another regulator called membrane cofactor protein, MCP (formerly; glycoprotein 45-70 or gp45-70) act as cofactors for cleavage of C4b and C3b by factor I. Clearance of C3b/C4b coated immune complexes and microorganisms occurs via interaction of CR1

and a second receptor, CR2 with the C3b and C4b fragments (Holers *et al* 1985).

1.1.4 Membrane Attack Complex (MAC)

The terminal complement pathway is quite remarkable in that the water soluble hydrophilic components C5b to C9 undergo a hydrophilic-amphiphilic transition via a non-enzymatic self-assembly mechanism to form the MAC complex. This structure creates transmembrane channels by displacing lipid molecules and plasma membrane constituents, resulting in disruption of the phospholipid bilayer of target cells. Cell lysis and death follow this disruption (review Muller-Eberhard 1986). The MAC pathway is initiated after proteolysis of C5 by C5 convertase (both classical and alternative pathway forms) resulting in C5a and b. The activated C5; C5b, loosely bound to C3b, develops a transient binding site for C6 leading to formation of a stable C5b-6 dimer. Subsequent binding of C7 results in formation of a C5b-7 complex. This complex appears to undergo a hydrophilic-amphiphilic transition resulting in the expression of a high affinity lipid binding site that mediates a strong, non-covalent interaction between the developing complex and target membranes (Muller-Eberhard 1986). The C5b-7 complex then dissociates from C3b. Failure of the C5b-7 complex to bind a membrane surface at this stage results in the loss of its potential cytolytic activity. Self aggregation of the complex then takes place in the fluid phase (Figure 1.2). Binding of the three subunit protein C8 to the C5b-7 complex, utilising one subunit, the C8β chain, (Monahan and Sodetz 1981) is thought to bring about a conformational change in C8. This change allows the disulfide-linked α and possibly γ chains to penetrate the hydrophobic core of the lipid bilayer of the membrane to which the C5b-7 complex is attached. The exact function of C8γ, however remains unclear (Haefliger *et al* 1991). The C5b-8 complex is capable of slowly lysing erythrocytes and some nucleated cells however it primarily functions as a receptor for C9. C5b-8 binds C9 and in so doing catalyses polymerisation of C9 resulting in typical membrane lesions as seen in target cells. The MAC is composed of one molecule of each of C5b, C6, C7, C8 and somewhere between one and 18 C9 molecules (review Reid

Figure 1.2 Assembly of the Membrane Attack Complex. Components five to nine self-aggregate to form transmembrane channels, causing cell lysis. Control of this process is via S-protein. The asterisks (*) denote metastable forms of C5b and C5b-7, respectively. Reproduced, with permission, from Muller-Eberhard 1986, fig 1.


Fluid Phase
C8
C9
C5 Convertase
S-PROTEIN
SC5b-7
SC5b-8
SC5b-9
C6
C7
C5
C5b*
C5b,6
C5b-7*
C8
nC9
C3b
C3b
C5b-7
C5b-8
C5b-8(poly C9)
ENZYMATIC INITIATION
MEMBRANE BINDING
TRANSMEMBRANE CHANNEL

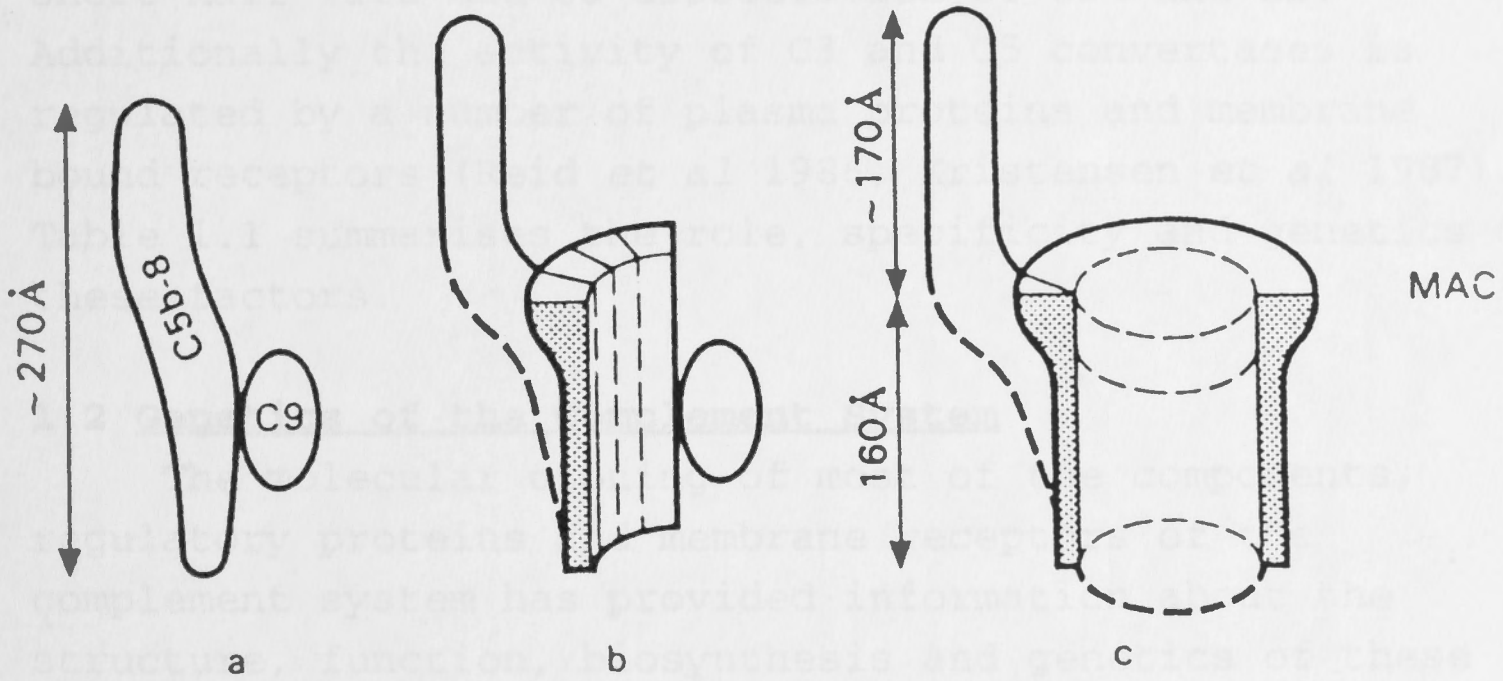
1988). Binding of the first molecule is through the C8 component but how the binding of one C9 molecule then allows high affinity C9-C9 interactions is not clear. It is thought that the relatively hydrophobic C-terminal half of each C9 molecule is inserted into the phospholipid membrane during lesion formation. The availability of monomeric C9 determines the type of lesion found, if sufficient C9 is present then discrete cylinder like membrane lesions are seen (Figure 1.3). Low levels of C9, relative to C5b-8, effectively prevent the formation of typical membrane lesions, however membrane damage can occur without the cylindrical lesions (Dankert and Esser 1985). It appears then that formation of the cylindrical complex is not a prerequisite for cell lysis.

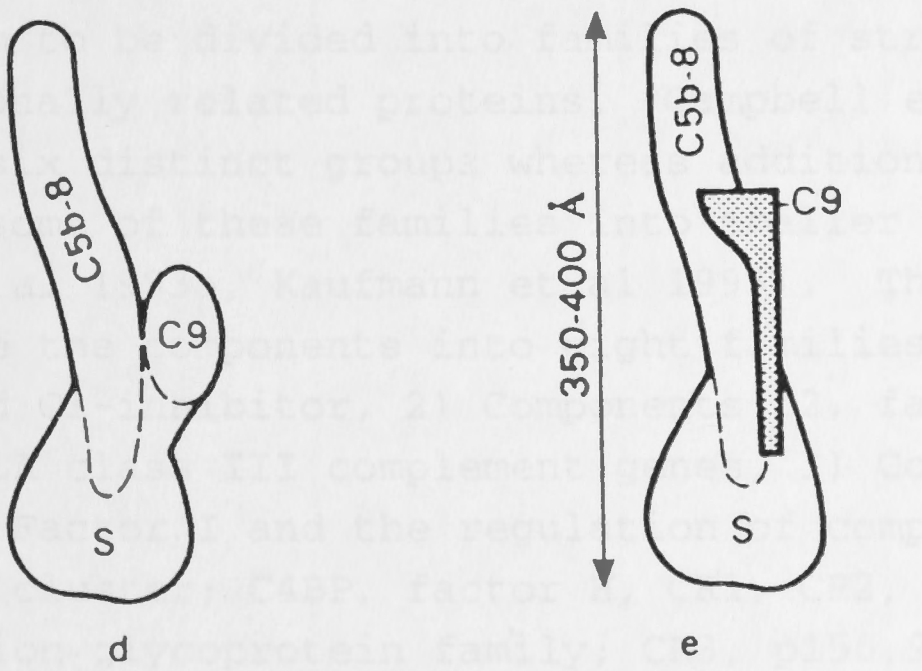
It is becoming clear that the MAC can cause some subtle and transient changes in nucleated cells without irreversibly damaging the cell. Several studies have shown that most cell types are relatively resistant to lysis by the MAC because of the presence of protective mechanisms (Koski *et al* 1983, Ramm *et al* 1983, Morgan *et al* 1984). MACs in the neutrophil are removed from the cell surface by endocytosis and vesiculation (Campbell and Morgan 1985, Morgan *et al* 1987) which allows these cells to recover from complement attack. Cell lysis is however not the only effect of the MAC since non-lethal amounts of the complex can cause activation of cellular processes, often resulting in production and release of toxic reactive oxygen metabolites and metabolites of arachiodonic acid (Morgan and Campbell 1985, Imagawa *et al* 1987).

1.1.4.1 Regulation of the MAC

A single chain plasma glycoprotein, called S-protein is thought to be the primary regulating protein of the C5b-9 complex. Up to three molecules of S-protein (or vitronectin) can bind to the C5b-7 complex, preventing the complex from binding to cell surfaces which protects neighbouring cells from lysis by the MAC. The resultant free S-C5b-7 complex can bind C8 and then C9, forming S-C5b-9, however polymerisation of C9 does not follow. Two possible mechanisms for S-protein regulation include; 1) preventing the conformational change in C9 that is

Figure 1.3 Models for C5b-9 and the fluid phase SC5b-9 complexes. Diagrams (a) to (c) represent the assembly of MAC C5b-8 complex in association with poly C9. Panels (d) and (e) represent models for SC5b-9 containing C9 in globular (d), or partially unfolded form (e). Reproduced, with permission, from Podack *et al* 1984, fig 7.


~270Å
C5b-8
C9
a
b
~170Å
160Å
MAC
c


C5b-8
C9
S
d
350-400 Å
C5b-8
C9
S
e
SC5b-9

apparently a prerequisite for polymerisation, or 2) allowing conformation changes to occur but causing steric hindrance of the polymerisation process (Podack *et al* 1984).

1.1.5 Complement Regulators and Receptors

Activation and control of the complement system is mediated in several different ways, some of which have been mentioned above. The C3 and C5 convertases have a short half life due to dissociation of C2a and Bb. Additionally the activity of C3 and C5 convertases is regulated by a number of plasma proteins and membrane bound receptors (Reid *et al* 1986, Kristensen *et al* 1987). Table 1.1 summarises the role, specificity and genetics of these factors.

1.2 Genetics of the Complement System

The molecular cloning of most of the components, regulatory proteins and membrane receptors of the complement system has provided information about the structure, function, biosynthesis and genetics of these glycoproteins. Such studies have also emphasised the similarities and differences between components which allows them to be divided into families of structurally and functionally related proteins. Campbell *et al* (1988) describes six distinct groups whereas additional work has separated some of these families into smaller groups (Hobart *et al* 1993a, Kaufmann et al 1993). This study will divide the components into eight families; 1) C1 complex and C1-inhibitor, 2) Components C2, factor B and C4 - the HLA class III complement genes, 3) Components C3 and C5, 4) Factor I and the regulation of complement activation cluster; C4BP, factor H, CR1, CR2, DAF, MCP, 5) Cell adhesion glycoprotein family; CR3, p150,95, and other receptor proteins, 6) Properdin, 7) MAC cluster I; C8α, C8β (C8γ), and 8) MAC cluster II; C6, C7 and C9,

1.2.1 C1 Complex and C1 Inhibitor

1.2.1.1 C1q

The C1 complex consists of subcomponents C1q, C1r and C1s. Molecular cloning of C1q A (Sellar *et al* 1987), B (Reid 1985) and C chains (Seller *et al* 1991) has provided cDNA probes which have been used to identify monocytes and

Table 1.1 Regulatory plasma proteins and membrane receptors of the complement system

Protein	Molecular weight (in kd)	Specificity	Chromosome location	Amino acids	Gene (kb)	Exons	Number SCRs	Role
C1-INH	110	C1r, C1s	11p11.2-q13	478	17	8	0	Inactivation of C1r and down regulation of complement activity
C4BP	540	C4b	1q32	3843 (7 x 549)	30	18	56	Cofactor for C4 degradation by factor I
MCP (CD46)	45-70	C4b, C3b	1q32	384	43	14	4	Similar cofactor activity to C4bp
DAF	70	C3 convertase	1q32	440/381	40	11	4	Regulation of C3 convertase
Properdin	220	C3bBb	Xp11.2-Xp11.3	442 (monomer)	6	10	6	Up regulation of C3bBb (C3 convertase)
Factor H	155	C3b	1q32	1213	90	?	20	Accelerates decay of C3bBb, acts as a cofactor for factor I

C1-INH = complement component 1 inhibitor, C4bp = complement component4 binding protein, MCP = membrane cofactor protein, DAF = decay accelerating factor,

Table 1.1 continued

Protein	Molecular weight (in kd)	Specificity	Chromosome location	Amino acids	Gene (kb)	Exons	Number SCRs	Role
CR1 A B C D	250] 290] 330] 210]	C3b, C4b	1q32	1800	133- 160	39 47 55? 31?	28	Regulation of C3b breakdown, binding immune complexes to erythrocytes phagocytosis, decay acceleration of C3, C5 convertases
CR2	145	C3dg, C3d	1q32	1013	30	19	16	Regulation of B cell functions, Epstein-Barr virus receptor
CR3	165 95	iC3b	β;21 α;16p	β;769	?		?	Aids cellular destruction of pathogens, enhances ingestion, possibly memory related
CR4	?	C3dg	?	?	?	?	?	Binds C3dg/C3d on neutrophils and platelets
S-protein (vitronectin)	80	C5b-9	?					Regulation of cytolytic damage by MAC

CR1, 2, 3, 4 = complement receptor type 1, 2, 3, and 4, ? = as yet unknown.

Table 1.1 continued

Protein	Molecular weight (in kd)	Specificity	Chromosome location	Amino acids	Gene (kb)	Exons	Number SCRs	Role
Factor I	88	C4b, C3b	4q25	583	25	?	0	Inactivates C4b/C3b convertase of alternative pathway
Factor D	24	Factor B-C3	?	228	?	?	?	Activates C3
p150,95	150,95	iC3b	?	?	?	?	?	Augments phagocytosis
C5a	45	C5a	?	350	?	?	?	Secretion, cellular activation, chemotaxis
C3e receptor	?	C3e	?	?	?	?	?	Leucocytosis
C3a receptor	?	C3a, C4a	?	?	?	?	?	Secretion, cellular activation
Factor H receptor	50	factor H	?	?	?	?	?	FactorI secretion blastogenesis, induction of respiration bursts

? = as yet unknown.
Data taken from review by Reid 1988 and a range of more recent research publications in text.

macrophages as the probable major source of serum C1q. Also arising from this work is the ability to identify mutations causing C1q deficiency. C1q deficiency can be classified into three forms; 1) partial, 2) complete and 3) complete functional deficiency (Loos and Heinz 1986, Reid 1989). A study by McAdam *et al* (1988) has identified the defect causing the deficiency in one case but it appears that C1q deficiency has several different causes. The genes for the A, B and C chains are in the order A-C-B on a 24 kb stretch of DNA on chromosome 1p. The A, B, and C chain genes are 2.5, 2.6 and 3.2 kb long respectively and each contains one intron. Sellar *et al* (1991) reports that there is only one gene per chain and no major gene rearrangements were found after RFLP analysis of C1q deficient individuals.

1.2.1.2 C1r and C1s

The C1r and C1s proteins are 166 kDa and are composed of an A and B chain. The genes for C1r and C1s have been cloned (Leytus *et al* 1986 and Tosi *et al* 1987) and localised to a 50 kb stretch of DNA on chromosome 12p13 (Tosi *et al* 1987). These proteins posess interesting repeating sequences, called short consensus repeats (SCR), which are found in a number of other complement components and regulators. These SCRs consist of approximately 60 amino acids and are characterised by a consensus of four conserved cysteine residues, highly conserved proline and glycine residues, a tryptophan residue and a tyrosine or phenylalanine residue.

Deficiencies of C1r and C1s have been reported (review Reid 1989) but are quite rare. As yet no genetic studies have described a mutation which may account for these deficiencies.

1.2.1.3 C1 Inhibitor (C1-INH)

C1-INH is a regulatory glycoprotein which has considerable homology with members of the serpin family of inhibitors of serine esterases. Genetic characteristics of C1-INH are listed in Table 1.1.

1.2.2 Components C2, Factor B and C4 - HLA Class III Complement Genes

C2, C4 and factor B genes form a tight cluster along with the 21-hydroxylase gene in the HLA region on chromosome 6p13 (Franke and Pellegrino 1977, Carroll *et al* 1988). The class I and II HLA genes on chromosome six code for highly polymorphic cell surface antigens involved in the immune response, the class III proteins C2, C4 and factor B are also polymorphic but show no structural similarity to class I and II gene products.

1.2.2.1 Complement Component Two (C2)

C2 is a 102 kDa serine protease consisting of 734 amino acids. The gene for C2 is approximately 18 kb (Campbell 1987) and consists of 18 exons (Ishii *et al* 1993). C2b possesses three SCR modules which are encoded by single exons. Deficiency of C2 is one of the most common human complement component deficiencies (Glass *et al* 1976) and more than half of the C2 deficient individuals have rheumatological disorders such as SLE, Henoch-Schonlein purpura and polymyositis. Johnson *et al* (1993) report that one common type of C2 deficiency is caused by a 28 bp genomic deletion that causes skipping of exon six during RNA splicing and subsequent generation of a premature termination codon. As yet the DNA mutation giving rise to the second type of C2 deficiency has not been identified.

1.2.2.2 Factor B

Factor B is a 92 kDa serine protease consisting of 739 amino acids. It is highly polymorphic with up to 18 alleles defined by differences in electrophoretic mobility in agarose gels (Alper *et al* 1972). Two of these alleles are common while the remaining 16 are rare. Campbell (1987) reports that there is only one amino acid residue difference between the two common alleles, consisting of a glutamine (F allele) to arginine (S allele) substitution at residue seven. The gene for factor B is six kb, cDNA 2388 bp and consists of 18 exons (Campbell *et al* 1985). Like C2, factor B is a mosaic structure made up of three domains apparently derived from three distinct gene superfamilies. Three exons exactly encode the three SCR modules within the Ba fragment. The middle domain of

factor B (residues 235-457) is structurally similar to the type A domain of von Willebrand factor (vWF)(Mancuso *et al* 1989). The carboxy terminal domain of factor B is structurally similar to serine proteases and contains eight exons which encode the active centre of the C3 and C5 convertases of the alternative pathway.

The homology of structure and function in the C2 and factor B proteins plus the close linkage of the genes suggests they arose from an ancestral gene by duplication. Despite these apparent similarities there is a ten fold difference in amounts of these proteins in plasma and a three fold difference in gene length. Homozygous deficiency of factor B has not been reported in man or other species, suggesting that it may be a lethal condition (Reid 1988) and reports of heterozygous deficiency are rare.

1.2.2.3 Complement Component Four (C4)

The C4 protein occurs as two isotypes, C4A and C4B which are distinguishable by their electrophoretic mobility on agarose gels and their relative hemolytic activity (Awdeh and Alper 1980). C4A and B are more than 99 percent homologous (Belt *et al* 1984) and the main difference between them appears to be the reactivity of their thiol esters (Bentley 1988). Four amino acids are responsible for this difference. The C4A and C4B proteins are highly polymorphic with at least 35 different alleles described (Mauff *et al* 1983). C4A and C4B are coded for by two genes in tandem, approximately 10 kb apart. Comparison of the C4A and C4B cDNA sequences revealed only 14 nucleotide differences (Belt *et al* 1984, 1985), and of these 12 were found in the C4d region, containing the potential thiolester site. These nucleotide changes give rise to nine amino acid substitutions (Belt *et al* 1984). One further difference in the C4d region (Belt *et al* 1985, Yu *et al* 1986) and six others determined by protein sequencing (Chakravarti *et al* 1983, 1987, Law and Gagnon 1985) suggest that 18 positions in the C4 molecule show amino acid substitutions. A nine bp deletion coding for residues 1397-1399 (Belt *et al* 1984) has been described in both C4A and C4B (Law and Gagnon 1985), but it is not clear whether this is real or a cloning artifact (Campbell

et al 1990). These amino acid changes relate to antigenic differences among allotypes, in addition to allelic differences. DNA sequencing of C4 alleles has shown that there are more alleles than was indicated with protein typing. For example two different C4A3 allotypes have been defined (Belt *et al* 1984, 1985, Yu *et al* 1986) which differ at position 1267 due to an alanine to serine substitution. Likewise two different C4B1 allotypes have been defined which differ at position 1157, due to an asparagine-serine substitution (Belt *et al* 1984, 1985) .

An interesting feature of the C4 genes is that they can differ in size. The gene at the first locus (C4A) is approximately 22 kb, whereas the gene at the second locus (C4B) can be either 22 kb or 16 kb in length (Yu *et al* 1986, Prentice *et al* 1986, Palsodittir *et al* 1987). This length difference is due to the presence or absence of six to seven kb of nucleotide sequence in an intron approximately 2.5 kb from the 5' end of the gene.

A further phenomenon in the genetics of C4 is the exceptionally high frequency of silent or null alleles. These null alleles (QO) result in the deficiency or partial deficiency of C4A or C4B protein. The frequency of the nulls in the population is around five to 15 percent for C4AQO and ten to 20 percent for C4BQO (Schendel *et al* 1984, Tokunaga *et al* 1985, Partanen and Koskimies 1986). The molecular basis of the null allele has been defined in several cases (Yu and Campbell 1987, Hauptmann *et al* 1987) and may help in our understanding of some autoimmune disease which are associated with this deficiency.

Cases of duplication of the C4 genes have also been reported (Rittner *et al* 1984a, Raum *et al* 1984, Uring Lambert *et al* 1984) at an estimated frequency of one to two percent, although this is probably an underestimate as three new kinds of C4 gene duplication have since been identified from protein and RFLP analysis (Schneider *et al* 1986).

1.2.3 Complement Components Three and Five (C3, C5)

C3 is the most abundant complement protein and it plays a major role in the activation of the system since it participates in both the classical and alternative pathways. The complete amino acid sequence has been deduced from the cDNA sequence (de Bruijn and Fey 1985). The C3 protein consists of two polypeptide chains, α; 992 residues and β; 645 residues.

The C3 gene is 41 kb and consists of 41 exons (Fong *et al* 1990). The C3 β chain spans 13 kb from exon one to exon 16. Exon 16 encodes both α and β chains. The α chain is 28 kb and contains 26 exons including exon 16. The C3 gene has been localised to chromosome 19 (Whitehead *et al* 1982)

C3 protein exhibits two common allotypic variants that are separated by gel electrophoresis and are called C3 fast (C3 F) and C3 slow (C3 S). In addition to these forms at least 20 rare allelic variants have been described. Botto *et al* (1990) report that the two alleles differ by a single change at nucleotide 364 with S alleles posessing an arginine and F alleles a glycine residue at this site. The molecular basis of a second structural polymorphism, defined by a monoclonal antibody (Koch and Behrendt 1986) was also characterised by Botto's group. This change occurred at codon 314 in exon nine of the β chain where a leucine residue was substituted for a proline molecule. Three RFLPs have also been mapped to introns in the C3 gene, but no allelic association between them has been found (Botto *et al* 1990). Homozygous deficiency of C3 in humans is relatively rare and is associated with pyogenic infections and nephritis.

Like C3, mature C5 is a disulphide linked dimer consisting of an α and β chain. Components C3, C4 and C5 along with α-macroglobin show a high degree of homology on comparison of amino acid sequences (Sottrup Jensen *et al* 1985). C5 however, differs from the others since it does not contain an internal thiolester bond and therefore lacks the covalent binding properties associated with activation.

The human C5 gene is 79 kb, cDNA 3309 bp, and is comprised of 41 exons (Carney *et al* 1991). The gene has been localised to chromosome 9q32-34 (Wetsel *et al* 1988). Comparison of C5 with C3 and C4 revealed strong similarities in exon size and number. The 5' flanking region of the gene contains sequences homologous with interferon, interleukin-6, gluccocorticoid, estrogen, NF-$_k$B and HNF-1, all of which are known responsive elements.

The C5 protein is not polymorphic in Americans, English, Portuguese or West Africans (Hobart *et al* 1981, Rosenfeld *et al* 1977) but is in Melanesians (Vaz-Guedes *et al* 1978). C5 deficiency has been reported in at least 19 individuals most of whom had recurrent infections (review Hauptmann 1989, review Morgan and Walport 1991). C5 deficient individuals lack bactericidal activity and have severely impaired ability to induce chemotaxis (Miller and Nilsson 1970, Nilsson *et al* 1974, Rosenfeld *et al* 1976). As yet no studies have determined the molecular basis for the C5 polymorphism or deficiency.

1.2.4 Factor I and Regulation of Complement Activation Group (RCA)

Factor I is a serine protease involved in the degredation of complement component three. Like the majority of complement proteins it is composed of a number of protein modules; two low density lipoprotein (LDL) receptor class A and B repeats, a factor H like SCR, and a CD5 module. Further genetic characterisics of factor I are listed in Table 1.1.

The proteins of the RCA family are all C3b/C4b binding proteins which share a structural organisation. These proteins are characterised by having their entire or almost their entire sequence composed of SCRs (Chung *et al* 1985, Kristensen *et al* 1986, Klickstein *et al* 1987, Caras *et al* 1987, Medof *et al* 1987, Weis *et al* 1988). These proteins are encoded by linked genes which have been assigned to chromosome 1q32 (Rodriguez de Cordoba *et al* 1984, 1985, Rey-Campos *et al* 1987, Weis *et al* 1987, Lublin *et al* 1987). The order of the genes is MCP, CR1, CR2, DAF, C4BP while the factor H gene is 6.9 cM away from these. Table 1.1 lists their genetic characteristics.

1.2.5 Cell Adhesion Glycoprotein Family and other Complement Proteins

CR3, p150,95 and LFA-1 are all members of a family of cell surface antigens that share a common 95 kDa β chain which is non-covalently linked to a distinct α chain. Each αβ complex participates in some form of cell adhesion activity (Springer *et al* 1987). The cDNA sequence of the common β subunit has been cloned and sequenced, showing that it is composed of 747 residues (Law *et al* 1987, Kishimoto *et al* 1987). The gene for this β chain is approximately 32 kb (Kishimoto *et al* 1987) and has been mapped to chromosome 21 (Marlin *et al* 1986). A DNA polymorphism has been described using RFLP analysis of Bgl II digests. Deficiency of these cell adhesion molecules usually includes all three members, suggesting a defect in the common β chain. Patients deficient in these molecules often experience recurrent bacterial and fungal infections (Springer *et al* 1987). Other less well characterised complement factors are also listed with available information in Table 1.1.

1.2.6 Properdin

Properdin is a glycoprotein which circulates in the blood in the form of oligomers of a 56 kd chain. (Smith *et al* 1984). The primary amino acid sequence of properdin is composed mainly of six SCRs. Genetic features of properdin are listed in Table 1.1. Deficiency of properdin is rare and is inherited as an X-linked recessive disorder. Deficient individuals are predisposed to life-threatening bacterial infections (Figueroa and Densen 1991) however early detection, utilising a linked polymorphic dinucleotide repeat, and immunisation are helpful.

1.2.7 Membrane Attack Cluster I - C8 (MAC I C8)

C8 is a 151 kDa glycoprotein constituent of the C5b-9 complex. C8 along with C9 has the ability to circulate in plasma as a hydrophilic protein but undergoes a hydrophilic to amphiphilic transition leading to interaction with target membranes. C8 has an unusual subunit structure consisting of three components; α (64 kd), β (64 kd) and γ (22 kd) (Klob and Muller-Eberhard 1976, Steckel *et al* 1980). These are arranged as a

disulphide-linked α–γ dimer that is noncovalently associated with β.

The asymmetrical arrangement of noncovalent and covalently associated C8 subunits is unusual for a serum protein but is consistent with the fact that each subunit arises from a separate gene (Ng *et al* 1987, Howard *et al* 1987, Rao *et al* 1987). The amino acid sequence of subunit α was derived from cDNA analysis and consists of 553 residues plus a leader sequence of 30 amino acids. The N- and C-termini are both cysteine rich while the central region is relatively free of cysteines. The N-terminus exhibits strong homology to the 40-residue repeat sequence found in the LDL receptor. The C-terminus exhibits homology to epidermal growth factor (EGF) precurser and several other proteins including urokinase, blood coagulation factors FIX and FX and tissue plasminogen activator (Rao *et al* 1987). It also has striking overall homology to human C9.

Amino acid sequence of C8 β was derived from a 2.0 kb cDNA sequence (Howard *et al* 1987, Haefliger et al 1987). This subunit has a 54 amino acid leader sequence, followed by 536 residues corresponding to mature β. Like C8 α the N- and C-terminii of C8 β are cysteine rich while the central region is poor in cysteines. The LDL receptor and EGF precursor consensus sequences are again present as are other segments of homology to α.

The C8 γ amino acid sequence has been derived from a 718 bp cDNA clone (Ng *et al* 1987). It consists of 182 residues preceeded by a 20 amino acid signal peptide. The sequence contains three cysteines, only one of which is linked to α (Haefliger *et al* 1991). C8 γ displays a striking amino acid sequence similarity to a group of proteins called lipocalins (Pervaiz and Brew 1987). The common feature of all lipocalin proteins is their molecular mass of around 22 kd and their ability to transport small ligands, suggesting a similar role for C8 γ.

As noted above the three subunits of C8 are coded for by three separate genes. The initial idea that this was

the case arose from studies of the C8 α–γ polymorphism (Raum *et al* 1979, Rittner *et al* 1984b) and the C8 β polymorphism (Alper *et al* 1983, Rogde *et al* 1985). In Caucasians C8A*A (previously named C81*A) and C8A*B (C81*B) occur at frequencies of approximately 0.588 and 0.374 respectively (Rittner *et al* 1993). The C8B*1 (previously C82*1) and C8B*2 (C82*2) alleles in Caucasians occur at frequencies of approximately 0.952 and 0.044 respectively (Alper *et al* 1983). Rare alleles for these loci also exist, with 12 being identified for C8A, mostly occurring in Japanese groups, and one for the C8B locus, designated C8B*A1 (C82*A1) (Alper *et al* 1983). Two major types of C8 deficiency are known at the protein level (Tedesco *et al* 1983). These two types are restricted to different ethnic groups with C8 α–γ deficiency found in Blacks while C8 β deficiency is found in Caucasians (Ross and Densen 1984). C8 deficient individuals often suffer from recurrent neisserial infections such as meningitis.

The genes for C8 subunits are designated C8A for α, C8B for β and C8G for γ. C8A and B are closely linked on chromosome 1p32 (Theriault *et al* 1992) while C8G is located on chromosome 9q (Kaufman *et al* 1989). The gene for C8B is approximately 40 kb, and consists of 12 exons ranging in size from 69 to 347 bp (Kaufmann *et al* 1993). All intron-exon junctions follow the GT-AG rule. Two RFLPs have been reported for the C8B gene (Hermann *et al* 1989, Rogde *et al* 1989a), neither of which correlates with the protein polymorphisms. Kaufmann *et al* (1993) study reports that the polymorphic Taq I site occurs in intron 11. A further study by the same group (Kaufmann *et al* 1993) describes a DNA mutation which is a major cause for C8 β deficiency. A single substitution of a thymine for cytosine in exon nine results in a premature stop codon. This mutant allele was observed in all C8 β deficient families investigated.

The genomic organisation of the C8A and C8G genes has not yet been reported. The C8A gene shows a Taq I RFLP and a possible polymorphic variable number of tandem repeats (VNTR) has been suggested (Rogde *et al* 1991a).

1.2.8 Membrane Attack Cluster II (C9, C7, C6)

1.2.8.1 Complement Component Nine

C9 is a single chain, 70 kDa, glycoprotein of 537 amino acids (DiScipio *et al* 1984). It consists of three cysteine rich domains; a thrombospondin (TSP) repeat and a LDL receptor domain at its N-terminus plus an EGF domain at its C-terminus. The central region of the C9 sequence has strong sequence similarities with perforin, a pore forming protein secreted from cytotoxic T lymphocytes (Shinkai *et al* 1988). This central region appears to possess some specific hydrophobic properties.

The C9 protein is not polymorphic, but deficiencies of C9 are well known. Japanese groups have been shown to have a high incidence of C9 deficiency, approximately 0.045 to 0.104 percent (Akagaki and Inai 1985 as reported by Nagata *et al* 1989). There is an association between C9 deficiency and meningoccocal meningitis, but it is not as strong as that for other MAC component deficiencies.

The C9 gene has been cloned (DiScipio *et al* 1984, Marazziti *et al* 1988) and found to have 11 exons. Analysis of the intron/exon boundaries showed good correlation between splice sites and surface features of the protein but not with the protein homology structures (Marazziti *et al* 1988). This was considered surprising since the same sequence in the LDL receptor is precisely bound by introns, suggesting that this sequence is present in both proteins as a result of exon shuffling.

The gene for C9 has been localised to chromosome 5p13, along with the C6 and C7 genes (Abbott *et al* 1989, Rogne *et al* 1989b) it is however located more than 2.5 Mb from these genes (Setien *et al* 1993). Several Taq I RFLPs have been described for the C9 gene (Rogne *et al* 1990, Coto *et al* 1990a). As yet the sites determining these RFLP patterns have not been located.

1.2.8.2 Complement Component Seven (C7)

Complement component seven is a 120 kDa single chain glycoprotein. The complete primary structure of C7 has been derived from the cDNA sequence (DiScipio *et al* 1988). C7 is a mosaic protein consisting of 821 amino acids, the

carboxy terminal third contains four cysteine-rich segments. The protein has 28 disulfide bonds and is glycosylated at two sites. Almost all of the cysteines are found in small units of 35-77 amino acids, these regions are homologous with various proteins including LDL receptor, EGF precursor, TSP, factor I-like domain (FIM), and C4 and blood coagulation factors IX and X.

The C7 protein was first found to be polymorphic in 1976 (Hobart and Lachmann) with three structural forms being identified. These were found to be controlled by three codominant alleles; C7*1, 2 and 3 (Hobart *et al* 1978). Four more alleles were soon identified; C7*4 (Nakamura *et al* 1984a), C7*6 (Zeng *et al* 1986), C7*7 (Washio *et al* 1986) and C7*8 (Komatsu *et al* 1989). Allele C7*5 was described by Nishimukai and Tamaki (1986) but was later shown to be identical to C7*3 (Washio *et al* 1986). All of these alleles were identified using isoelectric focusing and since then further polymorphisms (C7*9, C7*M and C7*N) of the C7 protein have been identified using monoclonal C7-allospecific antibody in an ELISA assay (Wurzner *et al* 1990, 1992b). Caucasian groups show one common autosomal codominant allele (C7*1) and two rare alleles with a frequency of less than 0.01 (Hobart *et al* 1978). In Japanese three common C7 alleles are found (Nakamura *et al* 1984a) and these correspond to those found in Caucasians. However the two rare alleles were found at higher frequencies (0.15) in Japanese. In Chinese one of the rare alleles C7*2 is also more frequent than in Caucasians (York *et al* 1986). Hobart (1990) reports that Oriental populations have both a higher frequency of variants and a greater range of protein allotypes compared with Caucasians.

Deficiency of the C7 protein has been reported in at least 20 families (review Wurzner *et al* 1992a). Neisserial infection is the most common clinical manifestation, however deficient individuals may be healthy. C7 deficiency has been seen to co-occur with C6 deficiency in two families (Lachmann *et al* 1978, Morgan *et al* 1989) and with C4B deficiency in four further cases (Chapel *et al* 1987, Lopez-Trascasa *et al* 1988, Wuillemin

et al 1991, Segurado *et al* 1992). As yet there has been no DNA mutation reported which explains the C7 deficiency.

The complete cDNA encoding the C7 protein was sequenced in 1988 (DiScipio *et al*). The genomic organisation of the C7 gene was investigated by Hobart *et al* (in preparation). The C7 gene is approximately 80 kb in length and consists of 17 exons, ranging in size from 86 to 244 bp. The intron sizes range from 0.5 to 8.5 kb. The C7 gene is very similar to that for C6, C8B and C9. All intron/exon boundary phases are identical for C6 and C7.

A Taq I RFLP for the C7 gene has been described (Coto *et al* 1990b) and identifies two alleles 4.2 and 4.0 kb. The estimated frequencies of the two alleles in a Spanish group are 0.32 and 0.68 for the 4.2 and 4.0 kb alleles respectively.

1.2.8.3 Complement Component Six (C6)

C6 is a single chain glycoprotein with a molecular weight of approximately 120 kD. The polypeptide chain contains 913 amino acids with a 21 residue signal peptide (DiScipio and Hugli 1989, Haefliger *et al* 1989). C6 is a mosaic protein which is composed of nine distinct modules. These homology units include three TSP type I, an LDL type A receptor, an EGF precursor, two SCRs and two FIM modules. The C6 polypeptide chain is cross-linked by 32 disulfide bonds which occur in short discrete segments. Like C8, C7 and C9 the C6 protein has a central cysteine-poor region.

The C6 protein was first found to be polymorphic in 1975 (Hobart *et al*) and has been shown to be so in all major ethnic groups. Nineteen different allotypes besides the two common ones have been described (review Tokunaga *et al* 1990). Table 3.1 (Chapter 3) gives the frequencies of the common alleles plus the combined frequency of the rare protein alleles in all the groups examined. As discussed in Chapter 3 deficiency of C6 is well documented and is often associated with recurrent neisserial infections, however several cases of an association with autoimmune diseases have also been reported (Tedesco *et al*

1981, Trapp *et al* 1987, Reinitz *et al* 1986, Wisnieski *et al* 1985). As yet there has been no report of the mechanisms determining C6 deficiency at the DNA level.

The gene for C6 was cloned and sequenced in 1989 (Haefliger *et al* 1989, DiScipio and Hugli 1989). The cDNA sequence is 3.3 kb while the genomic DNA covers approximately 85 kb (Hobart *et al* 1993a). The C6 gene is located on chromosome 5p13 (Jeremiah *et al* 1990, Abbott *et al* 1989) and is closely linked to C7 and C9 (Hobart *et al* 1978, 1993b, Tokunaga *et al* 1986, Abbott *et al* 1989, Rogne *et al* 1991b, Setien *et al* 1993). The intron-exon boundaries have been defined (Hobart *et al* 1993a) with the gene consisting of seventeen exons ranging from 91 bp to 774 bp, most being around 165 bp. Two RFLPs for the C6 gene have been described; one involves a polymorphic Taq I site (Coto *et al* 1991a) while the second occurs at an Msp I site (Coto *et al* 1991b). The Msp I RFLP has been associated with C6 deficiency (Fernie *et al* 1994). Two C6 and C7 sub-total deficient (SD) families, plus two families with C6 subtotal deficiency only, appear to possess an exon three (protein allele site) A allele , Msp I positive, and C6SD haplotype, while 16 of 17 C6Q0 individuals possessed an exon three B allele, Msp I negative, C6Q0 haplotype. It appears then that the C6Q0 and C6SD mutations arose independantly on two different allelic backgrounds (Fernie *et al* 1994).

1.3 Evolution of the Complement System

The complement system in mammals is a highly successful recognition and effector system. The classical pathway of complement can be triggered by antibody-antigen complexes and the alternative pathway has a major role in recognition of foreign organisms. While the complement system is highly effective in eliminating antigens, unlike immunoglubulin-type genes which have a vast array of possible rearrangements, the complement system is not able to quickly adapt or memorise antigens. This is a severe limitation in an environment where many antigens are evolving faster than their potential host. Thus it appears that the combination of systems would greatly enhance the survival of the individual, suggesting that

presence of both facets of immunity represents a highly evolved system.

There is some evidence that suggests primitive parts of the complement system occur in some invertebrates while vertebrates possess a more developed system. It is generally agreed that invertebrates do not possess rearrangeable immunoglobulin genes while all vertebrates examined do (Litman and Kehoe 1978). Therefore it appears that the ancestral complement system evolved before the adaptive immune recognition mechanisms (review Farries and Akinson 1991).

1.3.1 Alternative and Classical Pathway Evolution

A form of the alternative pathway has been found in several invertebrate phyla, including Arthropods, Echinoderms and Annelids (Ballow 1977). An equivalent protein to the alternative pathway component C3 has been found in birds (Kai *et al* 1983, Laursen and Koch 1989), reptiles (Petrella *et al* 1989), amphibians (Grossberger *et al* 1989, Avila and Lambris 1990), bony fish (Nonaka *et al* 1984b), sharks (Jensen *et al* 1981) and lampreys (Nonaka *et al* 1984a). Conserved features of this protein are the chain structure and internal thioester bond (Alsenz *et al* 1984). Thus it appears that the first functional complement component could have been an ancestral version of C3 (review Farries and Atkinson 1991). At some later time it is thought that a gene fusion event created a protease, a primitive factor B protein, with a binding site for activated C3. Evolution of the other components then allowed amplification and specificity of this pathway to develop.

Several studies show that plasma of birds, reptiles, amphibians and fish possess antibody-dependent cytolytic activity similar to the classical pathway of complement while the nurse shark possesses components equivalent to C1, C2 and C4 (Jensen *et al* 1981, Ross and Jensen 1973). Farries *et al* (1990) suggest that C1q evolved first as the link between antibodies and effector cells. Possibly an ancestral C1r-C1s complex then arose with the function of binding C1q to C3. Duplication of the C3 gene may then have occurred allowing specialisation of C4 and C1.

Duplication of the factor B gene could have given rise to C2 allowing further divergence and specialisation.

1.3.2 Evolution of the Terminal Components

Sharks have a lytic system which includes C8- and C9-like components (Jensen *et al* 1981), while C5 has been identified in trout (Nonaka *et al* 1981). Farries and Atkinson (1991) suggest this terminal cascade evolved around the same time as the classical pathway, more than 400 miliion years ago (Hobart *et al* 1993a). C6 and C7 have not been detected in the nurse shark and are thought to have evolved later (DiScipio and Hugli 1989, Hobart *et al* 1993a).

1.3.3 Genetic Mechanisms of Evolution of the Early Components of the Complement System

Insights into the origin of complement components can be gained from homologies with other proteins. Like numerous other proteins many complement components are mosaics and are made up of structural units similar to parts of other proteins. These individual module structures are usually encoded for by one or several exons with intron/exon boundaries coinciding with protein unit boundaries in the majority of cases (review Patthy 1987). Patthy (1985) suggests that exon shuffling is an important mechanism in the reassortment of these modules across different protein groups and these insertion/deletion events are facilitated if intron/exon boundaries correspond to protein module boundaries. This phenomenon is thought to be further facilitated if exons are in the same phase at their 5' and 3' ends (Patthy 1987). Through this mechanism of exon shuffling proteins with novel combinations of domains, and thus new functions, are thought to have been created. Since these basic units are so ancient and widespread it is not possible to deduce which recombinations have produced the new proteins. Some genetic events however, are simpler to reconstruct, for example gene duplication by unequal crossing over (Farries and Atkinson 1991) probably explains the similarities and linkage between factor B and C2, plus C4A and C4B, C1r and C1s and possibly the RCA genes. C3, C4 and C5 also share considerable structural similarities indicating gene duplication despite their lack of linkage.

Intragenic duplication is also thought to be an important mechanism in the evolution of this system, particularly for the RCA group. All these proteins contain tandem arrays of short consensus repeats and analysis of intron/exon boundaries indicates intragenic duplication of blocks of four SCRs from an ancestral four SCR structure (Fujisaku *et al* 1989).

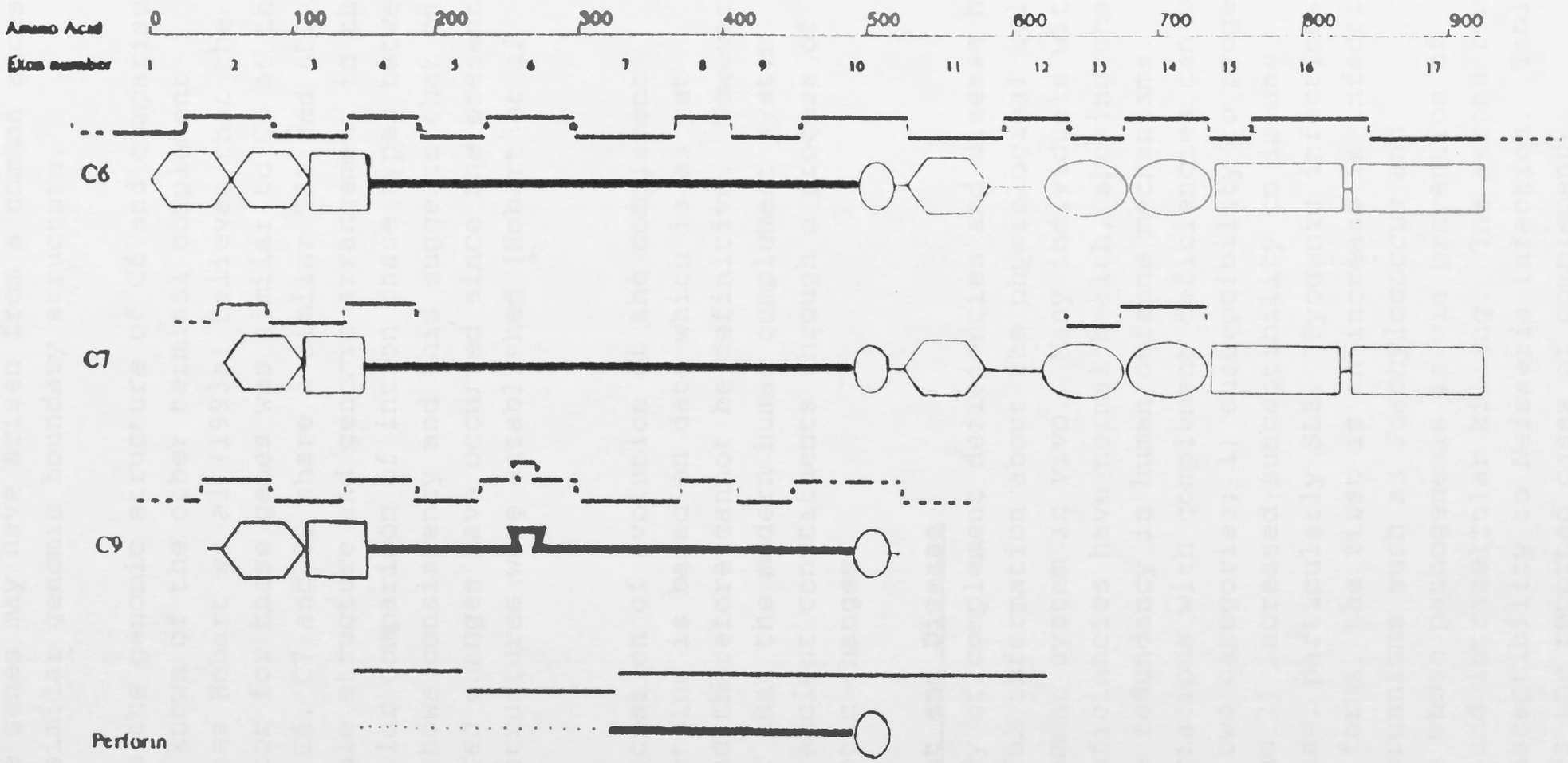
1.3.4 Genetic Mechanisms in the Evolution of Components of the Membrane Attack Complex

The common structural features of the terminal components suggest they evolved from a common ancestor. The structural arrangement of these proteins is shown in Figure 1.4. In general these proteins differ in their complexity with C6 being most complex and C9 least. All of these proteins possess a central cysteine-poor region which has similarities with perforin, a pore-forming protein that acts as an effector molecule for killer T-cells and natural killer (NK) cells (Shinkai *et al* 1988, Lichtenheld *et al* 1988). Terminal complement proteins also possess cysteine rich areas which have similarities with modules in the low density lipoprotein receptor type A and B, thrombospondin and the epidermal growth factor receptor (DiScipio *et al* 1984, Stanley *et al* 1985, Howard *et al* 1987, Rao *et al* 1987, Chakravarti *et al* 1988, 1989, DiScipio and Hugli 1989, Haefliger *et al* 1989). C6 and C7 also have SCR's. Another homology module occurring in the C6 and C7 proteins is the FIM domain (DiScipio and Hugli 1989, Haefliger *et al* 1989, DiScipio *et al* 1988). C6 is one of the most complex of known mosaic proteins with great diversity in constituent protein homology units (Hobart *et al* 1993a).

The mechanism of exon shuffling is thought to be important in evolution of the MAC components, but characteristics facilitating this mechanism are uncommon in the C6 and C7 genes (Hobart *et al* 1993a). In these genes only one of the protein elements (SCR2) follows the cotermination convention and phases of intron/exon boundaries are not consistent at their 5'and 3' ends.

Comparison of the intron/exon structures of the C6, C7 and C9 genes shows them to be very highly conserved

Figure 1.4 A comparative map of the genetic organisation of C6, C7, C9 and perforin. Amino acids and exon numbers for C6 are given. Protein homology units are depicted as open-boxed or circled regions. In general exons do not correspond to protein unit structures. Reproduced, with permission, from Hobart *et al* 1993a, fig 2.


Amino Acid
0
100
200
300
400
500
600
700
800
900
Exon number
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
C6
C7
C9
Perforin
Thrombospondin
LDL receptor
EGF precursor
SCR
Factor I
Exon, non-coding
Non-homologous sequence
Cysteine-poor region
Exon, coding
Assumed exon (C9)

(Hobart *et al* 1993a) and perhaps suggests alternative mechanisms of exon shuffling, not reliant on cotermination of DNA and protein units or phase consistency. For example these genes may have arisen from a common ancestor which had a similar genomic boundary structure.

Based on the genomic structure of C6 and comparisons with what is known of the other terminal complement component genes Hobart *et al* (1993a) believes that the common ancestor for these genes was similar to C6 at the 5' end since C6, C7 and C9 share a similar TSP and LDL receptor module structure and genomic arrangement in this area. A similar comparison of intron phase types between these genes shows consistency and this suggests that only very restricted changes have occurred since the present intron/exon structures were established (Hobart *et al* 1993a).

This discussion of evolution of the complement component proteins is based on data which is as yet incomplete and therefore cannot be definitive. However it appears clear that the modern human complement system evolved from ancient constituents through a process of feasible genetic changes.

1.4 Complement and Disease

The study of complement deficiencies and disease has provided useful information about the physiological role of the complement system *in vivo*. Many individuals with complement deficiencies have normal health, showing that there is some redundancy in human defence mechanisms. Disease associations with complement deficiencies can be grouped into two categories; 1) susceptibility to pyogenic infections and 2) increased susceptibility to immune complex disease, particularly SLE. Pyogenic infections can take two forms, the first is an increase in infections by pyogenic organisms such as *Staphylococcus* and *Streptococcus* whose pathogenesis is via prevention of phagocytosis and intracellular killing. The second form is through susceptibility to *Neisseria* infection. Table 1.2 summarises the reported cases of complement deficiencies and their disease associations. Complement deficiencies are uncommon yet their study has increased

Table 1.2 Inherited complement component deficiencies: chromosomal location of the genes and disease associations.

Complement component	Chromosome location	Number of cases	Reported associated disease
C1q A chain	1p	>40	SLE in majority, SLE-like disease, DLE, pyogenic infections including meningitis
B chain	1p		
C chain	1p		
C1r, C1s	12p13	C1r/C1s 16	SLE-like disease, pyogenic infections, often recurrent
		C1r and C1s 4	2 DLE and arthritis, 2 healthy
C2	6p13	>100	SLE, SLE-like disease, DLE, pyogenic infections, glomerulonephritis, rheumatoid purpura, vasculitis, healthy
C3	19	16	Pyogenic infections, glomerulonephritis, SLE
C4 A	6p13	35% pop. lack 1 allele	SLE, SLE-like disease, DLE, vasculitis,
B	6p13	8-10% lack 2 alleles	Rheumatoid purpura.
		1% lack 3 alleles	SLE, DLE, SLE and infections,
		17 cases lack 4 alleles	SLE and glomerulonephritis,
C5	9q32-34	30	Neisserial infections, SLE
C6	5p13	>100	Neisserial infections, SLE-like disease, or healthy
C7	5p13	>80	Neisserial infections, SLE, healthy schleroderna, Rheumatoid Arthritis,

Table 1.2 continued

Complement component	Chromosome location	Number of cases	Reported associated disease
C8 α	1		Neisserial infection, SLE-like disease,
β	1	>80	
γ	9p	α–γ <20	
C9	5p13	5 (Caucasians) >500 (Japanese)	Usually healthy Healthy, Neisserial infections
Factor B	6p13	0	-
Factor D	?X	2	Recurrent bacterial infections
C1-INH	11p11.2-q13	many, >500	Hereditary angioneurotic edema, SLE-like disease (rare)
Factor I	4	15	Pyogenic infections, glomerulonephritis SLE
Factor H	1q32	12	Pyogenic infections, glomerulohephritis SLE, haemolytic uraemic syndrome
Properdin	X	>50	Neisserial infections, rarely other pyogenic infections

Table 1.2

Complement component	Chromosome location	Number of cases	Reported associated disease
C4bp	1q32	3	Atypical Behcets disease
DAF	1q32	2	Deficiency on erythrocytes; paroxysmal nocturnal hemoglobinuria with deficiency of the CR1 receptors
CR3	21	complete 4 partial 15	Increased susceptibility to bacterial infections

SLE = systemic lupus erythematosus, DLE = discoid lupus erythematosus,
Data from Morgan and Walport 1991, Hauptman 1989 and references in text.

the knowledge of the role of complement in the human body. The importance of the membrane attack complex in host defense against *Neisseria* has been highlighted and a greater understanding of mechanisms of infection has developed.

The mechanism underlying the association with immune complex disease is not well established, however complement promotes the removal of immune complexes from the circulation and tissues. Deficiency of this function may allow accumulation of immune complexes causing inflammation and release of autoantigens. This in turn stimulates the production of autoantibodies and more immune complexes (Lachmann and Walport 1987). While homozygous complement deficiency is a rare cause of SLE it represents the strongest susceptibility genotype identified thus far with more than 80 percent of cases of homozygous deficiency of C1q and C4 having SLE (review Morgan and Walport 1991)

1.5 Systemic Lupus Erythematosus

Systemic lupus erythematosus (SLE) is a multisystem autoimmune disease of unknown etiology. It is characterised by excessive quantities of autoantibodies and immune complexes, leading to inflammation in many different organs (Steinberg 1985). The lupus erythematosus cell (LE cell) phenomenon was discovered in 1948 by Hargraves and was later shown to represent an antibody to deoxyribose nucleoprotein (DNP). Many other autoantibodies, particularly against nuclear antigens, have since been found in the sera of SLE patients. Consequently the fluorescent antinuclear antibody (ANA) test was developed and has been used as a routine diagnostic test for SLE since the late 1950's.

1.5.1 Incidence

SLE is a disease that primarily affects women, with 90 to 95 percent of patients in their teens, twenties and thirties being female. It can occur at any age, however 70 percent of patients experience onset of disease between ages 13 and 40 years (Steinberg 1985). SLE has been reported to occur in approximately one in 2000 Caucasoids (Engleman and Shearn 1983) while Meddings and Grennan (1980) report an incidence of one in 6780 New Zealand

Caucasoids. The disease is significantly more common in black Americans than American Caucasians and North American indian tribes the Sioux, Crow and Arapahoe show an even greater predisposition to SLE (Steinberg 1985). Estimates of the incidence of SLE are somewhat variable and possibly reflect both the heterogeneity of the disease and the ethnicity of populations from which samples were taken. As yet there is no large scale survey of the prevalence of SLE in the Australian population.

1.5.2 Etiology

The etiology of SLE is unknown, however the symptoms are thought to be due to excess autoantibodies that react with self components leading to inflammation. Initiation of this process may involve many factors both genetic and environmental and may differ substantially between patients.

Genetic factors in SLE have been implicated particularly through studies of family aggregation of the disease (Arnett 1987), twin studies (review Block *et al* 1975) and inherited disorders of the complement system (review Morgan and Walport 1991). Possible genetic contributors to the disease include genes that allow excessive antibody response after stimulation, genes that predispose to particular antibody responses, hormonal regulatory genes and genes governing metabolic control.

Several studies (DeHoratius and Messner 1975, DeHoratius *et al* 1975, Arnett and Schulman 1976; Kaplan 1984) have also shown that environmental factors contribute to the disease process. Suggested contributory environmental agents include certain foods, drugs, UV light and infections. These factors appear to act on the genetically determined immune system to predispose or protect an individual from SLE expression.

1.5.3 Diagnosis of SLE

SLE is a heterogenous disease in its expression varying from a mild cutaneous problem to a severe, life threatening, multi-organ disease. It may also involve stages of active disease and remission with a wide range of clinical symptoms. The diversity of symptoms seen in

SLE patients makes diagnosis difficult, particularly as many are non-specific and similar to manifestations of other connective tissue diseases. It is possible that SLE is a collection of several diseases and does not represent a uniform group of patients.

SLE patients examined in this study were diagnosed according to the criteria of Tan *et al* (1982). Antinuclear antibodies, anti-DNA antibodies, complement levels and complement activity have been shown as highly specific for the diagnosis of SLE (Passas *et al* 1985). Table 1.3 shows the classification criteria as taken from Tan *et al* (1982). For diagnosis of SLE, patients in this series had any four of the criteria either serially or simultaneously during the course of observation. These criteria are exclusive rather than inclusive and are designed to exclude other autoimmune connective tissue diseases, arthritides and systemic illnesses, unfortunately also excluding some individuals who are likely to have SLE (Tan *et al* 1982).

1.5.4 Clinical Manifestations of SLE

Clinical features of SLE are diverse, Table 1.4 shows a number of the symptoms seen and their incidence, these frequencies can vary between studies and may be influenced by how the patient was ascertained. Cutaneous disorders include malar rash, discoid lesions, photosensitivity and mouth and nasal ulcerations, these features are common with approximately 85 percent of SLE patients having a dermatological disorder (Rothfield 1985). Patients are particularly affected by the typical "butterfly" rash across the cheeks and nose and sensitivity to sunlight which may exacerbate the malar rash. Systemic manifestations of SLE involve the joints, lungs, kidneys, liver, central nervous system, spleen and gastrointestinal tract.

1.5.5 Pathogenesis

SLE is usually classified as an immune complex disorder, however this is thought to be an over simplification (Steinberg 1985). SLE is mediated primarily by antibodies but details of pathogenesis are not well proven. The production of injurious

Table 1.3 The 1982 revised criteria for classification of systemic lupus erythematosus.

Criterion		Definition
1. Malar rash		Fixed erythema, flat or raised, over the malar eminences, tending to spare the nasolabial folds
2. Discoid rash		Erythematous raised patches with adherent keratotic scaling and follicular plugging; atrophic scarring may occur in older lesions
3. Photosensitivity		Skin rash as a result of unusual reaction to sunlight, by patient history or physician observation
4. Oral ulcers		Oral or nasopharyngeal ulceration, usually painless, observed by a physician
5. Arthritis		Nonerosive arthritis involving 2 or more peripheral joints, characterized by tenderness, swelling, or effusion
6. Serositis	a)	Pleuritis—convincing history of pleuritic pain or rub heard by a physician or evidence of pleural effusion *OR*
	b)	Pericarditis—documented by ECG or rub or evidence of pericardial effusion
7. Renal disorder	a)	Persistent proteinuria greater than 0.5 grams per day or greater than 3+ if quantitation not peformed *OR*
	b)	Cellular casts—may be red cell, hemoglobin, granular, tubular, or mixed
8. Neurologic disorder	a)	Seizures—in the absence of offending drugs or known metabolic derangements, e.g., uremia, ketoacidosis, or electrolyte imbalance *OR*
	b)	Psychosis—in the absence of offending drugs or known metabolic derangements, e.g., uremia, ketoacidosis, or electrolyte imbalance
9. Hematologic disorder	a)	Hemolytic anemia—with reticulocytosis *OR*
	b)	Leukopenia—less than 4,000 mm^3 total on 2 or more occasions *OR*
	c)	Lymphopenia—less than 1,500 mm^3 on 2 or more occasions *OR*
	d)	Thrombocytopenia—less than 100,000 mm^3 in the absence of offending drugs
10. Immunologic disorder	a)	Positive LE cell preparation *OR*
	b)	Anti-DNA: antibody to native DNA in abnormal titer *OR*
	c)	Anti-Sm: presence of antibody to Sm nuclear antigen *OR*
	d)	False positive serologic test for syphilis known to be positive for at least 6 months and confirmed by *Treponema pallidum* immobilization or fluorescent treponemal antibody absorption test
11. Antinuclear antibody		An abnormal titer of antinuclear antibody by immunofluorescence or an equivalent assay at any point in time and in the absence of drugs known to be associated with "drug-induced lupus" syndrome

* The proposed classification is based on 11 criteria. For the purpose of identifying patients in clinical studies, a person shall be said to have systemic lupus erythematosus if any 4 or more of the 11 criteria are present, serially or simultaneously, during any interval of observation

Reproduced, with permission, from Tan *et al* 1982, table 2.

Table 1.4 Frequency of clinical features in SLE.

Feature	Percent
Fever	83
Weight loss	62
Arthritis, arthralgia	90
Skin	74
Butterfly rash	42
Photosensitivity	30
Mucous membrane lesions	12
Alopecia	27
Raynaud's phenomenon	17
Purpura	15
Uriticaria	8
Renal	53
Nephrosis	18
Gastrointestinal	38
Pulmonary	47
Pleurisy	45
Effusion	24
Pneumonia	29
Cardiac	46
Pericarditis	27
Murmurs	23
ECG changes	39
Lymphadenopathy	46
Splenomegaly	15
Hepatomegaly	25
Central nervous system	32
Psychosis	15
Convulsions	15

From Schur (1982).

autoantibodies is well demonstrated in the association of renal disease where the degree of pathology is largely determined by the extent of antibody deposition and the magnitude of the resultant inflammatory process. Continuous deposition and inflammation leads to irreversible tissue damage. Antibody deposition however can occur in other tissues without induction of an inflammatory response. Thus the affinity, isotypes and charge of an antibody along with the size and configuration of the antigen are critical to disease pathogenesis.

Antibodies in SLE patients also bind cell surfaces. Antibodies to erythrocytes, granulocytes, lymphocytes and macrophages are produced by SLE patients. These antibodies cause the cells to be removed from circulation or killed (Steinberg 1985). The inflammatory lesions characteristic of SLE are initiated by antibody and occur primarily in small vessels. Other problems of SLE patients; CNS disorders, hematologic, dermal and cardiac dysfunction have additional pathogenic mechanisms, as yet not all are well understood.

Richardson *et al* (1990) reports that DNA methylation in T-cells is impaired in patients with SLE and rheumatic arthritis (RA). They suggest a model in which T-cell DNA hypomethylation alters gene expression and generates autoreactive T-cells which then participate in the pathogenesis of SLE and RA.

1.5.6 Diagnostic Tests

As yet there is no perfect test for SLE however the presence of large amounts of antibodies to native DNA is a very useful diagnostic finding. Several other tests; the lupus band test, LE cell test and fluorescent antinuclear antibody test (ANA) have some usefulness in SLE diagnosis despite their concurrent detection of other disorders. At present it is possible to measure antibodies specifically reactive with various antigens, a procedure with potentially greater diagnostic use than older methods.

1.5.7 Treatment

The variable severity of SLE requires an individualised treatment. The disease is usually handled by treating symptoms as they arise. Management of major organ involvement is aimed at preservation of function and prevention of failure. Drugs such as corticosteroids, prednisone, methylprednisone, cyclophosphamide, azathiopine and several others have some use in specific circumstances (Lockshin 1991).

1.6 Aims

The aim of this study is to examine the complement component six protein allele distribution and C6 Taq I RFLP allele frequencies in Caucasian patients with SLE as compared to a healthy group of Caucasians. The protein and RFLP polymorphisms will be analysed for a possible association to determine whether RFLP typing can replace protein typing. C6 is a polymorphic protein and the molecular basis for the common and four rare protein alleles will be investigated. The presence of these DNA polymorphisms will be examined in several population groups from the Pacific region, to establish whether these polymorphisms arose prior to emergence of these ethnic groups. A small group of chimpanzee samples will also be examined for the presence of these polymorphisms in order to determine whether these sequence changes predate speciation.

Chapter 2

Chapter 2
Materials and Methods

2.1 Materials

2.1.1. Chemicals, enzymes, radioisotopes

The materials used in this project were obtained from the following suppliers:

Ajax chemicals - Auburn, NSW Australia
NaOAc, $MgCl_2$, HCl, KCl, NaCl, $MgSO_4$, EDTA, isoamyl alcohol, ethanol, CH_3COOH, isopropanol, bromophenol blue, $(NH_4)_2S_2O_8$, Na_2HPO_3, boric acid, KH_2PO_4, 1-butanol

Amersham International - North Ryde, NSW, Australia
Multiprime labelling system, Nick translation kit

Amresco - Gymea, NSW, Australia
Acrylamide

Applied Biosystems (a division of Perkin Elmer Corporation) - Burwood, Vic, Australia
Dye labelled primers, deoxy/dideoxynucleotide triphosphate mixes

Australian Medical and Scientific (distributors of Amicon products)- Chatswood, NSW, Australia.
Microconcentrators

Bacto Laboratories - Liverpool, NSW, Australia
Tryptone, yeast extract, agar

BDH Chemicals - Melbourne, Victoria, Australia
Sodium dodecyl sulphate ultra pure, NH_4OAc, polyethylene $glycol_{6000}$, dichloromethylsilane solution, amyl alcohol, tri-sodium citrate,

Bio-Rad Laboratories - North Ryde, NSW, Australia
N,N'-methylene-bis-acrylamide

Boehringer Mannheim - Castle Hill, NSW, Australia
Alkaline phoshatase, X-gal, Hae III, M buffer, Msp I, L buffer, Mae II, dNTPs

Bresatec - Adelaide, SA, Australia
$\alpha\ ^{32}P$ dATP, lambda Hind III/Hind III Eco RI geneclean kit

Calbiochem Novabiochem - Alexandria, NSW, Australia
C6 monoclonal antibody

Fluka Chemische Fabrik - Gymea, NSW, Australia
Formamide, dimethylformamide, taurine, riboflavin
Hanimex - Brookvale, NSW, Australia
X-ray film
Immunodiagnostics (distributors of Atlantic Antibodies) - Camperdown, NSW, Australia
Alkaline phosphatase - anti goat IgG (H + L)
Linbrook International (distributors for LKB-Bromma) - Crows Nest, NSW, Australia
Ampholine pH 5-7 and pH 7-9
Mallinckrodt - Victoria, Australia
Glycerol, glucose, $CaCl_2$
May and Baker Limited - Victoria, Australia
Na_2CO_4, NH_4Cl_2, $CHCl_3$
Merck - Kilsyth, Vic, Australia
NaOH, $NaBH_4$, gelatin
New England Biolabs - Arundel, Qld, Australia
Dde I, Bsm I, Aci I, Scr FI, Acc I, Mnl I, Bsr I,
Novachem - South Yarra, Vic, Australia
phenol
Pharmacia Biotechnology - North Ryde, NSW, Australia
Dextran sulfate, T4 polynucleotide kinase, T4 DNA ligase, dNTPs, ATP, M13mp18 RF1 DNA, lambda Hind III, One-phor-all-buffer-plus, Sma I, pBR322
Polaroid Australia - North Ryde, NSW, Australia
Polaroid film
Promega - Sydney, NSW, Australia
Thermus aquaticus (Taq) DNA polymerase, Taq I
Sigma-Aldrich Chemical Company - Castle Hill, NSW, Australia
Bovine serum albumin (BSA), Tris, agarose type II, proteinase K, mineral oil, DTT, ethidium bromide, IPTG, salmon sperm DNA, Tween 20, formaldehyde, nitro blue tetrazolium (NBT), $AgNO_3$, thiamine, 5-bromo-chloro-3-indoyl phosphate (BCIP), N-lauroyl sarcosine, $MnCl_2$

2.2 Methods

2.2.1 C6 isoelectric focusing gels

Native serum samples were subjected to isoelectric focusing (pH 5-9) in flat bed polyacrylamide gels (Hobart *et al* 1975) followed by immunoblotting (Whitehouse and Putt 1983). Proteins were transferred to polyvinylide difluoride membrane by capillary transfer for 2 hr. Membranes were then blocked with 0.5% (w/v) bovine serum albumin (BSA)-phosphate buffered saline (PBS) at 4°C overnight and then rinsed twice in 0.025% (v/v) Tween20-PBS. The first antibody, goat antihuman C6 diluted one in 700 with 2% (w/v) BSA-Tween20-PBS, was added and incubated for 90 min at room temperature. Membranes were washed with 0.025% (v/v) Tween20-PBS five times followed by incubation for 90 min at room temperature with the second antibody, anti goat IgG alkaline phosphatase conjugate diluted one in 1000 with 2% (w/v) BSA-Tween20-PBS. Membranes were stained using the 5-bromo-chloro-3-indoyl phosphate (BCIP) and nitro blue tetrazolium (NBT) detection system (Blake *et al* 1985)

2.2.2 DNA extractions

Genomic DNA was prepared from peripheral blood leucocytes, buffy coats and haemolysates. Most samples analysed in this study were prepared in previous studies using the routine method of phenol/chloroform extraction (Maniatis *et al* 1982). Additional samples were prepared using a salting-out procedure (Miller *et al* 1988). Cells were pelleted and resuspended in tris EDTA (TE) (10 mM Tris 1 M EDTA). Haemolysates required 4-6 washes in PBS before being pelleted and resuspended. 1% (w/v) SDS and 0.1 mg/100 µL proteinase K were added to the cells and incubation was at 37°C overnight. Saturated sodium acetate (250 µL/mL) was added and tubes were vortexed for 15 sec to allow precipitation of proteins. Samples were then centrifuged and supernatants transferred to fresh tubes. DNA was precipitated by addition of an equal volume of isopropyl alcohol, transferred to microfuge

tubes, dried and resuspended in double distilled water (ddH_2O).

2.2.3 Estimation of DNA concentration and purity

The concentration of DNA preparations was estimated using a Gilford spectrophotometer at 260 nm assuming that a 50 µg/mL solution of double-stranded DNA has an absorbance of 1.0 at 260 nm in a 1.0 cm path length cell. The purity of the DNA preparation was judged by examining the ratio of absorbance at 260 nm and 280 nm. Pure DNA has a 260:280 absorbance ratio of 1.8.

2.2.4 PCR amplification of DNA

The polymerase chain reaction (Saiki 1985) was used extensively in this study to amplify C6 exons from genomic DNA and cDNA fragments from a lambda clone containing C6 cDNA (from R Discipio CA, USA). Reactions were carried out according to the Taq polymerase manufacturers recommendations (Promega) in volumes from 10-100 µL, overlayed with mineral oil in 0.5 mL microfuge tubes. Radiolabelling PCR reactions were done in an Ericomp Twinblock (^{TM}) Thermal cycler (San Diego, USA) and non-radioactive PCRs were performed on an Innovonics Gene Machine (Bartelt Industries, Melbourne, Australia). Optimal reaction conditions for amplification with each primer pair (sequences of which are located in the Appendix) were determined experimentally. Cycles consisted of an initial denaturation at 96°C for 4 min with subsequent cycles 1 min, annealing for 1 min at the appropriate temperature and extensions at 72°C for 1 min. Annealing temperatures were calculated for each primer pair (Suggs *et al* 1981) and the lower temperature was chosen initially. Thirty five cycles were routinely performed. Monitoring of contamination was done by including a negative control in all sets of amplifications. This control contained all reaction components except template DNA. Reaction mixtures excluding DNA were set up in a DNA-free area that was regularly illuminated with UV radiation to reduce contamination risk.

2.2.5 Restriction endonuclease digestion of DNA

For restriction fragment length polymorphism (RFLP) analysis, approximately 10 µg of DNA was digested in a 50 µL volume containing 5 µL of stock buffer (supplied) and 50 units of the appropriate restriction endonuclease according to the manufacturers protocol for incubation time and temperature.

Digestion of PCR products was carried out as above but in a 10-20 µL volume using approximately 250 ng of amplified DNA and 1-5 units of enzyme. The restriction enzyme Aci I required the addition of 1% gelatin to the reaction mix and replacement of the recommended New England buffer 2 with New England buffer 4. Under these altered conditions the enzyme proved more reliable.

2.2.6 Gel electrophoresis of DNA

Agarose gels were prepared by dissolving the appropriate amount of agarose in Tris-acetate-EDTA (TAE) buffer (0.04 M Tris/acetic acid, 0.002 M EDTA pH 7.8). Ethidium bromide was added to a final concentration of 0.5 µg/mL.

The percentage of agarose in gels varied depending on the size of DNA fragments being examined. Small fragments, for example 100-300 bp, were run on 2% (w/v) gels. Total genomic digests were electrophoresed on 0.7% (w/v) gels in order to separate the larger fragments.

To examine PCR amplified C6 exons 5 µL of each reaction product plus 2 µL of loading dye (30% v/v glycerol, 0.25% w/v bromophenol blue) was electrophoresed in a 2% (w/v) agarose gel at 120 V for 30 min. After electrophoresis fragments were made visable using a UV transilluminator at wavelength 254 nm.

Acrylamide solution (40% w/v) was made up to a ratio of 19:1 acrylamide to N,N-methylene-bis-acrylamide. The solution was de-ionised by stirring with mixed bed resin then filtered to remove resin particles. It was then

degassed under vaccuum, protected from light and stored at 4°C. Ammonium persulphate solution (10% w/v) was made fortnightly and stored at 4°C.

Polyacrylamide gels were made by diluting the above acrylamide solution with TBE then polymerising it by addition of 0.625% (v/v) ammonium persulphate and 0.075% (v/v) Temed. Digested PCR products were separated on 160 x 160 x 1.5 mm polyacrylamide gels using Hoefer supplied plates and tank apparatus in TBE buffer. Gels were run at 30 mA for 2 to 4 hours, then stained with ethidium bromide and photographed. In cases where band sizes were very small it was necessary to silver stain the gel instead. The method used was an adaptation of Merril *et al* 1981 and involves soaking the gel in 10% ethanol/0.5% acetic acid for 3 min twice followed by 10 min in 0.1% $AgNO_3$. The gel was then washed twice in ddH_2O and soaked in 1.5% NaOH, 0.01% $NaBH_4$, 0.15% formaldehyde until bands became visible. The reaction was then stopped by a 10 min incubation in Na_2CO_3. Gels were photographed and if desired dried and stored.

2.2.7 Southern blot and restriction fragment length polymorphism analysis

2.2.7.1 Transfer of DNA on to membrane

Digested DNA was transferred from the agarose gel onto nylon membrane (Biotrace HP) according to the method of Southern (1975) with modifications by Reed and Mann (1985). Prior to transfer, gels were de-purinated in 0.25 M HCl for 10 min to facilitate transfer of large fragments. Membranes were rinsed in ddH_2O then alkali transferred using capillary blotting. Gels were blotted overnight and then rinsed in 2 x SSC (0.3 M NaCl, 0.03 M sodium citrate [pH 7.0]) air dried, fixed under UV light (254 nm) for 3 min then dried for 2 hr at 65°C.

2.2.7.2 Hybridisation of DNA on nylon membranes

According to the protocol of Nasmyth (1982) three membranes and a nylon sheet were soaked in 2 x SSC, rolled

and placed into glass bottles. Pre-hybridisation mix containing 50 µg/mL of pre-boiled sonicated salmon sperm DNA, 10 mL Nasmyth solution and 10 mL ddH_2O, was added to the hybridisation bottles. Membranes were pre-hybridised at 65°C for approximately 18 hours on first use, subsequent pre-hybridisation periods were 2 hr. Nasmyth solutions (1.1 M NaCl; 0.333 M Na_2HPO_4; 0.111 M EDTA [pH 6.2]; 1.85% N-laurylsarcosine and 18.5% dextran sulfate) were made fortnightly and stored at 4°C.

Hybridisation solutions were essentially identical to pre-hybridisation solutions. Sonicated salmon sperm DNA, 10 mL ddH_2O and labelled probe were boiled for 10 min. 15 mL of fresh, warmed Nasmyth solution was then added to the probe mix. The pre-hybridisation mix was discarded and hybridisation mix was added quickly to the bottles. Membranes were incubated at 65°C for 12 to 48 hours in a Hybaid oven.

Hybridisation solutions were discarded following incubation and membranes were washed first at room temperature for 10 min in 2 x SSC, 0.1% SDS. Consecutive washes in 1 x SSC, 0.1% SDS and 0.5 x SSC 0.1% SDS were carried out at 65° for 10-30 min depending on radioactivity levels. Membranes were checked after each wash step. When counts reached approximately 2-20 cps the membranes were briefly air dried then wrapped in plastic. X-ray film was exposed to the membranes for 2-14 days at -70°C. Intensifying screens, Lighting Plus (Du Pont) were included in the cassette to enhance results. Development was done in a Kodak X-Omat M20 automatic processor.

Membranes were stripped using 0.1% (w/v) SDS and 0.1 x SSC. 500 mL of solution was boiled and poured over membranes. When the solution cooled to room temperature the procedure was repeated until counts were reduced to background. Moist membranes were stored in plastic wrap and used several times.

2.2.7.3 Radio-labelling of DNA

Probes were PCR-amplified fragments of C6 cDNA and were prepared using random priming (Feinberg and Vogelstein 1983,1984). Reactions were performed using the Amersham Multiprime DNA labelling systems following the supplier's protocol. Reaction mixes contained 0.25 µg PCR product (boiled for 2 min and then kept on ice),4 µL of each unlabelled dCTP, dGTP, dTTP, 5 µL α ^{32}P dATP (3000 mCi/mM) 5 µL reaction buffer, 5 µL hexanucleotide primer buffer, 2 units DNA polymerase I - Klenow fragment and ddH_2O to 50 µL. Reaction mixes were incubated at 37°C for 1 hr. Labelled probes were then precipitated by addition of 2 volumes of ethanol, 0.1 volumes NH_4Ac and resuspended in 250 µL TE.

2.2.7.4 Molecular weight DNA marker preparation

Radio-labelled markers were added to hybridisation mixes for RFLP analysis and loaded on to SSCP gels for orientation and size comparisons. Lambda phage DNA cut with Hind III/Eco RI and Hind III was labelled with α ^{32}P by nick translation (Kelly *et al* 1970) using the Amersham nick translation kit according to the suppliers specifications: 0.25 µg DNA, 3.3 µL of each of dCTP, dGTP, dTTP, 2.5 µL α ^{32}P dATP, 5 µL (2.5 units) DNA polymerase made up to 50 µL with ddH_2O. Reactions were incubated at 15°C for 1 hr, ethanol precipitated and resuspended in 250 µL TE then stored at -20°C until needed. A volume with approximately 200 cps at a distance of 2 cm from a radiation monitor was used in each hybridisation. pBR322 cut with Hae III was radiolabelled using T4 polynucleotide kinase: 1 µg pBR322 Hae III, 2.5 µL one-phor-all-buffer-plus, 2 µL (20units) T4 polynucleotide kinase, 2 µL γ ^{32}P dATP, and ddH_2O to 25 µL incubated at 37°C for 30 min. The mixture was then diluted 1 in 25 with ddH_2O and stored at -20°C until needed.

2.2.8 Single-stranded conformation polymorphism analysis (SSCP)

SSCP analysis was carried out according to Orita *et al* (1991) with modifications by Triggs Raine *et al* (1991). Specifically; radio-labelling of each C6 exon was done by incorporation of α ^{32}P dATP in the PCR amplification reaction. On completion of PCR reactions the mixes were diluted 1 in 50 with 0.1% (v/v) SDS, 10 mM EDTA. 10 µL of the diluted product was mixed with 10 µL of formamide loading dye (95% formamide, 20 mM EDTA, 0.25% w/v xylene cyanol, 0.25% w/v bromophenol blue), boiled for 2 min then chilled on ice before loading. Samples were run on a 6% polyacrylamide gel (33 cm x 42 cm x 0.35 mm) containing 10% (v/v) glycerol as well as on a second gel without glycerol. Electrophoresis of glycerol gels was at 40 watts for 4.5 hr in a 4°C room and non-glycerol gels were run at 30 watts for 2.5 hr also in a 4°C room. On completion of electrophoresis, gels were fixed in 10% (v/v) methanol, 10% (v/v) acetic acid for 10 min, transferred to 3MM Whatman paper and dried at 80°C under vacuum. Autoradiography was carried out from 4-24 hr at room temperature without intensifying screens. Films were developed as described previously.

2.2.9 Cloning of DNA fragments

2.2.9.1 Preparation of insert DNA fragments

PCR products were extracted with a phenol, chloroform mixture (50 % phenol) and then ethanol precipitated using 2 volumes of 100% ethanol and 0.1 volumes of 2 M NaOAc at -20°C overnight. The DNA was pelleted in a microcentrifuge for 15 min then resuspended in 13 µL ddH_2O. This product was phosphorylated using 9.2 units of polynucleotide kinase, 2 mM ATP and kinase buffer (50 mM Tris/HCl [pH 7.5], 20 mM $MgCl_2$, 2 mM DTT). The mix was incubated at 37°C for 45 min. Products were purified and concentrated to 5µL using microconcentrators (Amicon). 1 µL of this was run on a 2% (w/v) agarose gel to estimate concentration. Products were then appropriately diluted or used directly in ligation reactions.

2.2.9.2 Preparation of blunt ended vector

Two µg M13mp18 RF1 DNA was digested with 30 units of Sma I in a 50 µL volume containing one-phor-all-buffer plus and ddH_2O at 30°C for 1.5 hr. The DNA was dephosphorylated with 1 unit of calf intestinal alkaline phosphatase and CIP buffer (supplied) in a final volume of 100 µL. Incubation was carried out at 37°C for 30 min. The vector was then treated with proteinase K (100 ng/µL) in the presence of 0.25% (w/v) SDS and 5 mM EDTA at 56°C for 30 min. The DNA was next extracted with phenol/chloroform and ethanol precipitated (as described above). The DNA was pelleted in a microcentrifuge and resuspended in 20 µL ddH_2O, such that the final concentration was approximately 150-200 ng/µL. 1 µL of digested vector was run on a 1% (w/v) agarose gel along with a sample of undigested M13mp18 to check digestion and concentration.

2.2.9.3 Blunt end ligation of insert and vector

100 ng of linearised M13mp18 was incubated with approximately 30-50 ng of fragment DNA overnight at room temperature in a reaction mixture containing 50 mM Tris/HCl [pH 7.6], 10 mM $MgCl_2$, 5% (w/v) polyethylene glycol $(PEG)_{6000}$, 1 mM ATP, 1 mM DTT and 1 unit of T4 DNA ligase.

2.2.9.4 Preparation of E.coli TG1 competent cells

A single colony of E coli TG1 cells was inoculated into 10 mL Luria broth (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% NaCl) and shaken overnight at 37°C. The overnight culture was then diluted one in 100 into 200 mL of fresh sterile Luria broth and shaken at 37°C until the absorbance at 260 nm was 0.5. Cells were collected by centrifugation at 7000 rpm for 5 min in a pre-cooled centrifuge then resuspended in 40 mL CM solution1 (10 mM NaOAc, 50 mM $MnCl_2$, 5 mM NaCl). Cells were again centrifuged at 7000 rpm for 5 min then gently resuspended in 4 mL CM solution2 (10 mM NaOAc, 5% glycerol, 70 mM $CaCl_2$, 5 mM $MnCl_2$). The mixture was

incubated on ice for 20 min and then 100 µL aliquots were placed into cold tubes and stored at -70°C until needed.

2.2.9.5 Transformation and plating of competent cells

Competent cells were removed from -70°C and placed on ice to thaw slowly for 20 min. Ligation mixtures were then added to 100 µL of competent cells and incubated on ice for a further 45 min. Cells were heat shocked for 2.5 min at 42°C and then kept on ice. To each sample 200 µL of fresh exponentially-growing TG1 cells, 0.3 mM IPTG, 0.02% (w/v) X-gal and 3 mL of warmed H-top agar (1% w/v tryptone, 0.5% w/v yeast extract, 0.8% w/v NaCl, 0.8% w/v agar) were added. The mixtures were gently rolled between palms to mix, poured onto minimal media plates and grown overnight at 37°C

2.2.9.6 Screening clones and orienting fragments

Clear plaques were grown in 2 mL Luria broth with 20 µL of fresh exponentially-growing TG1 cells for approximately 7 hr in a shaking 37°C incubator. Cultures were then decanted into microfuge tubes and centrifuged for 5 min to form a pellet of cells. To check for the presence of inserts 20 µL of the supernatant and 5 µL DIGE dye (3% w/v SDS, 30% v/v glycerol, 0.25% w/v bromophenol blue) were incubated at 65°C for 10 min. Samples, along with control cultures, were run on a 0.7% (w/v) agarose gel at 60 V for 4-6 hr until the bromophenol blue had run off the end of the gel. Clones containing inserts run more slowly than those without. Positive cultures were then tested for orientation of the insert using the complement test (C-test) method. 10 µL of supernatant from two clones was mixed with 5 µL of C-test dye (3% w/v SDS, 30% v/v glycerol, 1.25 M NaCl, 0.25% (w/v) bromophenol blue) and incubated at 65°C for 50 min. Samples were then run on a 0.7% (w/v) agarose gel at 100 V for 1 hr. Clones with inserts in different orientations ran as two bands while inserts in the same orientation ran as one band.

2.2.10 Automated DNA sequencing

2.2.10.1 Preparatation of single-stranded DNA template

M13 single-stranded DNA was prepared from DIGE and C-tested clear plaques. 1.2 mL of plaque culture supernatant was removed and incubated with 300 µL PEG/NaCl solution (20% w/v PEG_{6000}, 2.5 M NaCl) on ice for 30 min. Tubes were then centrifuged for 15 min and all traces of PEG/NaCl solution were removed. The pellet was resuspended in 100 µL TE and extracted with equal volumes of phenol and chloroform-IAA. DNA was precipitated from the aqueous phase by addition of 2 volumes of 100% ethanol and 0.1 volumes of 3 M NaOAc (pH 5.3). The pellet was recovered after centrifugation and rinsed in 70% ethanol before being vacuum dried and resuspended in 8-15 µL ddH_2O.

2.2.10.2 Estimation of single-stranded DNA concentration

Yields of single-stranded template were estimated by running 1 µL of the DNA on a 1% (w/v) agarose gel along with 100 ng/µL of single-stranded M13mp8. Band intensities were compared and usually yields of single-stranded DNA were in the range 100-250 ng/µL.

2.2.10.3 Sequencing protocol

Approximately 600 ng of single-stranded DNA template was sequenced using an automated DNA sequencing system (ABI). Reactions for each A and C base involved PCR cycling 100 ng of template, 1 µL Taq buffer (supplied), 1 µL M13 dye primer Joe or Fam (0.4 pmol/µL), 1 µL d/ddATP or d/ddCTP and 1.25 units Taq polymerase. G and T bases required M13 Tamra and Rox dye labelled primers respectively and twice the volume of all reaction components. Tubes were then overlayed with approximately 10 µL of mineral oil and placed in a PCR cycling machine. Denaturation at 95°C for 3 min (cycle 1), and 30 sec (cycles 2-30), annealing at 55°C for 30 sec and extension at 70°C for 1 min for 30 cycles. After PCR cycling was

complete the four A,C,G,T reactions were pooled and precipitated using 2.5 volumes of 100% ethanol and 0.1 volumes 3 M sodium acetate (pH 5.3) at -20°C for several hr. DNA was pelleted in a microfuge for 30 mins, rinsed in 70% ethanol and vacuum dried. The pellet was resuspended in 5 µL deionised formamide, 1 µL 50 mM EDTA (pH 8.0) heated for 2 min at 90°C and electrophoresed on a 6% denaturing polyacrylamide gel, using an ABI 373 automated DNA sequencer. The ABI version 1.3 software was used to perform the automatic fluoresence analysis and base calling.

2.2.11 Statistical and computing methods

Restriction digest analysis of C6 exons was carried out using the software package PC/Gene (™, Intelligenetics Inc., U.S.A.). Primers were designed using the PCR primer selection software (Epicentre Software, Pasadena, U.S.A.). Calculation of protein pI values was carried out using the Isoelectric computer program (Genetics Computer Group Sequence Analysis Software package version 7.2). Statistical analyses were performed using the INSTAT computer program, P values of 0.05 or smaller were accepted as significant. Linkage disequilibrium estimates were calculated using three methods, these are described further in chapters 3 and 6. One method utilised log linear modelling techniques with the program GLIM 3.77 update 0 (copyright) 1985 Royal Statistical Society, London. A second method used Baur and Danilovs (1980) equation for linkage, and a specifically written computer program (X Gao, Human Genetics Group, John Curtin School of Medical Reseach) for the calculation, this method incorporated both possible haplotypes for double heterozygotes.

Chapter 3

Chapter 3

C6 Protein Typing and Taq I RFLP analysis

3.1 Introduction

Complement component six was first found to be polymorphic in 1975 (Hobart *et al*) using isoelectric focusing (IEF). Three common patterns are seen and designated A, B and AB, the A pattern being the more anodic. Additionally rare patterns occur which are characterised by strong bands at some fraction of the normal band interval. To date there are at least 19 variant allotypes (review Tokunaga *et al* 1990).

The inheritance of these patterns conforms to an autosomal codominant model (Hobart *et al* 1975). It was suggested by Alper *et al* (1975) that the A and B patterns differed by a single unit of charge causing a frameshift of the A pattern and this has since been proven (Fernie *et al* 1993, Chapter 4, Dewald *et al* 1993). Since the initial report of C6 IEF polymorphisms many population groups have been studied. Table 3.1 shows a summary of these findings. The main difference appears to be that the C6 A allotype is the more common form in Caucasians, while C6 B is more common in Oriental groups who are also characterised by a relatively high frequency of the "rare" C6*B2 allele. In all studies the gene frequencies conform to Hardy-Weinberg expectations, suggesting that there is little or no selective pressure on C6 alleles (Hobart *et al* 1975).

Several studies have shown that C6 and its functionally sequential component C7 are structurally and biochemically very similar (Podack *et al* 1979, DiScipio and Gagnon 1982), and that the genes for these proteins are closely linked (Hobart *et al* 1979, Tokunaga *et al* 1986), along with C9 on chromosome 5p13 (Jeremiah *et al* 1990). Many animal studies have shown that C6 is polymorphic and inherited as an autosomal codominant gene, in rabbits (Kunstman and Mauff 1980, Goldman *et al* 1982,

Table 3.1 Published C6 protein allele frequencies for various racial groups.

Population - Country(group)	C6 Protein Allele Frequencies A	B	Rare	Sample Number	Reference
Caucasian - England, USA	0.610	0.370	0.015	99	Hobart and Lachmann 1975
USA	0.646	0.333	0.021	264	Alper *et al* 1975
England, USA	0.621	0.360	0.018	189	Hobart amd Lachmann 1976
Norway	0.610	0.390	0.000	81	Olving *et al* 1977
West Germany	0.613	0.379	0.008	194	Rittner *et al* 1979
West Germany	0.601	0.388	0.011	709	Kunstmann *et al* 1980
Germany	0.646	0.346	0.008	254	Kuhnl and Kreckel 1980
Australia (IDDM)	0.590	0.404	0.006	162	Kirk *et al* 1980
Australia (NIDDM)	0.612	0.362	0.025	178	Kirk *et al* 1980
Australia	0.642	0.345	0.013	116	Kirk *et al* 1980
Norway (Lapps)	0.530	0.470	0.000	167	Olving *et al* 1980
Norway (Non-Lapps)	0.590	0.410	0.004	1623	Olving *et al* 1980
England	0.683	0.307	0.010	202	Whitehouse and Putt 1983
Italy	0.657	0.334	0.009	439	Scacchi *et al* 1992
Sardinia	0.706	0.294	0.000	240	Scacchi *et al* 1992
Russia	0.625	0.350	0.010	?	Kucher *et al* 1993

	A	B	Rare	Total Number
Combined Caucasian data	0.624	0.366	0.010	>4917
Range	0.530-0.706	0.294-0.470	0.000-0.025	
SD	(0.041)	(0.043)	(0.007)	

Table 3.1 continued.

Population - Country (area)	**C6 Protein Allele Frequencies** A	B	Rare	Sample Number	Reference
Oriental - N. America	0.590	0.350	0.050	51	Hobart and Lachmann 1975
Japan	0.427	0.483	0.090	288	Tokunaga *et al* 1983
Japan	0.423	0.510	0.067	495	Tokunaga *et al* 1984
Japan	0.432	0.503	0.065	565	Nakamura *et al* 1984
Japan	0.467	0.481	0.052	135	Nishimukai *et al* 1985
China (Beijing)	0.416	0.532	0.052	155	Zeng *et al* 1986
China (Guangzhou)	0.445	0.518	0.037	255	Zeng *et al* 1986
Japan (Ibaraki)	0.450	0.477	0.073	351	Washio *et al* 1986
Japan	0.446	0.466	0.088	278	Tokunaga *et al* 1986
Korea (Seol)	0.433	0.523	0.044	490	Park *et al* 1988
China (Fujiang)	0.463	0.500	0.037	?	Zhang and Du 1990
China (Sichuan)	0.498	0.448	0.055	?	Zhang and Du 1990
China (Heilongjiang)	0.471	0.522	0.007	?	Zhang and Du 1990
China (Zhengzho Han)	0.452	0.523	0.025	?	Hu and Du 1992
China (Lanzhou Han)	0.461	0.522	0.017	?	Hu and Du 1992
China (Huhhot Han)	0.445	0.529	0.026	?	Hu and Du 1992
China (Xi'an Han)	0.490	0.487	0.023	?	Hu and Du 1992
China (Guangdong)	0.457	0.515	0.028	?	Hu and Du 1992
China (Liaoning)	0.470	0.505	0.025	203	Tsunenari *et al* 1992

	A	B	Rare	Total number
Combined Oriental data	0.460	0.494	0.038	>3266
Range	0.416-0.590	0.350-0.532	0.007-0.090	
Standard deviation	0.038	0.042	0.023	

Table 3.1 continued.

Population - Country	**C6 Protein Allele Frequencies** A	B	Rare	Sample Number	Reference
Blacks -	0.560	0.380	0.060	59	Hobart and Lachmann 1975
N. America, W. Indes, Nigeria	0.551	0.403	0.045	165	Hobart and Lachmann 1976
Indians - India	0.500	0.500	0.000	13	Hobart and Lachmann 1976
Fiji	0.612	0.388	0.000	115	Ranford *et al* 1982
India	0.633	0.366	0.000	86	Corbo *et al* 1991
Micronesians - Nauru	0.446	0.452	0.102	186	Ranford *et al* 1982
	0.375	0.568	0.055	440	Ranford 1989
Polynesians - W.Samoa	0.629	0.359	0.006	245	Ranford *et al* 1982
Rarotonga	0.708	0.287	0.005	94	Ranford *et al* 1982
New Caledonia	0.600	0.396	0.004	120	Ranford *et al* 1982
	0.638	0.347	0.015	852	Ranford 1989
Melanesia - New Caledonia	0.650	0.336	0.014	70	Ranford *et al* 1982
Uvea	0.637	0.345	0.018	139	Ranford *et al* 1982
Fiji	0.693	0.307	0.000	186	Ranford *et al* 1982
	0.679	0.315	0.006	319	Ranford 1989
Australia - N. W. Aborigines	0.546	0.404	0.050	99	Ranford 1989
C. Aborigines	0.481	0.420	0.099	106	Ranford 1989
New Guinea - E. Highlands	0.578	0.402	0.020	97	Ranford 1989
New Britain	0.624	0.376	0.000	105	Ranford 1989
Rhesus Monkeys	0.592	0.354	0.005	103	Hall and Alper 1977

Rother 1986), rats (Granadose *et al* 1984), manx shearwater (Whitehouse 1982), pigs (Shibata *et al* 1993), and rhesus monkeys (Hall and Alper 1977). The close genetic linkage of C6 and C7 seen in man has also been found in marmoset (Whitehouse 1984), mouse (Orren *et al* 1985), dog (Eldridge *et al* 1983) and opposum (van-Oorschot *et al* 1993). This suggests that C6 and C7 are strongly conserved throughout species and probably of great antiquity.

Genetically controlled deficiency states have been identified for C6, firstly by Leddy and coworkers in 1974; by 1992 more than 100 C6 deficient individuals had been identified (review Wurzner *et al* 1992a). Work by Orren *et al* (1992) suggests that there may be two groups within the C6 deficiency category. In one group individuals have very low levels of C6; 0.3-3 µg/mL compared with the normal range of 20-80 µg/mL. The C6 protein in these samples is hemolytically active, able to incorporate into the terminal complement complex, has a molecular weight 79-86 percent of normal C6 and has only one weak anodal IEF band. Individuals with this C6 pattern are termed C6 subtotal deficients (C6SD). Samples in the second group show no detectable levels of C6 protein in their sera and are termed C6Q0 for quantitative zero.

At least 17 studies have suggested that deficiency of C6 (C6Q0) or C7 may lead to increased susceptibility to neisserial infection (partial review Hauptman 1989, Potter *et al* 1990, Densen 1991). In addition meningococcaemia has been reported in a case of acquired complement component deficiency (C3-C5, C8, C9 and factor B and I) due to hepatic failure (Ellison *et al* 1986). Interestingly Wurzner *et al* (1991) study of C6SD individuals showed them to have no history of susceptibity to neisserial infections. Therefore it appears that even very low levels of functionally active C6 and C7 may provide protection against these infections.

Reports of an association between toxoplasma and C6 deficiency (C6Q0) have also been described (Feldman and

Schreiber 1977, Delage *et al* 1979, Morgan *et al* 1989) as has an association between hyperthyroidism and C6 deficiency (Trapp *et al* 1987, Morgan *et al* 1989) and recurrent bronchopulmonary infections (Glass *et al* 1978).

C6 deficiency (C6Q0) has not been limited to associations with infective disease states but also with autoimmune disorders. Individual cases of C6 and C7 deficiency have been found in patients with systemic (Tedesco *et al* 1981, Zeitz *et al* 1981) or discoid lupus erythematosus (Trapp *et al* 1987), Sjogrens syndrome (Trapp *et al* 1987), rheumatic illness (Boyer *et al* 1975), arthritis with antinuclear antibodies (Reinitz *et al* 1986), haematuria (Sakano *et al* 1988) and other diseases. A significant association between the C6*B2 allele and glomerulonephritis has also been reported (Nishimuka *et al* 1985). To date two studies (Kirk and Ranford 1980, Yukiyama *et al* 1985) have examined C6 protein associations in autoimmune disease patients. Kirk and Ranford (1980) showed that 340 Caucasian diabetes patients had C6*A and C6*B frequencies similar to controls. Yukiyama and coworkers (1985) showed that Japanese SLE patients had reduced C6*A and C6*B2 frequencies compared with controls.

Complete and partial deficiencies of complement components C2 and C4 are associated with systemic lupus erythematosus (SLE) (Glass *et al* 1976, Schur 1978, Fielder *et al* 1983, Christiansen *et al* 1983, Dunckley *et al* 1987) with C4A deficiency being a genetic determinant of the disease in approximately one third of SLE patients (Dunckley *et al* 1987). It is of interest then to determine whether complete or partial deficiencies of C6 are similar genetic determinants of SLE in Caucasians, as C4A deficiency.

Additionally a C6 DNA polymorphism, a Taq I restriction fragment length polymorphism (RFLP) has been reported (Coto *et al* 1991a) but not extensively studied in either patient groups or healthy controls. Prior to widespread implementation of the polymerase chain reaction

(PCR), DNA analysis using RFLP was used to detect protein polymorphisms, often relying on linkage disequilibrium as well as direct detection of the polymorphism. Jazwinska *et al* (1988) showed that the serologically detectable Gm allotypes could be confidently defined by RFLP analysis. Similarly a study by Kohonen-Corish and Serjeantson (1986) showed that HLA-DR protein phenotypes correlated with the HLA-DRβ/Taq I RFLP. The present study examines whether Taq I RFLP analysis could replace traditional C6 protein typing.

In this study C6 protein analysis was carried out on 403 Caucasian SLE patients and 50 healthy Caucasian blood donors and laboratory staff using the conventional IEF gel technique (Hobart *et al* 1975) followed by passive transfer of protein to a membrane (Tokunaga *et al* 1984) and immunological detection (Whitehouse and Putt 1983). This combination of methods was chosen for its simplicity and availability of reagents. Taq I RFLP studies were carried out on 99 SLE patients and 46 healthy blood donors and laboratory staff. Numbers for RFLP analysis were much reduced in the SLE patient group due to the limited supply of large amounts of DNA, but numbers were considered adequate for the identification of an absolute correlation with protein data.

3.2 Aims

To examine complement component six protein and Taq I RFLP phenotypes and allele frequencies in a Caucasian SLE patient group and healthy white Australians.

3.3 Materials and Methods

3.3.1 SLE Patients

The SLE group examined consisted of patients who had been diagnosed with SLE according to the 1982 revised criteria for classification of SLE (Tan *et al* 1982). Patients came from Sydney where they had been reviewed by Dr Paul Gatenby (Clinical Immunologist, Royal Prince Alfred Hospital). Serum samples had been stored at -70°C for up to eight years. DNA was extracted for previous

studies between 1985 and 1989 using the standard phenol/chloroform method (Maniatis *et al* 1984) and stored at 4°C since then.

3.3.2 Controls

Serum samples were stored at -70°C. DNA was extracted for previous studies from healthy Canberran blood donors, laboratory staff and students and kept at 4°C.

3.3.3 Methods

Serum samples were C6 protein typed and DNA samples RFLP typed according to the methods described in Chapter 2. Some Taq I RFLP results were obtained by probing pre-existing nylon filters from previous studies by E Jazwinska and H Dunkley. This was done to conserve the limited supplies of DNA. Statistical analysis methods and linkage disequilibrium estimation methods are described and referenced in Chapter 2.

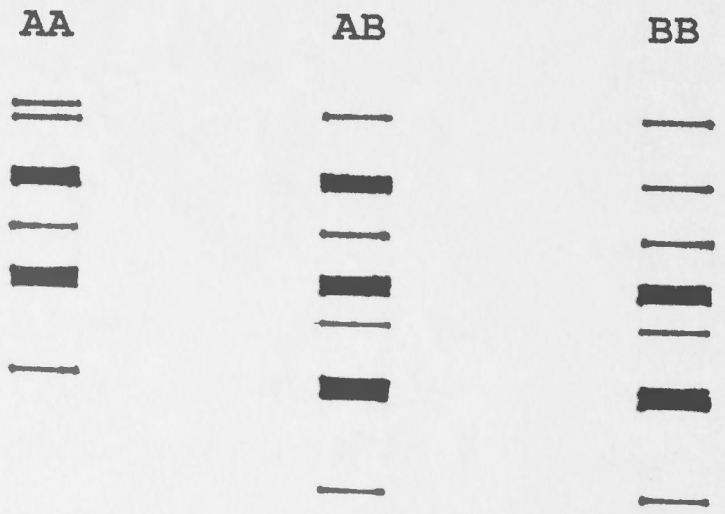
3.4 Results

3.4.1. Frequency of protein types

Figure 3.1 shows a C6 protein gel with the three common patterns AA, AB, BB and the heterozygous rare protein types found in this study; A1A, AB1, BB2, AB21 and BM91. C6 protein phenotypes and allele frequencies of SLE patients and controls are shown in Table 3.2. SLE patients and healthy controls showed very similar C6 protein allele frequencies (χ^2 = 0.002, P = 0.990). Both patient and control group C6 allele frequencies did not differ from expected Hardy-Weinberg equilibrium values (χ^2 = 0.160, P = 0.997, χ^2 = 0.005, P = 0.997 respectively). The SLE patients were European Caucasians as were the healthy control group and as such a comparison with pooled published Caucasian data was appropriate. No significant difference between the groups was found (SLE vs published mean; χ^2 = 0.769, P = 0.768, Healthy vs published mean; χ^2 = 0.755, P = 0.686).

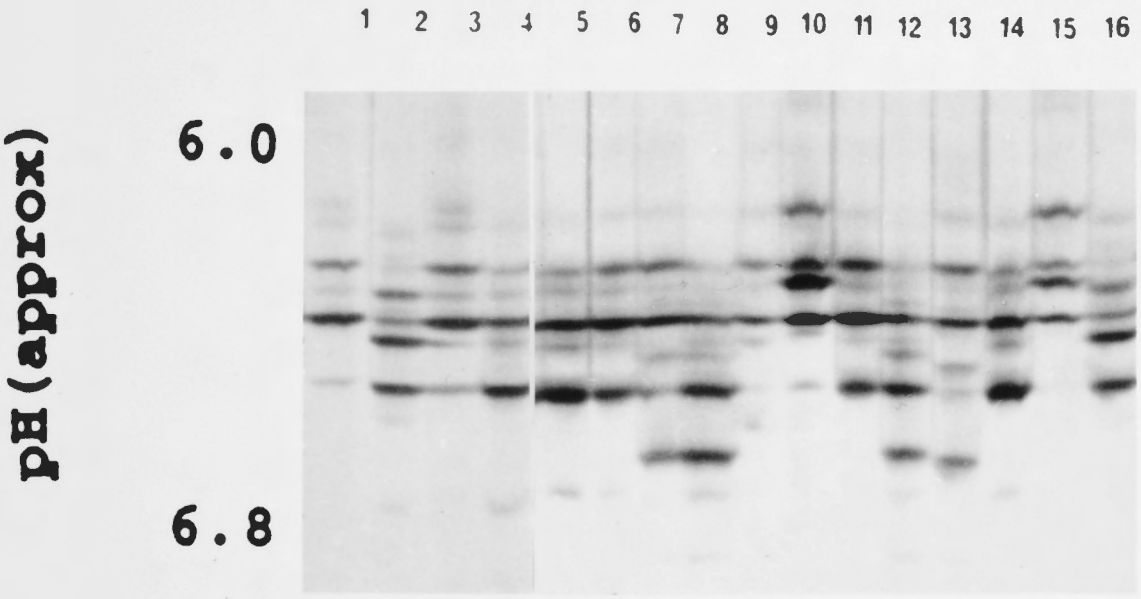
Figure 3.1 A C6 protein gel showing the three common and four rare protein types. Lane 1 AA, lane 2 and 16 BM91, lanes 3 and 11 AB, lanes 4, 5, 6 and 14 BB, lane 7 AB2, lanes 8 and 12 BB2, lane 9 AB1, lanes 10 and 15 A1A, lane 13 AB21.

Schematic Representation of C6 Protein Gel



C6 Isoelectric Focusing Gel.

Anode



Cathode

Table 3.2 The distribution of C6 phenotypes and C6 allele frequencies in an SLE patient group and a healthy Caucasian group.

Population	C6 Protein Phenotype										
	A		**AB**		**B**		**Arare**		**Brare**		
	Obs (%)	Exp	Obs (%)	Exp	Obs (%)	Exp	Obs (%)	Exp	Obs (%)	Exp	**Total**
SLE - Caucasian	135 (33.5)	132.3	189 (46.9)	194.4	74 (18.4)	71.4	3 (0.7)	2.8	2 (0.5)	2.0[a]	403
Healthy - Caucasian	15 (30.0)	15.7	25 (50.0)	24.1	9 (18.0)	9.2	1 (2.0)	0.6	0 (0.0)	0[b]	50

Population	C6 Allele Frequencies			
	C6*A	C6*B	C6*R	Number Tested
SLE - Caucasian	0.573	0.421	0.006	403
Healthy - Caucasian	0.560	0.430	0.010[d]	50
Pooled Published Caucasian data[c]	0.615	0.374	0.011[e,f]	4968

Obs = observed sample numbers, Exp = expected sample numbers

[a] $\chi^2 = 0.160$ df = 4 P = 0.997 [b] $\chi^2 = 0.005$ df = 3 P = 0.997

[c] See Table 3.1 for references

[d] C6 allele frequencies SLE vs Healthy $\chi^2 = 0.002$ df = 2 P = 0.990

[e] C6 allele frequencies SLE vs Published $\chi^2 = 0.769$ df = 2 P = 0.768

[f] C6 allele frequencies Healthy vs Published $\chi^2 = 0.755$ df = 2 P = 0.686

Four rare protein types in five patients were found and one in the control group. These variants were tentatively named using published patterns as a reference and later confirmed by Y Tokunaga, Blood Transfusion Service, University Hospital, Tokyo, Japan.

3.4.2 Frequency of RFLP patterns

RFLP phenotypes of SLE patients and controls as well as calculated allele frequencies are shown in Table 3.3. The two polymorphic bands detected with a 3' 1.28 kb fragment of C6 cDNA (R DiScipio, Research Institute of Scripps Clinic, La Jolla, California, USA) are an 8.0 kb and a 7.0 kb fragment (Figure 3.2). SLE and healthy control data show very similar RFLP allele distributions, (χ^2 = 0.002, P = 0.990). Insufficient quantities of DNA, or absence of pre-existing membranes, prevented RFLP analysis of three of the six rare-protein type samples. However the remaining three rare-protein type samples were RFLP typed. BM91 was an 8.0/7.0 kb RFLP heterozygote, BB2 was a 7.0 kb RFLP homozygote and AB1 was an 8.0 kb homozygote. RFLP patterns for these individuals were identical to patterns found in samples with common-protein phenotypes. RFLP allele frequencies of SLE, healthy Caucasians and published Spanish data are similar (SLE vs Spanish; χ^2 = 0.412 P = 0.814, healthy vs Spanish; χ^2 = 0.212 P = 0.899) and do not differ from expected Hardy-Weinberg equilibrium values (SLE; χ^2 = 0.009, P = 0.955, Healthy; χ^2 = 0.186, P = 0.911)

3.4.3 Association of protein and RFLP types

No absolute correlation between protein and RFLP types was found. Analysis of this data involves some degree of uncertainty since haplotypes of double heterozygotes cannot be definitively identified. Three methods of examining the data are available; 1) exclusion of double heterozygotes, 2) double heterozygotes counted twice, ie both possible haplotypes included and 3) double heterozygotes distributed among the possible haplotypes based on the allele frequencies of the two sites. Tables 3.4 (a-c) show the haplotype frequencies for the three

Figure 3.2 C6 1.28 kb cDNA (1689-2971 bp) hybridisation of Taq I digested genomic DNA. Lane 1 homozygote 7.0 kb, lane 2 heterozygote 7.0 and 8.0 kb, and lane 3 homozygote 8.0 kb.

C6 - 1.28kb cDNA (1689-2971bp) x TaqI

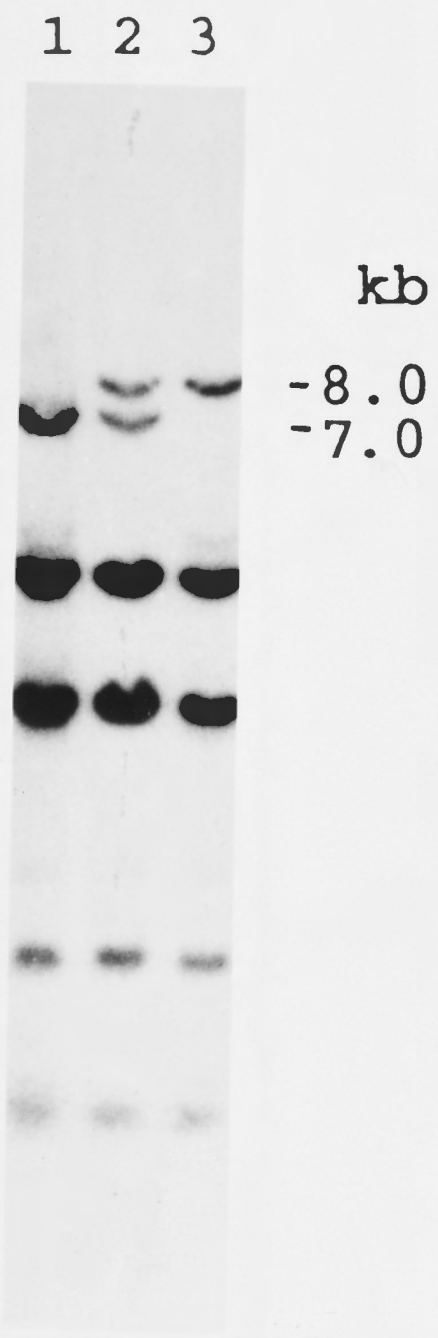


Table 3.3 C6 Taq I RFLP phenotypes and allele frequencies in an SLE patient group and healthy Caucasian group

Population	**C6 Taq I RFLP Type** (kb)						
	8.0		8.0/7.0		7.0		
	Obs (%)	Exp	Obs (%)	Exp	Obs (%)	Exp	**Total**
SLE - Caucasian	15 (15.2)	13.8	44 (44.4)	46.3	40 (40.4)	38.8[a]	99
Healthy - Caucasian	7 (15.2)	7.8	24 (52.2)	22.3	15 (32.6)	15.9[b]	46

Population	**C6 Taq I RFLP Allele Frequencies**		
	C6*8.0	C6*7.0	**Total**
SLE - Caucasian	0.374	0.626	99
Healthy - Caucasian	0.413	0.587[c]	46
Spanish data[f]	0.400	0.600[d,e]	62

Obs = observed sample numbers, Exp = expected sample numbers

[a] SLE, testing Hardy-Weinberg equilibrium $\chi^2 = 0.009$ df = 2 P = 0.955

[b] Healthy, testing Hardy-Weinberg equilibrium $\chi^2 = 0.186$ df = 2 P = 0.911

[c] C6 RFLP allele frequencies SLE vs Healthy $\chi^2 = 0.688$ df = 2 P = 0.710

[d] C6 RFLP allele frequencies SLE vs Spanish $\chi^2 = 0.412$ df = 2 P = 0.814

[e] C6 RFLP allele frequencies Healthy vs Spanish $\chi^2 = 0.212$ df = 2 P = 0.899

[f] Coto *et al* (1991a)

Table 3.4a C6 protein and C6 Taq I RFLP data pooled from the SLE patient group and healthy Caucasians.

Method 1; Haplotypes of combined SLE and healthy Caucasians protein and Taq I RFLP data excluding double heterozygotes and rare-protein alleles.

Genotypes Observed (N)	Haplotype Assignment A-8.0	B-8.0	A-7.0	B-7.0
AA:				
8.0/8.0 (15)	30			
8.0/7.0 (19)	19		19	
7.0/7.0 (11)			22	
AB:				
8.0/8.0 (4)	4	4		
8.0/7.0 (43)	_*	_*	*	_*
7.0/7.0 (22)			22	22
BB:				
8.0/8.0 (2)		4		
8.0/7.0 (4)		4		4
7.0/7.0 (20)				40
Total (140)	53	12	63	66
Haplotype frequencies	0.273	0.062	0.325	0.340

* haplotypes not discernable

Linkage disequilibrium; D^a = (freq of A-8.0)(freq of B-7.0) - (freq of A-7.0)(freq of B-8.0)

$D^a = 0.075$

D^b = Baur and Danilovs method (1980)

C6*A and C6*8.0 $D^b = 0.077$, $\chi^2 = 21.73$, $P < 0.001$

C6*B and C6*7.0 $D^b = 0.063$, $\chi^2 = 14.08$, $P = 0.035$

Table 3.4b C6 protein and C6 Taq I RFLP data pooled from the SLE patient group and healthy Caucasians.

Method 2; Haplotypes of combined SLE and healthy Caucasians protein and Taq I RFLP data including double heterozygotes, counted twice for both possible haplotypes and excluding rare-protein alleles

	Haplotype Assignment			
Genotypes Observed (N)	**A-8.0**	**B-8.0**	**A-7.0**	**B-7.0**
AA:				
8.0/8.0 (15)	30			
8.0/7.0 (19)	19		19	
7.0/7.0 (11)			22	
AB:				
8.0/8.0 (4)	4	4		
8.0/7.0 (43)	43*	43*	43*	43*
7.0/7.0 (22)			22	22
BB:				
8.0/8.0 (2)		4		
8.0/7.0 (4)		4		4
7.0/7.0 (20)				40
Total (140)	96	55	106	109
Haplotype frequency	0.262	0.150	0.290	0.298

* haplotypes not discernible

Linkage disequilibrium; $D^a = 0.0346$

C6*A and C6*8.0 $D^b = 0.119$ $\chi^2 = 64.41$, $P < 0.001$

C6*B and C6*7.0 $D^b = 0.101$, $\chi^2 = 47.00$, $P < 0.001$

Table 3.4c C6 protein and C6 Taq I RFLP data pooled from the SLE patient group and healthy Caucasians.

Method 3; Predicted haplotypes of combined SLE and healthy Caucasians protein and Taq I RFLP data using gene counting for all but the double heterozygotes, with these distributed according to allele frequencies for the two sites, rare-protein alleles excluded.

	Haplotype Assignment			
Genotypes Observed (N)	**A-8.0**	**B-8.0**	**A-7.0**	**B-7.0**
AA:				
8.0/8.0 (15)	30			
8.0/7.0 (19)	19		19	
7.0/7.0 (11)			22	
AB:				
8.0/8.0 (4)	4	4		
8.0/7.0 (43)	18.43*	30.85*	13.54*	22.66*
7.0/7.0 (22)			22	22
BB:				
8.0/8.0 (2)		4		
8.0/7.0 (4)		4		4
7.0/7.0 (20)				40
Total (140)	71	43	77	89
Haplotype frequencies	0.254	0.154	0.275	0.318

* haplotypes not discernible

Linkage disequilibrium using the conventional equation;

$D^a = 0.038$

methods described. Analysis of these results (Table 3.5) shows there to be no significant difference between them; (1 vs 2; χ^2 = 4.38, df = 3, P = 0.223, 1 vs 3; χ^2 = 4.40, df = 3, P = 0.230, 2 vs 3; χ^2 = 0.102, df = 3, P = 0.992).

Numerous methods for estimation of linkage disequilibrium (D) between two loci are available, the simplest being;

D^a = (freq A-8.0)(freq B-7.0) - (freq A-7.0)(freq B-8.0).

Results of this calculation for the three different haplotype distribution methods are given in Tables 3.4 (a-c). This linkage disequilibrium calculation for method 1) (discarding double heterozygotes) gives a value of D^a = 0.075, suggesting linkage disequilibrium between these sites. The alternative methods 2) and 3) give D^a = 0.035, and 0.038 respectively, concluding a random association. A second linkage disequilibrium estimation method (Baur and Danilovs 1980) is considered better as it attempts to normalise the equation for allele frequencies which the equation given above does not. This calculation however could only be carried out for the exclusion of double heterozygotes and double-counting methods because the computer program used to calculate this linkage disequilibrium value was designed for the double-counting method of haplotype analysis. The linkage values estimated for both of the haplotype methods showed a significant association between the protein A and RFLP 8.0 alleles as well as between the protein B and RFLP 7.0 alleles (Tables 3.4 a and b).

A third and further refined method of measuring an association between two loci is a log-linear model (Weir and Wilson 1986). This method discards double heterozygotes where the haplotypes can not be determined, since the authors believe their inclusion introduces ambiguities in the data. The result obtained from this analysis gives a scaled deviance of 35.713 , df = 4, P < 0.001, showing that there is significant deviation from random association of the alleles. The conclusion reached

Table 3.5 Comparison of haplotype distribution methods for analysis of double heterozygotes for the C6 protein determining loci and the C6 Taq I RFLP site.

Method	% Haplotypes A-8.0	B-8.0	A-7.0	A-7.0	Comparisons	χ^2	df	p
1. Discard double heterozygotes	27.3	6.2	32.5	34.0	1 vs 2	4.38	3	0.223
					1 vs 3	4.40	3	0.230
2. Double counting	26.2	15.0	29.0	29.8	2 vs 3	0.102	3	0.992
3. Distribute using	25.4	15.4	27.5	31.8				

from these linkage disequlilibrium analyses suggests there is linkage disequilibrium between the C6 protein and Taq I RFLP alleles.

3.5. Discussion

Results in this study contrast with Yukiyamas results in that there is no difference in distribution of C6 protein alleles between a large Caucasian SLE group, control group and published Caucasian data. The difference in these two studies probably arose from differences in sample size and diversity within SLE patients. Additionally there was no evidence that any patients or controls were C6 deficient. C6 protein allele frequencies for both SLE patients and controls are in good agreement with Hardy-Weinberg equilibrium thus indicating that all genotypes were distinguished and in particular that null alleles and heterozygous null samples were not missed. The C6 monoclonal antibody used in the C6 protein typing reacts with the end portion of the factor-I-like module II (FIM II) of C6 (Wurzner *et al* 1993) which is encoded by exon 15 (Hobart *et al* 1993a). While this analysis did not determine whether the protein in these samples was functional, no bands outside of the normal migratory range were noted and no samples showed any loss of constant bands. Thus there was no evidence of a truncated or aberrantly spliced protein in any of the samples examined. In addition little variability in band intensity was found across the different samples and there was no evidence that samples had degraded. So it appears that C6 is a relatively stable protein and is unaffected by appropriate long term storage.

Taq I RFLP studies confirmed the protein results in that no association between C6 and SLE was evident. Controls and patients had identical phenotypes and similar allele frequencies. Comparisons with published data continued the trend of similarity. Of the 145 samples RFLP typed none showed unusual sized bands or were missing bands. Band intensities varied only as was expected with varying amounts of DNA; band intensities were consistent within any one sample. Some partial digests were noted but were easily distinguished. From the results it appears that no major gene rearrangements have occurred in

the samples examined. RFLP analysis is a relatively gross level examination and from what is known of the conservation of C6 throughout species it was unlikely that a major gene rearrangement would be found.

Analysis of protein and Taq I RFLP data showed there to be no absolute correlation between the two techniques, and as such Taq I RFLP studies could not replace the isoelectric focusing protein technique. Further studies (Chapter 4, Fernie *et al* 1993, Dewald *et al* 1993) have since shown that the A and B protein alleles differ by a single unit of charge resulting from a point substitution at amino acid 98 (numbering from DiScipio and Hugli 1989). This polymorphism does not affect a Taq I restriction enzyme site in genomic DNA and thus is not responsible for the polymorphic RFLP pattern. The amino acid 98 polymorphism causes a change from glutamic acid (negatively charged) to alanine (neutrally charged) residue and thus determines whether the protein will have an A or B pattern on IEF gels. The polymorphism affects a Dde I restriction enzyme site and had this enzyme been used in RFLP analysis it might have resulted in an informative RFLP with perfect correlation to protein phenotype.

Three of the six rare-protein typed individuals were Taq I RFLP typed. These samples showed no differences in RFLP phenotype compared with other individuals. On the basis of the C6*A and C6*B protein allelism it is most likely that these rare alleles will be determined by point mutations in the DNA. Larger DNA changes such as several base pair insertions or deletions could result in frameshift changes consequently altering protein translation. Fernie *et al* (in preparation) suggest that the polymorphic Taq I site maps to the intron between exons 16 and 17, approximately 15 kb from exon 12; however they do not rule out that the site lies beyond the 3' untranslated region. It thus appears that the Taq I polymorphism would not affect protein phenotype since it probably occurs in a non-coding region of the gene.

In order to determine whether there is an association between the protein and Taq I RFLP alleles, haplotypes of the individuals must be determined. This is a simple process when samples are homozygous at either of the two sites but impossible when both sites are heterozygous and no family data are available. Many researchers advocate discarding double heterozygotes in this situation. This is considered valid although somewhat inefficient particularly as double heterozygotes in this study constitute the largest single class of genotypes in the data. Alternative procedures include these data either counting each double heterozygote twice, incorporating both possible haplotypes, or by allocating haplotypes based on allele frequencies for the two loci. Any method however, involves a degree of inaccuracy and a universal consensus as to which method is least affected by inacuracies has not yet been reached. Statistical analysis of the resultant haplotype frequencies in the present study for these three methods, suggests there is no significant difference between them. For this data set it therefore appears that any of the methods will represent the data to a similar degree of accuracy.

A further area of debate is which method should be employed to calculate a possible association between two loci. The simplest method, found in basic genetics texts, results in a conclusion of random association of these alleles for haplotype methods 2) and 3) but not when method 1) is utilised. This linkage disequilibrium calculation is considered overly simplified and very dependent on allele frequencies (Lewontin 1988), thus more refined statistical models have been developed. The further two calculations employed (Baur and Danilovs 1980, Weir and Wilson 1986) are considered better estimators of linkage disequilibrium as they attempt to normalise the equation such that it is less affected by allele frequencies. The conclusion reached from these analyses is that the protein and RFLP alleles are not associated at random. The validity of measuring an association between

two or more alleles is an issue not universally agreed upon and will be discussed further in Chapter 6.

The occurrence of linkage disequilibrium in this gene is not an expected outcome. The C6 gene is not thought to be greatly selected for since people deficient in the C6 protein are often perfectly healthy. Additionally neither the protein nor RFLP polymorphisms are thought to alter the functional efficiency of the C6 protein and any specific combination of alleles for the two sites is not known to incur any advantage or disadvantage to the individual. The two polymorphic sites being examined are at least 60 kb apart (Hobart *et al* 1993a) and therefore recombination events are possible in this region. As yet there have been no studies examining the Taq I polymorphism in other organisms and so it is impossible to determine if this polymorphism is as old as the C6 protein allelism, which has been described in other animals and is quite ancient. It is possible that the Taq I RFLP site is a recent polymorphism and insufficient time has passed for recombination to produce linkage equilibrium.

Further examination of the C6 gene (Chapter 5) shows that the common C6 protein allele determining site and one further DNA polymorphism occur in chimpanzees. This result implies that these polymorphisms arose prior to divergence of chimpanzees and humans and have since been maintained in the two populations. For this maintenance of DNA sequence to occur it is usually assumed that some form of selection has been involved (Takhata and Nei 1989, Golding 1992). It is possible that the selective advantage of these sites, as well as the apparently associated common protein allele determining loci and RFLP site, is no longer present or that it has not yet been identified. More extensive studies examining the relationship between structure and function of C6 may elucidate possible advantages of particular DNA sequences.

As noted earlier several studies reported individual cases of C6 and C7 deficiency with systemic (Tedesco *et al* 1981, Zeitz *et al* 1981) and discoid lupus erythematosus (Trapp *et al* 1987). The suggested mechanism by which complement contributed to development of SLE was in its affect on solubilisation of immune complexes. SLE is well associated with a reduced ability to solubilise immune complexes and it was thought that defects in the complement system might contribute to this. Zeitz *et al* (1981) study suggested that deficiencies of the terminal complement components and the resultant absence of functional membrane attack complex (MAC) did not prevent solubilisation of immune complexes. This evidence in addition to the present study appears to confirm that absence of MAC function caused, for example by C6 deficiency, is not a characteristic of SLE patients and that other mechanisms must be involved in solubilisation of immune complexes if that is truly a contributing factor to SLE development. It is likely, based on the heterogenous nature of SLE, that a number of factors both genetic and environmental contribute more to the disease process than C6.

Considerable evidence suggests that C6 deficiency (C6Q0) leads to an increased susceptibility to neisserial infection. Absence of C6 prevents the formation of the membrane attack complex (MAC) which consequently reduces bactericidal and cytolytic activity in the sera of these individuals (Densen 1991, Lim *et al* 1976). This lytic mechanism appears to be crucial in the immune system's ability to control neisserial infection. Again however there appears to be large heterogeneity in infective susceptibility as many C6 and C7 deficient individuals remain healthy (Lachmann *et al* 1978).

3.6 Conclusions

(1.) This study shows that neither C6 protein deficiency nor a specific C6 protein phenotype is associated with SLE. Additionally there is no evidence

that an abnormal C6 protein occurs in any of the SLE patients or controls.

(2.) No association between SLE and the C6 polymorphic Taq I site is present since SLE patients, controls and published data show similar allele frequencies. Nor is there evidence for any major gene rearrangements or deletions in the C6 gene of any SLE patients or controls.

(3.) No absolute correlation between C6 protein type and Taq I RFLP pattern is present. Therefore Taq I RFLP analysis cannot replace the C6 IEF protein typing protocol.

Chapter 4

Chapter 4

Mutation Detection In the C6 Gene using Single-Stranded Conformation Polymorphism Analysis

4.1 Introduction

The field of mutation detection is becoming increasingly important particularly as more disease-causing genes and their mutations are identified. Restriction fragment length polymorphism (RFLP) analysis and linkage studies have been widely used over the years to identify possible associations between a disease and particular polymorphisms. One early example of linkage between a disease and RFLP polymorphism was for sickle cell anemia where the mutation was linked to a variable Hpa I site (Kan and Doxy 1978). RFLP analysis is very useful in detecting major gene rearrangements and smaller mutations which alter a restriction enzyme recognition site. However many point mutations do not affect an enzyme site and thus are missed with this method. Neither the common C6 protein types (AA, AB and BB) nor the three rare-protein typed individuals (AB1, BB2 and BM91) found in this study (Chapter 3) showed any evidence for further polymorphisms in the Taq I RFLP study. This demonstrates that the changes causing these protein phenotypes do not involve a Taq I restriction enzyme recognition site.

It is of interest to identify and characterise the DNA site which determines the C6 protein allelism because it could enable protein typing by isoelectric focusing to be replaced by a more robust, faster and simpler DNA test. Isoelectric focusing is often inconvenient because samples must be carefully stored to avoid degradation of the labile proteins. The highly polymorphic nature of complement component six makes it a very useful marker in population studies, particularly for comparison of Oriental and Caucasian population groups. A simpler test that does not require careful storage of serum samples would greatly increase the usefulness of C6 as a tool for population studies.

A number of relatively new methods of mutation detection are now available and most are aimed at detecting single base-pair changes with very high efficiency. Several factors must be considered before adopting a technique for a particular study. The length of the coding region of a gene is particularly important especially if the use of mRNA-cDNA is not an option. The number of samples to be screened and the probability of heterozygosity of mutations must be considered. Cost, ease of procedure, safety and time are always important factors for almost all laboratories.

DNA from the rare-protein typed control was not available and only limited supplies of DNA from the SLE patients with protein variant phenotypes were available with no possibilty for further samples from three people as these patients were deceased or uncontactable. It was therefore not an option to use mRNA-cDNA and the study had to be optimised so as not to exhaust limited DNA samples. A screening procedure to identify heterozygous single-base changes in individual exons PCR-amplified from genomic DNA was necessary. Based on the number of exons and samples, the heterozygous nature of the prospective mutations, time and expense, sequencing every exon in each rare-protein typed individual, was not considered.

Single-stranded conformation polymorphism (SSCP) analysis is a mutation detection screening method which examines relatively short stretches of DNA, ideal for individual exon analysis. It first requires the target DNA to be amplified using PCR, utilising either incorporation of radiolabel or labelled primers in the PCR reaction. The PCR product is then diluted in a denaturing solution and heat dissociated before being run on a non-denaturing polyacrylamide gel (Orita *et al* 1989). Under these conditions single-stranded DNA fragments fold into specific conformations based on their primary sequence because these structures are stabilised by intramolecular interactions. Single-base substitutions in a PCR product

can induce a conformational change in the single-stranded DNA that is detected as an altered migratory pattern in a polyacrylamide gel. This technique is simple, inexpensive, relatively safe and fast. Its reported sensitivity is between 35 (Sarkar *et al* 1992) and 100 percent (Orita *et al* 1989, 1990, Michaud *et al* 1992) depending on how vigorously parameters such as gel temperature, ionic strength of the running buffer and glycerol content are optimised for each study (Glavac and Dean 1993).

Alternative screening methods commonly employed include denaturing gradient gel electrophoresis (DGGE), mismatch cleavage and heteroduplex analysis. DGGE relies on the differing melting behaviour of various sequences under denaturing conditions. This method appears to have good sensitivity, ranging from approximately 50 (Sheffield *et al* 1989) to 100 percent (Gottardi *et al* 1992) particularly when a GC clamp is utilised (Sheffield *et al* 1989). The method is simple, safe and fast but requires specialised apparatus and expensive GC extended primers.

Mismatch cleavage relies on identifying sequence mismatches in heteroduplexed molecules by chemical cleavage or enzymatic attack of the single-strand at the site of the mismatch. RNase A methods are less sensitive for particular mismatches (Myers *et al* 1985), and while chemical cleavage methods can detect a high percentage of genetic variation, they require expertise to carry out the biochemical steps involved (Cotton *et al* 1988), and great care must be taken with the toxic chemicals.

Heteroduplex analysis relies on the differing electophoretic mobility of perfectly matched DNA fragments compared with similar fragments containing a mismatch, or heteroduplex. It is a simple, fast, safe and relatively inexpensive method of screening mutations, but as yet only limited published data are available to establish its sensitivity.

Originally the strategy for identifying mutations which cause IEF protein variants was based on the idea that IEF only detects charge changes in the protein being examined. Thus only DNA mutations causing amino acid changes which result in charge changes would be candidates. This simplifies identification should several polymorphisms be found. However on further investigation it appears that neutral amino acid substitutions may be detected with careful application of IEF conditions (Whitney *et al* 1979) as they can change the conformation or folding of the protein and hence alter its migratory pattern in a polyacrylamide gel. From this it appears that these protein variants may not be exclusively charge variants and if several polymorphisms were found it may not be possible to definitively identify which of the mutations contributed to the variant pattern seen on an IEF gel. DNA analysis of a number of samples with the same protein variant could however resolve this problem.

Exon-specific PCR products from the C6 gene of five rare-protein typed SLE samples and several different ethnic population samples were examined using SSCP analysis. Any sample showing an altered SSCP pattern compared with the majority of others was investigated by sequencing cloned products. Ten clones from the variant samples were compared with the published sequence and wherever the same sequence change occurred in more than one clone a computer-based restriction site analysis was carried out, usually identifying an altered enzyme recognition sequence. Sequence changes were then either confirmed or disproven after the affected PCR product was digested with the appropriate enzyme and examined, along side controls, on a polyacryalmide gel.

4.2 Aims

To locate the DNA site which determines the common C6 protein alleles A and B and to examine all of the C6 coding exons in samples with a known rare-protein type to determine the DNA sequence changes that cause these variant protein phenotypes.

4.3 Materials and Methods

The methods of analysis used were restriction enzyme digestion, PCR-RFLP, single-stranded conformation polymorphism analysis, blunt end cloning and automated sequencing. Details are outlined in Chapter 2. Samples examined were (a) the five SLE patients found to have a rare-protein type, all of whom were heterozygous for the rare allele, (b) a homozygous WP3 (B2) sample found in a previous study (Ranford 1989), (c) one AA, AB and BB protein typed sample from the Caucasian SLE group and (d) one AA, AB and BB protein typed sample from each of four different ethnic groups; Aboriginal Australian, Caucasian, Micronesian and Polynesian groups.

4.4 Results

4.4.1 Common protein allele determination site

The two published C6 cDNA sequences (DiScipio and Hugli 1989, Haefliger *et al* 1989) differed by a single base change at amino acid 98, cDNA position 413 (numbering according to DiScipio and Hugli 1989), of exon three (Hobart *et al* 1993a). This change affected a Dde I restriction enzyme site which thus allowed PCR-RFLP to be employed to examine the site. The resultant digest patterns correlated perfectly with the common protein alleles. The C6*A had an additional Dde I site compared with C6*B.

Figure 4.1 shows the results of a Dde I digest and description of the DNA sequence, amino acid change and predicted pI values for the two protein alleles. This mutation causes an amino acid change giving rise to a charge change in the protein. This correlates well with what is seen on an IEF gel because the A allele has a predicted pI value of 6.64, which is slightly more acidic and therefore will migrate to a more anodic position than the B allele, which has a predicted pI value of 6.71.

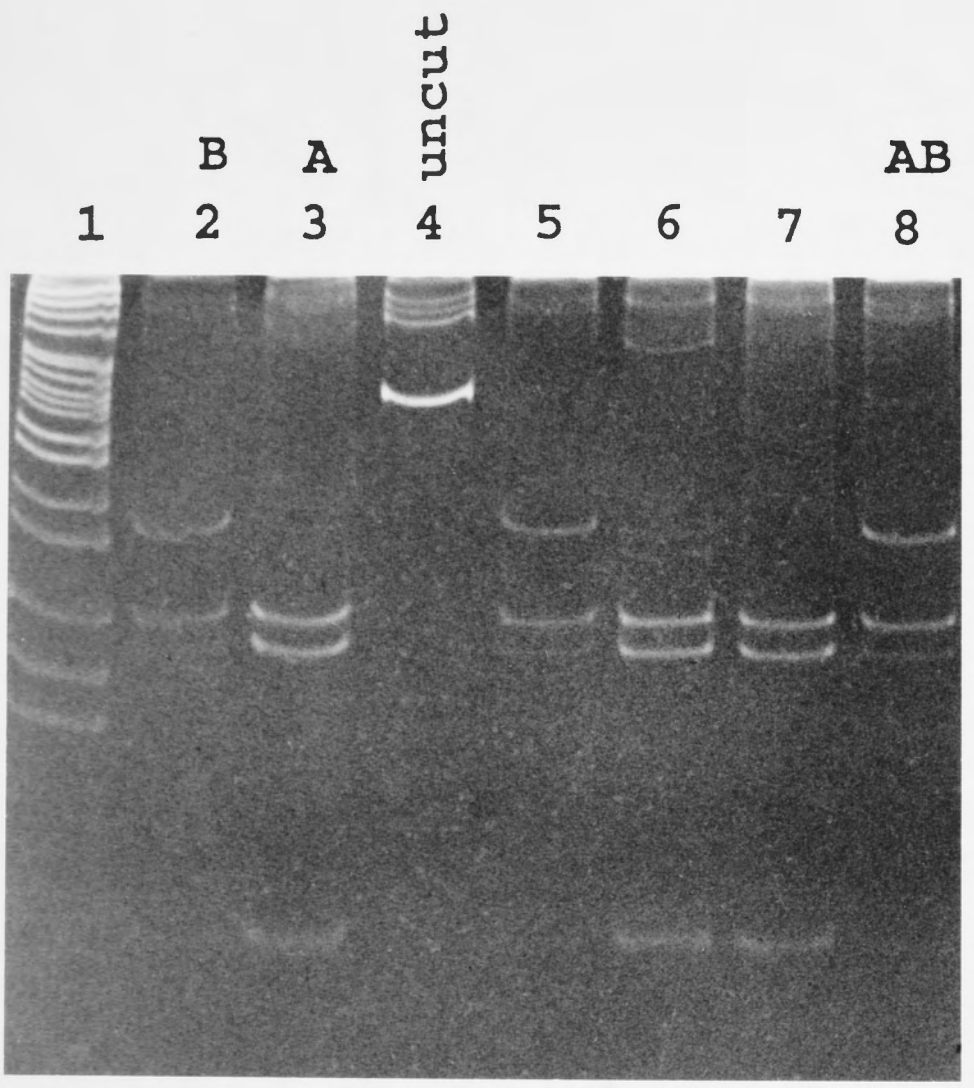
4.4.2 Variant SSCP patterns

Figure 4.2 shows the SSCP patterns found for all of the C6 exons. Table 4.1 summarises the variant SSCP

Figure 4.1 Amplified restriction fragment length polymorphism typing of the common C6 protein phenotypes utilising the exon 3, Dde I site **a** Polyacrylamide gel after electrophoresis of uncut 199 bp amplification product (lane 4) and Dde I fragments from C6 AA homozygote (lanes 3,6, 7), AB heterozygote (lane 8) and BB homozygote (lanes 2 and 5). The fragment sizes, in bp of a DNA standard (Msp I digested pBR322) are shown in lane 1. **b** Position, sequence, amino acid, restriction site and pI values for the two common C6 protein types.

C6 Exon 3 Dde I Digest

a



b

Polymorphism at cDNA 413, amino acid 98, exon 3.056

Phenotype;	A	B
Codon;	GAG	GCG
Amino acid;	Glu	Ala
Dde I site;	+ (C'TNAG)	- (C'TNAG)
pI;	6.64	6.71

Figure 4.2 SSCP patterns for all of the C6 exons examined. Non-glycerol and glycerol gels shown; variant bands are those which differ from the pattern seen in the majority of other samples. In exon 11 two equally frequent patterns were seen, these were labelled "2" for the two band pattern seen and "3" for the three band pattern since it was inappropriate to label either as "variant".

SSCP Gels

Exon 1.

Non-Glycerol variant

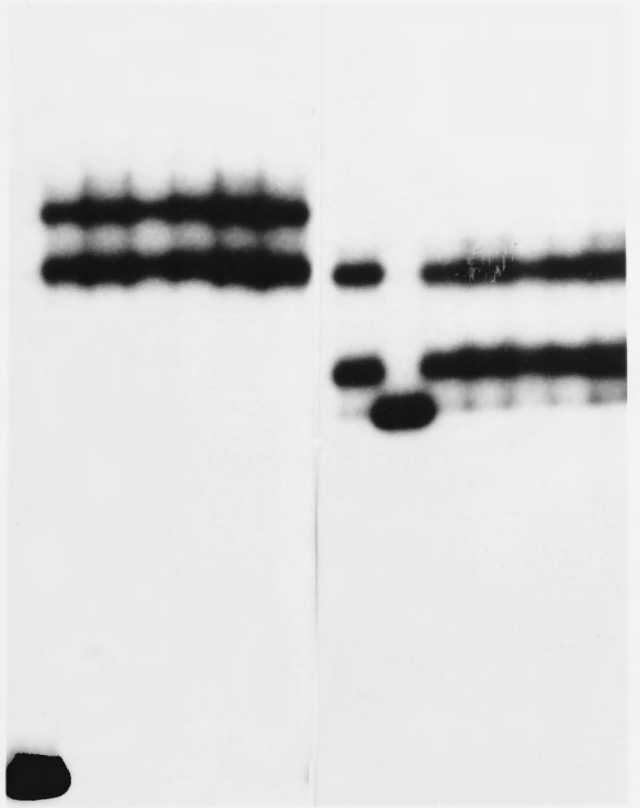
10% Glycerol variant

Non denatured

Non denatured

Exon 2.

Non-Glycerol 10% Glycerol

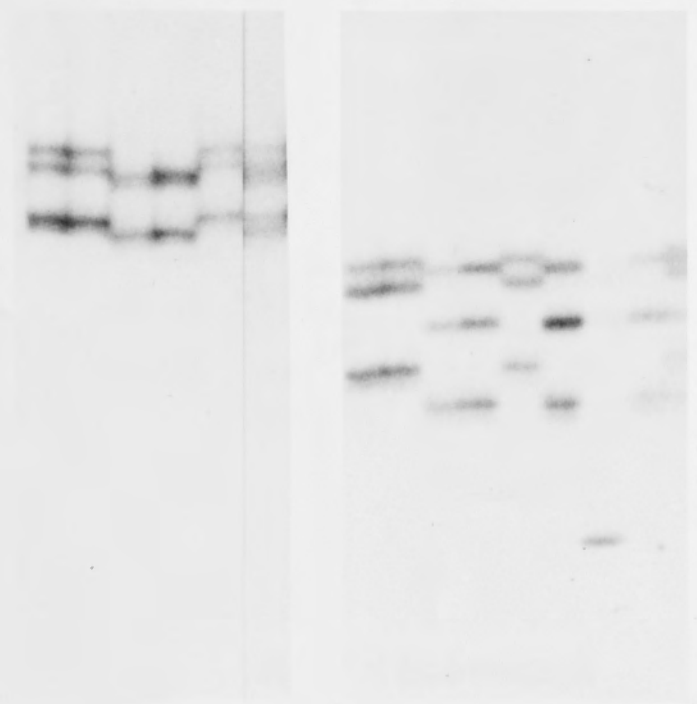


Non denatured

Exon 3.

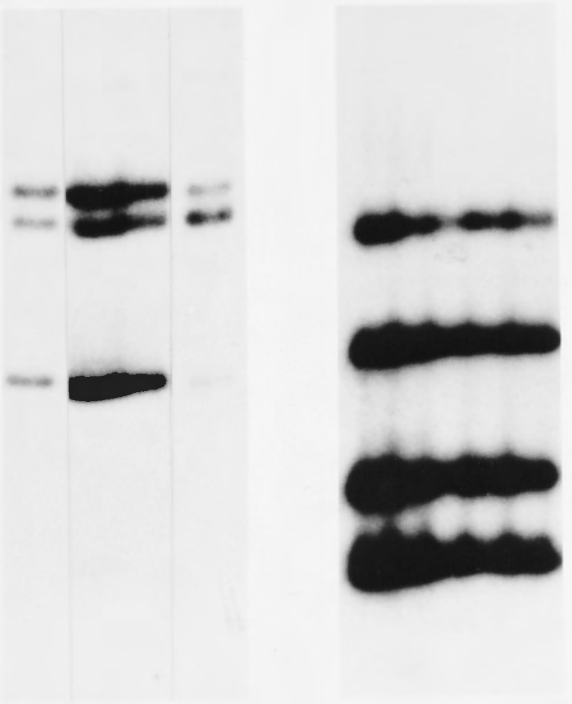
Non-Glycerol 10% Glycerol

A B AB A B AB



Exon 4.

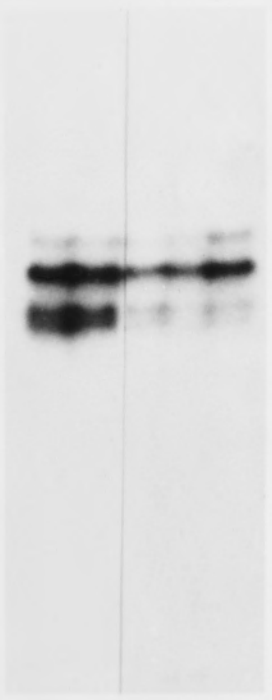
Non-Glycerol 10% Glycerol



SSCP Gels

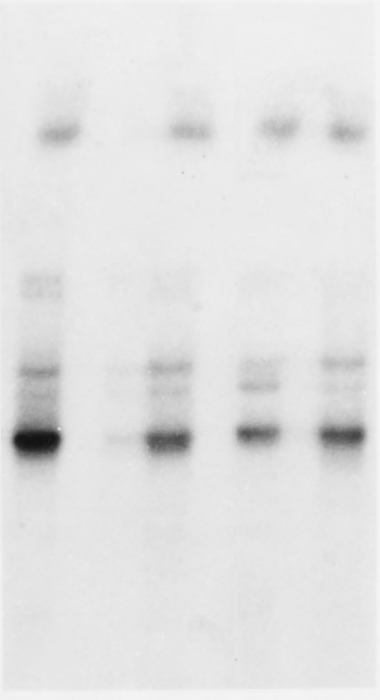
Exon 5.

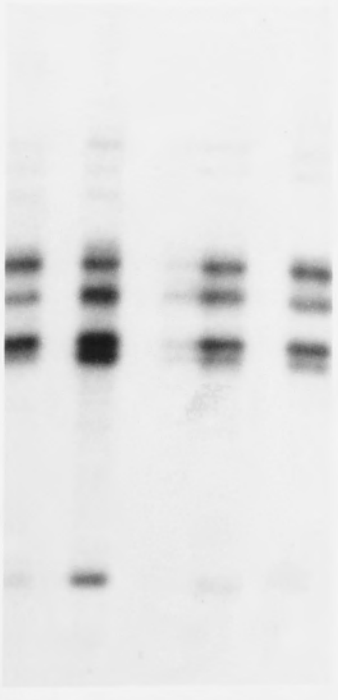
Non-Glycerol 10% Glycerol



Exon 6.

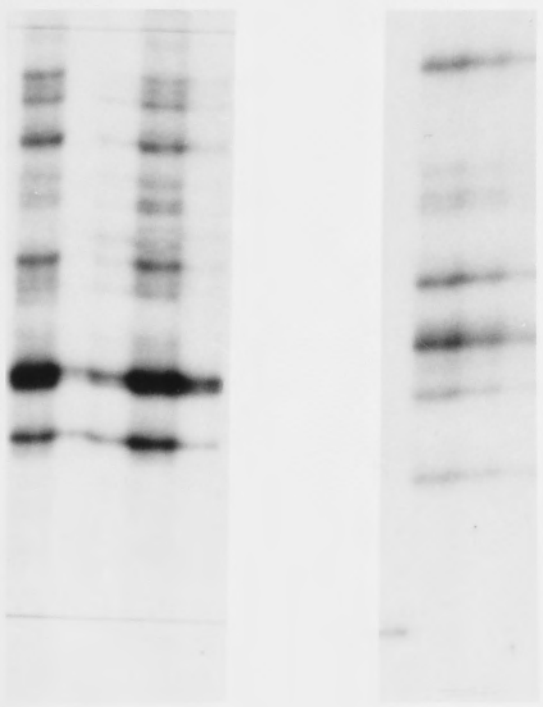
Non-Glycerol 10% Glycerol





Exon 7.

Non-Glycerol 10% Glycerol



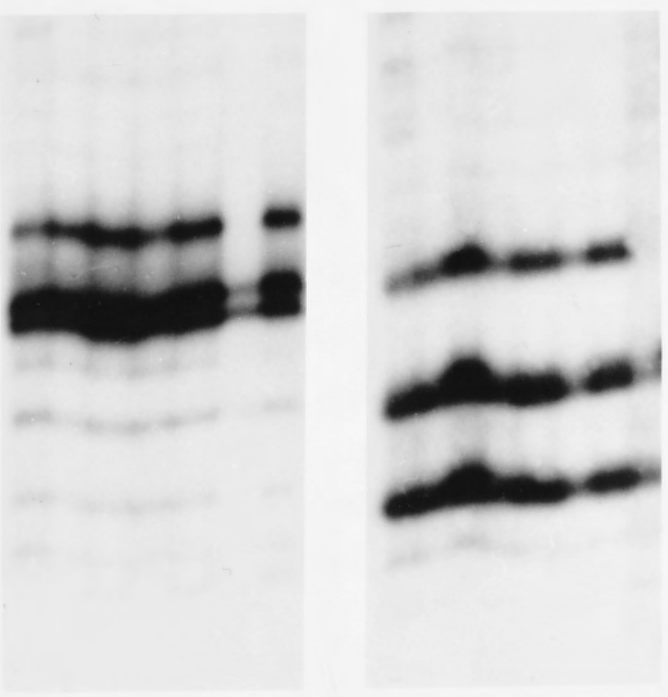
Non denatured

Exon 8.

Non-Glycerol 10% Glycerol

variant

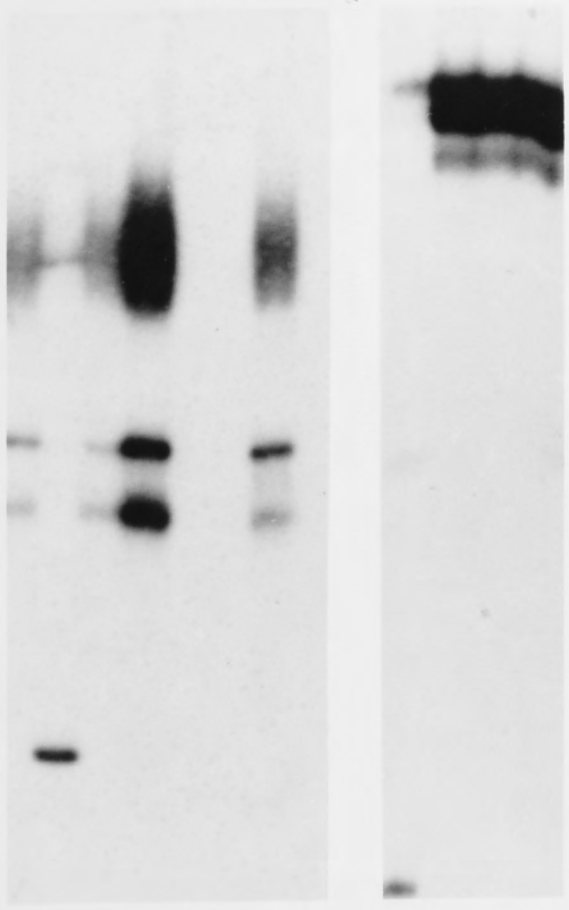
variant



SSCP Gels

Exon 9.

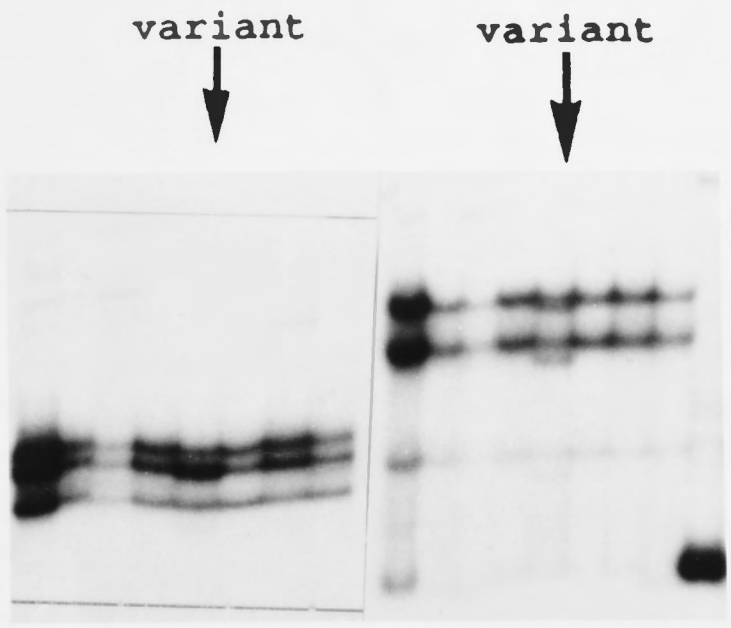
Non-Glycerol 10% Glycerol



Non denatured

Exon 10.

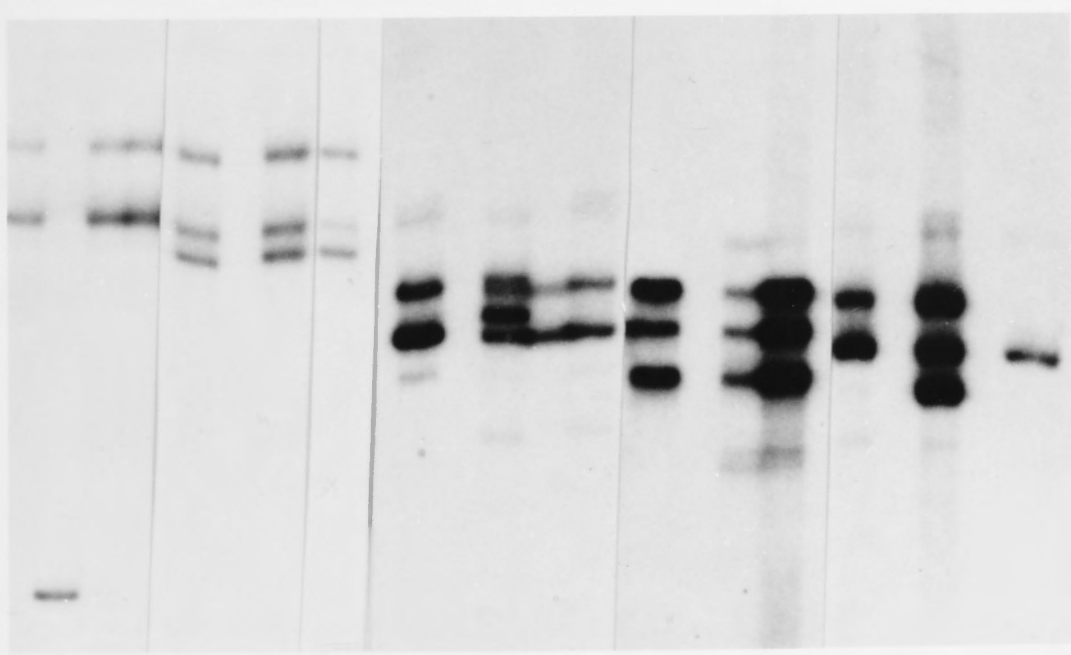
Non-Glycerol 10% Glycerol



Non denatured

Exon 11.

Non-Glycerol 10% Glycerol



Non denatured

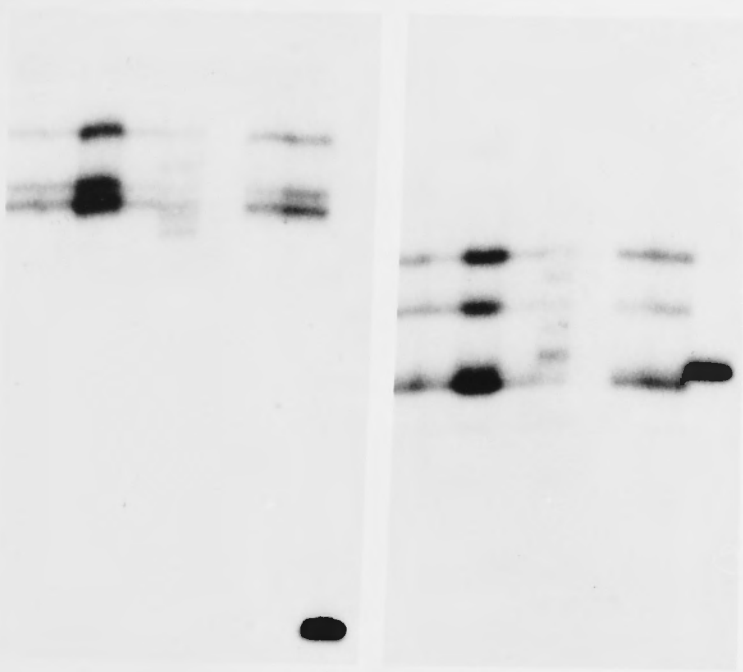
Non denatured

SSCP Gels

Exon 12.

Non-Glycerol 10% Glycerol

variant variant

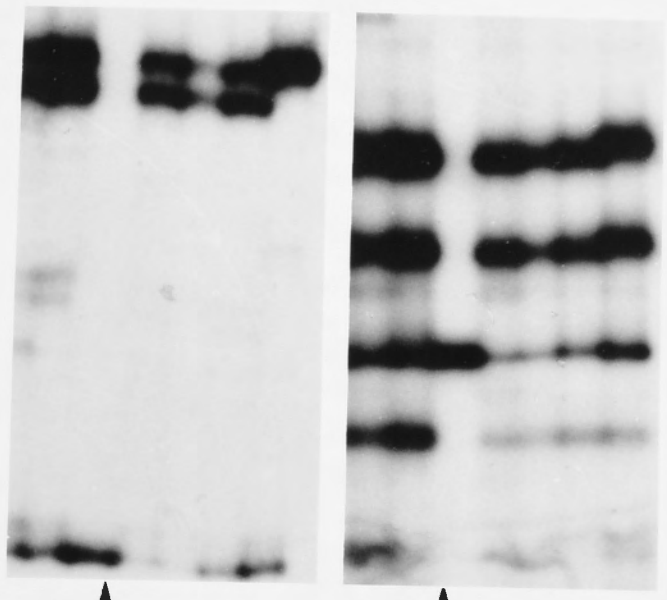


Non denatured

Exon 13.

Non-Glycerol 10% Glycerol

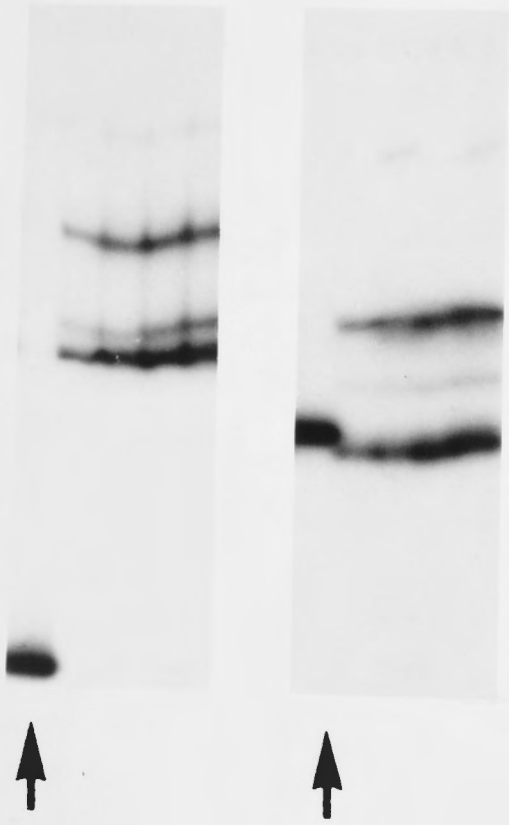
variant variant



Non denatured

Exon 14.

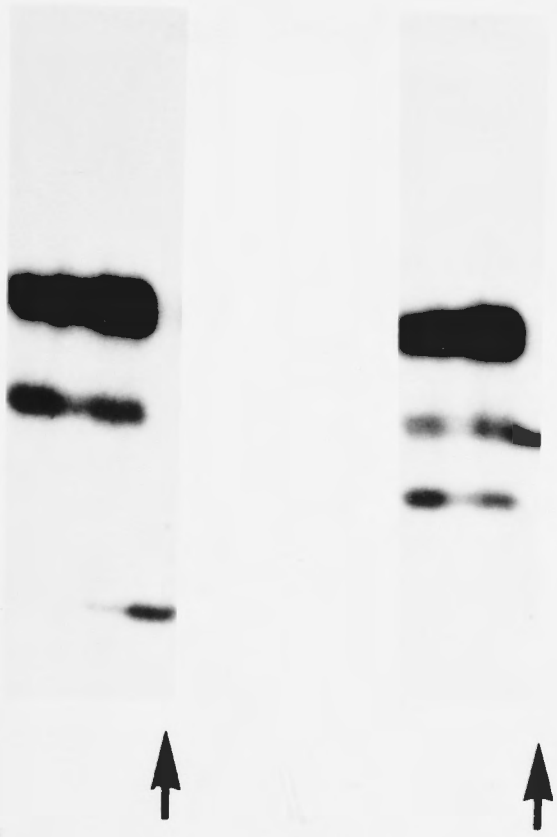
Non-Glycerol 10% Glycerol



Non denatured

Exon 15.

Non-Glycerol 10% Glycerol

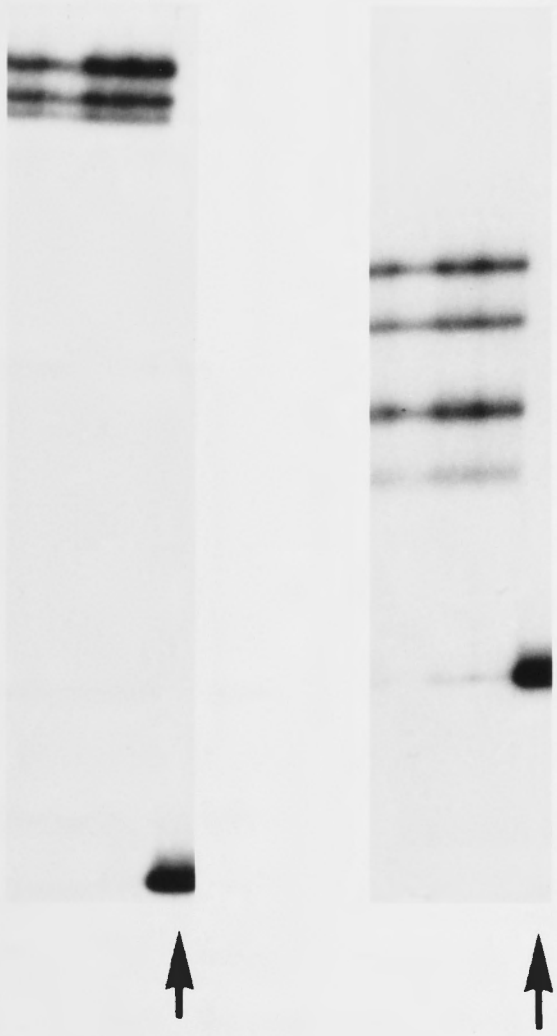


Non denatured

SSCP Gels

Exon 16.

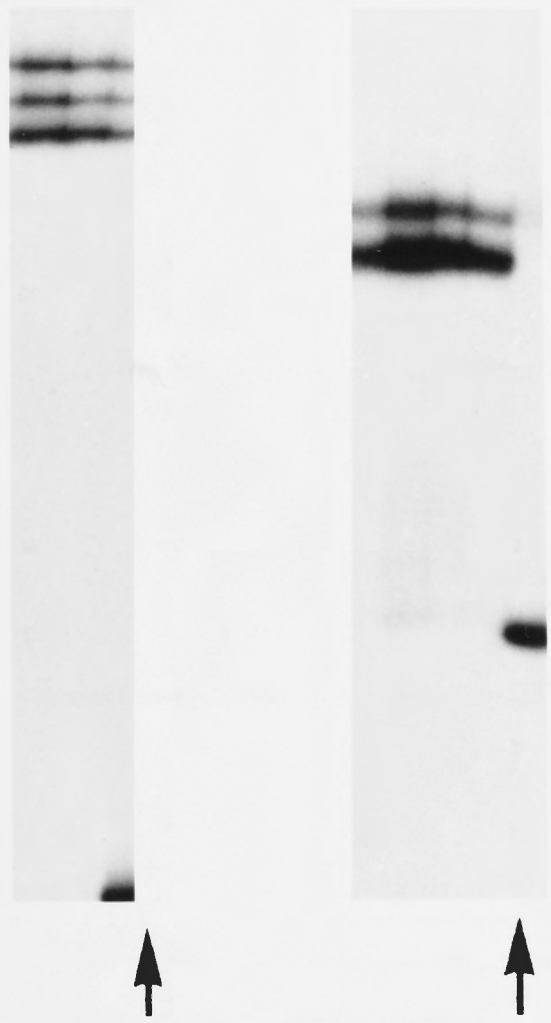
Non-Glycerol 10% Glycerol



Non denatured

Exon 17.

Non-Glycerol 10% Glycerol



Non denatured

Table 4.1 A summary of the variant SSCP patterns found in the C6 exons in four different protein variants and samples from Aboriginal Australians, Caucasians, Micronesians and Polynesians.

Protein Type	C6 Exons[#] 1	3	8	10	11	12	13
A1A[a]	-	√	√	√	2	√	√
	√	√	√	√	-	√	√
BB2	-	√	√	*	2	*	√
B2B2(WP3)	√	√	√	-	3	√	*
AB21	*	√	√	√	-	-	√
BM91	√	√	*	√	2	√	√
Caucasians[b]							
AA	√	*	√	√	2	√	√
BB	√	*	√	√	2	√	√
AB	√	*	√	√	2	√	√
Aboriginal[c]	√	√	√	√	3	√	√
Micronesian[c]	√	√	√	√	2	√	√
Polynesian[c]	√	√	√	√	3	√	√

[#] Exons 2, 4, 5, 6, 7, 9, 14, 15, 16 and 17 showed no variant SSCP patterns in any of the samples examined.

Symbols: * = unique SSCP pattern.
√ = most common SSCP pattern.
- = no SSCP result
2 = two SSCP bands seen.
3 = three SSCP bands seen.

[a] Two individuals had protein type A1A.

[b] AA, AB, and BB results are summarised from two Caucasian individuals.

[c] Micronesian, Polynesian and Aboriginal Australian patterns are combined results from three individuals.

patterns found in the samples examined and Table 4.2 shows the location, amino acid, codon sequence and protein phenotype of the samples from which sequence changes were identified. In one case a sample gave no SSCP result for two exons, that is PCR amplification failed for that sample in the radiolabelling experiment and it was decided to sequence these exons rather than repeat the radiolabelling PCRs and SSCP gels.

4.4.2.1 A1A Protein Type

No variant SSCP patterns were detected in these two samples for any of the C6 exons on either glycerol or non-glycerol gels. Exons one and 11 failed to amplify in one of the two samples, but since no variant pattern was found in the sample which did amplify, it was not considered necessary to investigate these exons with sequencing. Therefore the mutation causing this protein type was not detectable with the SSCP conditions utilised in this study.

4.4.2.2 BM91 Protein Type

Only exon eight showed a variant SSCP pattern for this sample. It was distinguishable since it showed the loss of a major band. Sequencing ten clones of this individual yielded no sequences different from published data. It is theoretically possible, although statistically unlikely, that ten normal clones were inadvertently selected over mutants. The same primer pair was used for SSCP analysis and generation of insert for cloning; therefore it cannot be argued that the mutation occurred in the primer binding site thus amplifying only the normal allele, because in that case no variant SSCP pattern would be seen. It appears that either the mutation was missed in sequencing or the real mutation was undetected under the conditions used for SSCP analysis and the variant pattern seen was a false positive result.

4.4.2.3 AB21 Protein Type

Two polymorphisms or mutations were found in this individual. The first was indicated by SSCP analysis and

Table 4.2 Polymorphisms detected by SSCP analysis and DNA sequencing.

Exon	Amino Acid Number	Amino Acid Change	Charge Change	Sequence Published	Sequence New	Enzyme site	Protein Type of proband sample
1.	NA	NA	NA	GGC	AGC	none	AB21
3.	98	Glu (A) ► Ala (NP)	yes	GAG	GCG	Dde I	A/B
10.	518	Cys (P) ► Cys (P)	no	TGT	TGC	Bsm I	BB2
11.	544	Asp (A) ► Gly (P)	yes	GAC	GGC	Aci I	BB2
[	544	Asp (A) ► Asp (A)	no	GAC	GAT	Acc I	BB2]
[	546	Gln (P) ► His (B)	yes	CAG	CAC	Bsr I	BB2]
	575	Arg (B) ► Gln (P)	yes	CGA	CAA	Mnl I	AB21
12.	630	Glu (A) ► Gly (P)	yes	GAA	GGA	Scr FI	BB2
13.	680	Arg (B) ► Trp (NP)	yes	CGG	TGG	Mae II	B2B2

P = polar amino acid (neutral charge)

NP = non polar amino acid (neutral charge)

A = acidic amino acid (negative charge)

B = basic amino acid (positive charge)

[Denotes that these two mutations invariably occurred together.

confirmed by sequencing; it occurred 74 bp upstream of the initiation codon. Its position suggests that it has no effect on the mature protein. The second sequence change was found only after sequencing as no SSCP result was obtained for this individual for exons 11 and 12. The sequence change occurred in exon 11, causing replacement of an arginine residue (positively charged) with a glutamine residue (neutral). The sequence change altered an Mnl I restriction enzyme site and thus was able to be confirmed with digest analysis (Figure 4.3). Using the computer program Isoelectric the predicted pI for this allele, assuming it has the background of a standard A allele, is 6.57 (Table 4.3). The standard A allele has a pI of 6.64 and the B allele a pI of 6.71. This suggests that the variant protein would migrate more anodally than the standard A allele. However this is not what was seen on an IEF gel; this allele migrated more cathodally than both a standard A and B allele (Chapter 3, Figure 3.1). This individual has been protein typed as AB21, but with Dde I digest analysis it appears as BB. Of all 143 Dde I and protein typed samples examined (Chapter 5) this individual is the only one in which the two tests gave conflicting results. The predicted pI assuming a B allele background is 6.64, the same as that seen for the standard A allele, suggesting that this individual possesses two mutations. One alters a standard B allele such that it migrates similarly to an A allele with IEF, while a second mutation might cause the protein to migrate further towards the cathode than a standard B allele.

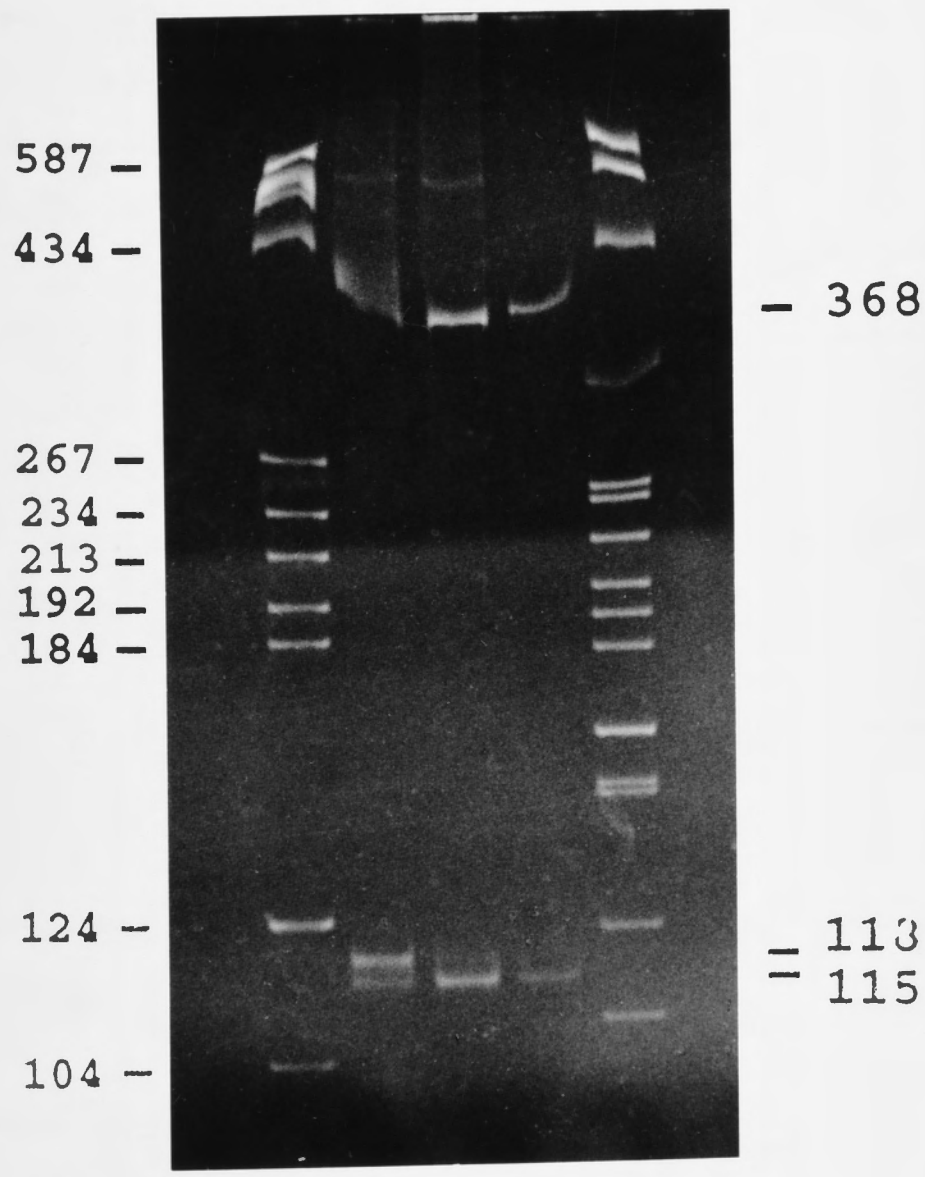
4.4.2.4 B2B2 Protein Type

This sample was originally found in a Nauruan population and designated WP3 for Western Pacific three allele. It was subsequently typed by Y Tokunaga (Blood Transfusion Service, University Hospital, Tokyo, Japan) and designated C6*B2. One variant SSCP pattern was identified in exon 13 and ten clones of this exon were examined. All sequences showed an arginine (positively charged) to tryptophan (neutral) substitution at amino acid 680. This mutation results in the protein having a

Figure 4.3 Amplified restriction fragment length polymorphism typing of C6 exon 11-12, Mnl I site. **a** Polyacrylamide gel after electrophoresis of Mnl I digested fragments of C6 exon 11-12 PCR products. An individual heterozygous for this polymorphism (lane 2), and individuals homozygous for the published sequence (lanes 3 and 4). The fragment sizes, in bp of a DNA standard (Hae III and Msp I digested pBR322) are shown in lanes 1 and 5 respectively. **b** Position, sequence, amino acid, restriction site and pI values for published and new alleles.

C6 Exon 11-12 Mnl I Digest

a



b

Polymorphism at cDNA 1845, amino acid 575, exon 11.103

Phenotype;	Published	New Allele
Codon;	CGA	CAA
Amino acid;	Arg	Gln
Mnl I site;	+ ($CCTCN_7/_7$)	- ($CCTCN_7/_7$)
pI (+A, +B);	6.64, 6.71	6.57, 6.64

Table 4.3 Calculated pI values# for the DNA mutations causing charge changes in the protein found in the rare-protein typed SLE patients.

Amino Acid Position	Name	pI	Underlying protein type	Change in pI*
98	Glutamic acid (AA)	6.64		
	Alanine (BB)	6.71		0.07
544	Glycine	6.71	A	0.07
		6.77	B	0.06
546	Histidine	6.67	A	0.03
		6.73	B	0.02
575	Glutamine	6.57	A	0.07
		6.64	B	0.07
630	Glycine	6.71	A	0.07
		6.77	B	0.06
680	Tryptophan	6.57	A	0.07
		6.64	B	0.07
544 Gly + 630 Gly		6.76	A	0.12
544 Gly + 630 Gly		6.83	B	0.12
546 His + 630 Gly		6.72	A	0.08
546 His + 630 Gly		6.79	B	0.08

* The change in pI is the difference between the standard A (pI = 6.64) or B allele (pI = 6.71) and the mutated allele.

\# pI values were calculated using the computer programme "Isoelectric" (see Chapter 2).

predicted pI of 6.64 (Table 4.3), identical to the standard A allele. This result does not correlate well with the migration pattern of this protein allele seen on an IEF gel as the sample runs more toward the cathode than an A allele (Chapter 3, Figure 3.1).

This sequence change destroyed a Mae II restriction enzyme site and so the homozygous nature of the mutation was confirmed with restiction digest analysis (Chapter 5, Figure 5.4). Fortunately one other WP3 (B2) DNA sample was available for analysis. This sample had the WP3 (B2) allele in the heterozygous form and after Mae II digest analysis it showed this mutation, also in a heterozygous form. Four Nauruan individuals with the "Nau" protein allele, also thought to be a C6*B2 allele (Ranford 1982, Tokunaga *et al* 1990), were available for testing as was the BB2 phenotype SLE patient. Mae II digestion analysis showed that two of the Nauruan samples and the BB2 SLE sample had the mutation in a heterozygous form and two lacked the mutation. These results were difficult to resolve; however analysis of further samples (Chapter 5) showed that 20 samples with common-protein phenotypes carried this mutation. One of these was homozygous for this change and two samples lacked the mutation but possesed the C6*B2 allele. It therefore appears that this mutation is not responsible for the B2 protein phenotype but is simply a relatively common polymorphism that was coincidently found in a homozygous form in this WP3 (B2) protein type sample.

4.4.2.5 BB2 Protein Type

Two variant SSCP patterns were detected for this individual. These exons were cloned, sequenced and compared with published data. A neutral polymorphism (Table 4.2) in exon 10 was detected and confirmed with Bsm I digestion (Chapter 5, Figure 5.1). Exon 12 showed a mutation at amino acid 630 which caused a glutamic acid (negatively charged) to glycine (neutral) amino acid substitution. This sequence change altered an Scr FI restriction enzyme recognition site and was thus able to

be confirmed with digest analysis (Figure 4.4). Using the Isoelectric program a pI of 6.77 was calculated (Table 4.3) for this protein allele, assuming a standard B allele background. This suggests that the protein would migrate further toward the cathode than a standard B allele which is in good agreement with what was seen on an IEF gel (Chapter 3, Figure 3.1). It appears then that this polymorphism is a good candidate to explain the B2 protein phenotype. Three Nauruan B2 protein typed samples and the homozygous WP3 protein typed sample were examined at this site in order to clarify whether this mutation is responsible for the B2 (Nauru and WP3) allele. Two Nauruan samples proved to have this change in the heterozygous form and the WP3 sample appeared to be homozygous for this change, which agrees with expectations based on their protein types. One heterozygous Nauruan B2 sample however, lacked the mutation.

Interestingly the BB2 sample proved to possess the exon 13 mutation identified from the variant SSCP pattern in the WP3 (B2) sample. No variant SSCP pattern was seen for exon 13 of this individual and yet digest analysis proved its presence. It is possible that the heterozygous form of this mutation explains the difficulty in determining its variant status by SSCP analysis.

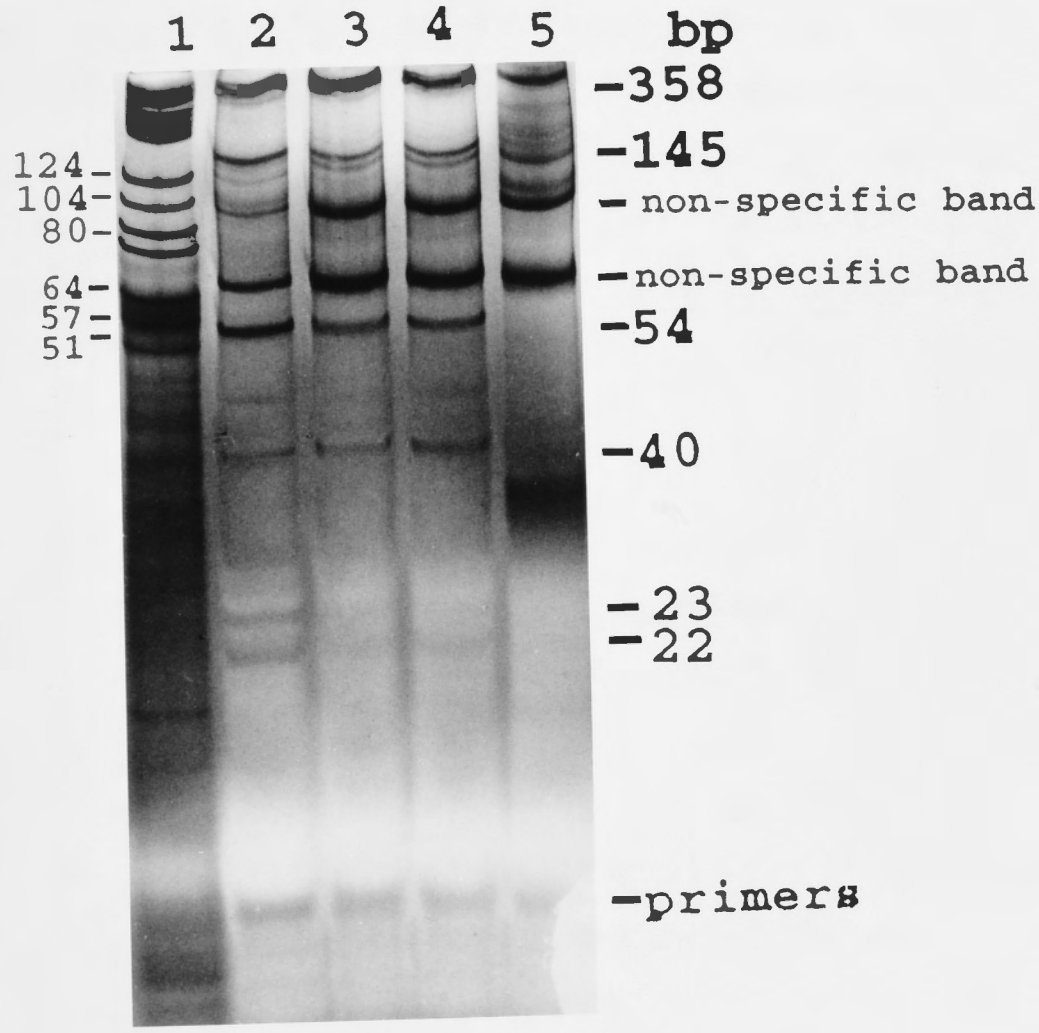
4.4.2.6 Exon 11

Exon 11 showed two relatively common SSCP patterns, one with a three band and another with a two band pattern (Figure 4.2). Hence an example of each pattern was examined by cloning and then sequencing. A Nauruan sample with the three-band SSCP pattern and a common protein type (AA) was chosen for examination but after sequencing no resolvable changes were identified. However all clones contained an ambiguous base (an N on automated sequence data) in the codon for amino acid 557 which was either an adenine or guanine but was not able to be definitively identified from the chromatographs obtained. Should this change prove real it would change a negative aspartic acid residue to a neutral asparagine residue. This charge

Figure 4.4 Amplified restriction fragment length polymorphism typing of exon 11-12, Scr FI site. **a** Polyacrylamide gel after electrophoresis of Scr FI fragments from samples heterozygous for the exon 12 polymorphism (lane 2), samples homozygous for the published sequence (lanes 3 and 4) and uncut PCR product (lane 5). The fragment sizes, in bp, of a DNA standard (Hae III digested pBR322) are shown in lane 1. **b** Position, sequence, amino acid, restriction site and pI values for the published and new exon 13 alleles.

C6 Exon 12 Scr FI Digest

a



b

Polymorphism at cDNA 2010, amino acid 630, exon 12.096

Phenotype;	Published	New Allele
Codon;	GAA	GGA
Amino acid;	Glu	Gly
Scr FI site;	- (CC'NGG)	+ (CC'NGG)
pI (+A, +B);	6.64, 6.71	6.71, 6.77

change was not indicated with IEF however other cases of charge changes not affecting IEF patterns have been identified (see further exon 11 results). Time constraints prevented further investigation of this sequence, so it was not equivically proven that this site was the explanation for the three-band pattern seen with SSCP. Further studies would examine this site with restriction enzyme Sfa NI to determine whether this mutation destroyed an Sfa NI recognition sequence and if so confirm it as real. Correlation of Sfa NI digest results with SSCP patterns for a number of individuals would identify whether this site was the explanation for the exon 11 two- and three-band SSCP pattern obtained in this study.

Exons 11 and 12 were amplified together for sequence analysis because the exon 12 forward primer was problematic. Since this insert was already cloned for the BB2 protein typed individual it was logical to use it to examine the two-band SSCP pattern found in exon 11. Sequence analysis gave interesting results. Ten clones were examined and of these two showed no changes at all, four showed a change from GAC (aspartic acid) to GGC (glycine) at amino acid 544, and four clones showed a change from GAC (aspartic acid) to GAT (aspartic acid) also at amino acid 544 along with a change from CAG (glutamine) to CAC (histidine) at amino acid 546. All of these changes affected restriction enzyme sites and thus were able to be confirmed. The GAC to GGC change at amino acid 544 created an additional Aci I site and proved to be real and not uncommon in other samples examined (Chapter 5). The predicted pI for this change alone was 6.77, assuming a standard B allele background (Table 4.3). It is possible that this mutation occurs on the same haplotype as the exon 12 mutation. This would give rise to the production of a protein with a predicted pI of 6.83, also assuming a standard B allele background. The change at amino acid 546 created a Bsr I site and also proved to be real and found in many other samples examined (Chapter 5). The pI value for the altered protein resulting from this

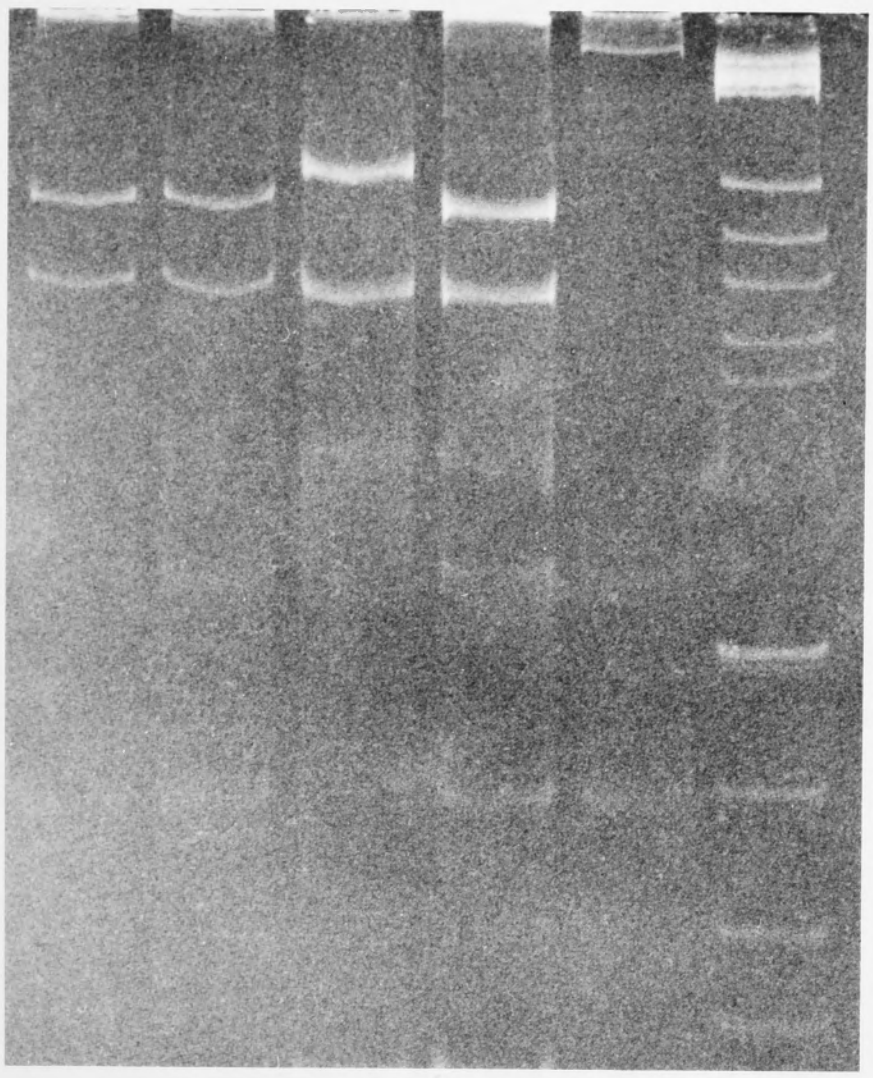
mutation alone was 6.73. If this mutation occurred with the exon 12 mutation a pI of 6.79 for the protein would result, again assuming a B allele background. The other change at amino acid 544 was a neutral polymorphism which destroyed an Acc I site, (Figure 4.5) therefore the normal allele could be confirmed. DNA from this individual was exhausted during the study and as such exons 11-12 PCR product for Acc I digestion were generated from unsequenced M13 cultures from 10 clones. Acc I restriction enzyme digestion demonstrated that there was no wild type exon 11 allele present in this PCR product. It is possible that no further wild type alleles were present in the M13 cultures which would explain this negative result. Also possible is that the two clones which gave wild type alleles for this individual were either sequencing or PCR errors, or resulted from a contaminated PCR product used for generation of the insert for cloning. If two clones were the result of PCR error this would indicate an unusually high degree of error for Taq I polymerase.

Interestingly Bsr I and Aci I digestion of a number of other samples suggested that there are other mutations in this region not described above. Chapter 5 deals with these new alleles in further detail. Based on Bsr I and Aci I digest results of a number of samples (Chapter 5), several of which had been SSCP typed at exon 11, the two- and three-band pattern did not correlate with either of these mutations nor a specific combination of these alleles. It thus appears that these mutations do not cause an altered migratory pattern with SSCP analysis under the conditions utilised in this study, and therefore would have been missed entirely had the unidentified mutation not caused a migratory change prompting sequence examination. The mutation causing the three-band SSCP pattern appears likely to be the unconfirmed mutation at amino acid 557 suggested after sequence examination of the Nauruan sample.

Figure 4.5 Amplified restriction fragment length polymorphism typing of C6 exon 11-12, Acc I site. **a** Polyacrylamide gel after electrophoresis of uncut 619 bp amplification product (lane 5) and Acc I fragments from an individual homozygous for the polymorphism (lane 3) and individuals homozygous for the published exon 11 sequence (lanes 1, 2 and 4). The fragment sizes, in bp of a standard DNA (Hae III digested pBR322) are shown in lane 6. **b** Position, sequence, amino acid, restriction site and pI values for the normal and new exon 11 alleles.

C6 Exon 11-12 Acc I Digest

a



b

Polymorphism at cDNA 1753, amino acid 544 exon 11.011

Phenotype;	Published	New Allele
Codon;	GAC	GAT
Amino acid;	Asp	Asp
Acc I site;	+ (GT(A/C)(T/G)AC)	- (GT(A/C)(T/G)AC)
pI (+A, +B);	6.64, 6.71	6.64, 6.71

4.4.3 SSCP Detection Analysis

It is important to note that some SSCP patterns showed more than the expected two bands for a homozygote or four bands for a heterozygote (Figure 4.2). This is a fairly common occurrence with SSCP gels as some single-stranded DNA fragments may have more than one stable conformation under the conditions at which the gel was run. This problem or artifact may have been overcome if each SSCP gel had been optimised for each specific exon. Initially the SSCP conditions were optimised for the exon three polymorphism; identical conditions were then used for each exon. With hindsight this process may not have been optimal but less information on the parameters affecting SSCP was available at the commencement of this analysis.

Overall SSCP analysis in this study detected sequence changes in exons one, three, 10, 11, 12 and 13. One possible false positive was also found and at least two mutations, those causing the BM91 and A1A protein alleles, were missed with this screening study. Interestingly four mutations were found and confirmed in exon 11, while two others were indicated by digest analysis. None of these correlated with a particular variant SSCP pattern, suggesting that they did not alter the migratory pattern of the single-stranded DNA under the conditions used in this study. The mutations detected with SSCP analysis included both transitions and transversions and are listed in Table 4.4.

4.5 Discussion

Examination of the two published cDNA sequences for the C6 gene gave an absolute correlation between a sequence difference and the protein phenotypes seen with isoelectric focusing. This change leads to a unit charge change in the resultant protein correlating well with the difference seen on an IEF gel. It is still possible that this difference is not the sole cause of the protein polymorphism and that perhaps it is only one of a number of substitutions which are in linkage disequilibrium.

Table 4.4 A review of the mutations detected with SSCP analysis

Exon	Sequence change	Type	Enzyme site	SSCP Detection
1.	G → A	transition	none	yes
3.	A → C	transversion	Dde I	yes
10.	T → C	transition	Bsm I	yes
11	A → G	transition	Aci I	no
11.	C → T	transition	Acc I	no
11.	G → C	transversion	Bsr I	no
11.	G → A	transition	Mnl I	-
12.	A → G	transition	Scr FI	yes
13.	C → T	transition	Mae II	yes

- this sample failed to amplify for the SSCP gel and therefore no SSCP result was available, the mutation was found after sequencing.

This however appears unlikely based on; (1) the absolute correlation found in 143 samples examined in this study (Chapter 5), (2) the lack of other variant SSCP patterns correlating with protein type and (3) that it is the only difference found in two complete cDNA's. The replacement of protein typing with a simple and robust DNA-based analysis allows more convenient examination of the common C6 polymorphism, particularly for studies of evolution and differences between ethnic groups.

Unfortunately SSCP analysis did not indicate a potential site for mutations responsible for the two protein types A1A and BM91. No variant patterns were detected for the two A1A protein type samples whereas a variant pattern was found for the BM91 sample but sequencing revealed no mutation, suggesting a false positive result. Examples of false positives with SSCP have been reported (Hongyo *et al* 1993) but are thought to be rare (Gaidano *et al* 1991, Sarkar *et al* 1992). One explanation is that false positives are the result of semi-stable transitional conformations which disappear when the sample is run under different conditions (Michaud *et al* 1992). This is probably unlikely in the present situation since the variant pattern had lost a major band not gained additional bands. An alternative and highly possible explanation is a sequence change caused by PCR error. Triggs Raine *et al* (1991) report that errors incorporated into a PCR product by Taq polymerase (Tindall and Kunkel 1988) do not form a large enough proportion of the product to visibly interfere with the SSCP analysis whereas Hongyo *et al* (1993) report that mutant SSCP bands could be clearly discerned when as little as three percent of the gene copies contained the mutation or error. This conflict probably arises because of the variability in optimisation of the SSCP gel running conditions and the differing nature of the sequence being examined. As such it remains possible that the false positive variant pattern arose through PCR error. Repetition of the PCR reaction and SSCP gel may have proven this true since it is highly unlikely the same error would arise a second time.

The sensitivity of SSCP analysis is well known to vary, commonly from around 70 to 100 percent depending on optimisation of gel running conditions (Glavac and Dean 1993). The present study detected two of five mutations expected to contribute to protein phenotype, representing a 40 percent efficiency rate for detection of expected mutations. SSCP analysis also detected four unexpected mutations which therefore suggests the technique is of considerable value in screening for polymorphisms. SSCP analysis has been well utilised since initiation of the present study and greater knowledge of the affect of important parameters has been gained. Temperature, glycerol content, gel concentration, buffer ionic strength and fragment size are all important parameters affecting the sensitivity of SSCP. It was also assumed that the type of mutation would influence its detection rate with transversions expected to cause a greater alteration in the secondary DNA structure than transitions. This however appears not to be true since this and several other studies (Glavac and Dean 1993, Sheffield *et al* 1993) show good detection rates for both transitions and transversions. Glavac and Dean (1993) along with Sheffield *et al* (1993) suggest that the position and type of mutation are not correlated with the rate of detection but that the base sequence around the mutation appears to have a greater affect on mobility. In view of these findings it could be that the mutations causing the A1A, B21 and BM91 protein phenotypes occur in areas of the gene which are perhaps GC rich and may not be amenable to detection with SSCP analysis. Fraser *et al* (1992) have used the computer program Squiggles (University of Wisconsin, Genetics Computer Group) to predict secondary structure of single-stranded DNA, based on RNA structure. They have found good agreement between experimental findings of mobility shifts, resulting from altered conformation, and predicted conformation changes based on these models. This technology may become very useful when utilising SSCP for identification of known mutations as it may allow design of very effective PCR primers.

The sample with protein type AB21 was shown to have two mutations. One in the 5' untranslated region which would not be expected to contribute to the variant protein phenotype while the second appeared in exon 11. This mutation resulted in an overall charge change with the replacement of a positively charged arginine residue for an uncharged glutamine residue. The predicted pI for this change is 6.57, less than a standard A allele (pI = 6.64); however IEF patterns suggest it has a pI greater than the standard B allele (6.71). Exon three results are also conflicting for this individual, suggesting it is homozygous C6*B at the DNA level. A possible explanation for these results is that the individual is homozygous B at the exon three site and carries two polymorphisms, one as yet unidentified. The identified polymorphism changes a normal B allele such thats its pI is 6.64, identical to that of a standard A allele. A second sequence change, on a standard B allele background, could produce an amino acid change resulting in a pI of > 6.71. This hypothesis would result in a total protein IEF pattern identical to that seen for this individual. Other explanations include the possibility that exon three results are incorrect and this mutation is one of several changes in the B allele, the combination of which give a pI value greater than 6.71. This however appears unlikely since the change in pI would have to be greater than 0.14, requiring several additional charge-changing mutations. Alternatively this mutation may significantly change the conformation of the protein, consequently altering its migratory pattern in a polyacrylamide gel matrix. The change from an arginine residue to a glutamine adds two CH_2 molecules and while this would not appear to greatly alter the protein's secondary structure, it is difficult to predict the consequences of this change. Protein structures are dependant on many chemical bonds and small changes in constituent molecules may alter bond interactions substantially and hence markedly change the folding pattern of the protein.

No further samples with this protein type were available so it was not possible to examine this site in other individuals to determine whether this mutation co-occurs with the variant protein phenotype. Future studies would involve obtaining a number of AB21 and BB21 protein-typed individuals plus samples with common protein types and carrying out Mnl I digest analysis on exon 11 along with further mutation screening in the B21 samples. Should B21 samples exclusively possess the exon 11 mutation compared with common-protein type samples it would be reasonable to assume that it is associated with this protein variant. It is anticipated that another change would be found for this protein variant and examination of this site in other B21 samples would confirm its contribution to this phenotype. There is however the possibility that a B21 protein phenotype as seen on an IEF gel is actually a heterogenous group of proteins with very similar pI values, and that examining a large number of samples would show that not all had the same sequence change as the sample examined in the present study.

The WP3 sample, also described as B2, was homozygous for a sequence change in exon 13. This change resulted in an overall loss of positive residues because the mutation caused the substitution of a positively charged arginine residue for a neutral tryptophan. The predicted pI for this sample; 6.64 does not correlate with what is observed on an IEF gel as the sample runs more towards the cathode than a normal A allele. Digest analysis of 63 samples, (Chapter 5) including six samples with the B2 allele, showed that this mutation occurred in samples with normal protein types and also that it was not present in some samples that possessed the B2 allele. Therefore it was concluded that this mutation is a common polymorphism of the C6 gene and is not responsible for the B2 or any other rare-protein allele.

Six mutations were found in the BB2 protein typed sample, with two being neutral polymorphisms. The

remaining mutations all caused charge changes. The exon 13 mutation is described above and appears not to be responsible for the B2 allele. The amino acid 544 (Aci I site) and amino acid 546 (Bsr I site) changes were found in relatively high frequencies in samples not showing variant-protein patterns (Chapter 5) and therefore appear not to be responsible for this protein allele. The exon 12 mutation at amino acid 630 (Scr FI site) was identified in two other Nauruan B2 type samples, however one C6*B2 sample did not show the Scr FI site mutation. This result suggests that the exon 12 mutation may be one cause of the B2 (Nau, WP3) protein phenotype, but that at least one other mutation also results in the characteristic B2 IEF pattern. It therefore appears that the B2 protein phenotype represents a group of proteins which are heterogenous in their underlying DNA mutations. Similar findings have been reported for the C4 gene (Belt *et al* 1984, 1985, Yu *et al* 1986) and it therefore appears likely that this will also be true for C6.

Many of the arguments for whether a specific sequence change is the cause of a protein phenotype are based on the predicted pI value, calculated using the computer program Isoelectric. While these predicted values are thought to correspond closely with what is seen on an IEF gel (Assis Prof W Hildebrand, personal communications) they still remain estimates. Despite the large number of C6 protein studies reported, few have attempted to approximate the pI of C6 alleles. An early study by Alper *et al* (1975) gave an approximate range of 6.0 to 6.5 for the common protein types. The values obtained with the computer package give slightly higher values of 6.64 and 6.71 for C6*A and C6*B respectively. This difference is considered insignificant after consideration of the error involved in estimating the values experimentally. It thus appears that the calculated pI values are in reasonable agreement with previous estimates and as such the predicted pI values for new alleles should be relatively accurate.

The cluster of mutations found in exon 11 of the C6 gene suggests that this area is a hot spot for mutation. Exon 11 codes for the entire third thrombospondin (TSP) repeat and some intervening protein sequence between this and the first short consensus repeat (SCR) in the C6 protein molecule. Interestingly the exon three protein allele determining site also occurs in a thrombospondin repeat module. Haefliger *et al* (1989) report that the C5b binding site on the C6 protein occurs in the carboxyl terminal portion of the protein, encoded for by the 5' end of exon 12 through to exon 17. Nakano *et al* (1991) further localised the C5b binding site to within the two short consensus repeats, encoded for primarily by exons 12 to 14. This region includes the exon 12 and 13 sequence changes identified but not the four exon 11 changes found. The present study produced no evidence that these sequence changes alter the function of the protein because a non-functional detection system for C6 identification was used. Other studies however have utilised a hemolytic C6 detection system which is a functional assay and there has been no suggestion that the variant protein phenotypes affect the function of this protein. There are no studies available that have examined whether these rare C6 protein types incorporate as normal into the MAC. It would be interesting to determine whether these rare-protein phenotypes reduce or enhance binding of C5b or C7.

From this study it is clear that single-stranded conformation polymorphism is a valuable tool for mutation detection; however particular care must be taken to optimise gel conditions as these are crucial for detection efficiency. Although this technique did not detect all DNA changes, there is no certainty that any screening method would. Even sequencing a number of clones from each exon may not detect a polymorphism present in a heterozygous state, as was possibly the case in one sample in this study. Efficiency of detection of mutations using the more common screening methods of SSCP and DGGE along with heteroduplex analysis will depend to a large extent on the sequence surrounding the mutation site. It may be

that some sites will not be amenable to detection with any one of these methods but it is possible that a combination of these techniques might consistently increase detection rates to 100 percent. Despite varying efficiency rates screening methods remain valuable as time- and money-saving devices compared with sequencing.

4.6 Conclusions

(1.) The site determining the common C6 protein alleles A and B occurs at amino acid 98, encoded for by exon three. Dde I digestion of PCR amplified C6 exon three can be used as a simple and reliable alternative method for C6 protein phenotyping.

(2.) SSCP analysis is a simple and fast method of screening for mutations in small DNA fragments; however considerable care must be taken to optimise gel conditions for each DNA fragment.

(3.) A total of nine new polymorphisms were detected in the C6 gene but only one of these convincingly correlated with a rare-protein allele. This suggests that the C6 gene is significantly more polymorphic than protein data suggests.

Chapter 5

Chapter 5

Population Studies of C6 DNA Polymorphisms

5.1 Introduction

The study of DNA polymorphisms across different species and populations enables workers to determine whether a gene is highly conserved or has diverged significantly since separation of the different groups. Golding (1992) considers it more likely that identical DNA sequences result from their being present in the ancestral group than through separate mutational events, however other workers (Sigurdardottir *et al* 1991, Anderson *et al* 1991) consider convergent evolution, ie independent mutational events, a more likely explanation for sequence identity across species. Thus information about the presence of identical DNA sequences among different groups leads to theories about evolution, migration and divergence of these populations.

Identification of polymorphisms shared across different species has led to the theory of trans-species evolution (Arden *et al* 1980, Klein 1980, Arden and Klein 1982). These polymorphisms are inferred to have been present in the ancestor of the species and to have survived within both species throughout their divergence. This sharing of polymorphisms across several different species requires a selective explanation (Takahata and Nei 1990, Golding 1992).

Trans-species polymorphisms have been identified across different species of mice (McConnell *et al* 1988, Sagai *et al* 1989, Figueroa *et al* 1988) and rats (Figueroa *et al* 1988) and between chimpanzees and humans (Lawlor *et al* 1988, Mayer *et al* 1988, Gyllensten and Erlich 1989). Parham *et al* (1989) suggest that more than 50 percent of 328 HLA substitutions may be shared by humans and chimpanzees. Takahata and Nei (1990) show that these shared polymorphisms require strong selection because the probability of observing this phenomenon for selectively neutral alleles is exceedingly small. Most neutral polymorphisms are retained for only an average of 4Ne

The effect of selection is also inferred from the ratio of amino acid replacements to silent nucleotide substitutions in comparison of alleles between species (Gyllensten and Erlich 1989). A larger number of amino acid replacements relative to silent and intronic changes suggests that selection maintains these polymorphisms (Hughes and Nei 1988, Gyllensten and Erlich 1989). The strength of the selection implied is surprising, particularly as many of the polymorphisms studied are of unknown or disputed selective value to the individual (Golding 1992). It has been speculated that these polymorphisms may periodically be under strong selection, however evidence for this is still lacking (Golding 1992). Even with the enormous amount of published work aimed at identifying the selective force associated with the A-B-O polymorphism, Cavalli-Sforza and Bodmer (1971) report that no real understanding of such forces has yet emerged

A further consequence of the idea of trans-species polymorphisms is that there would have to have been ancient allelic diversity (Klein 1987). This implies that a large group of individuals gave rise to present day humans, as opposed to a single line (Klein 1987, Takahata 1990). If this were not the case the polymorphisms would have been lost by random drift in the bottleneck stage (Figueroa *et al* 1988).

The polymorphic nature of complement component six has proved useful as a tool for examining different ethnic human groups. A study of polymorphisms in the C6 gene, combined with further molecular studies, could give an indication of similarities and differences in the genetic makeup of different populations and in tracing population origins and movements. Examination of the C6 protein alleles shows that Polynesians, Melanesians and Caucasians have a C6*A frequency of approximately 0.60 and C6*B 0.40 (Chapter 3, Table 3.1) whereas Japanese and Chinese population groups have C6*B frequencies higher than C6*A. In Micronesians C6*B is by far the most common allele (Ranford *et al* 1982, Ranford 1989) and Australian

Aborigines have approximately equal C6*A and B frequencies. C6*B2, the most prevalent rare allele, was first identified by Hobart and Lachmann (1976) and was designated C6*B2 by Mauff *et al* (1980). It occurs in the Micronesian people from Nauru with a frequency of approximately 0.07 (Ranford *et al* 1982) and has since been identified in all but one of the Pacific area populations that have been C6 protein typed. C6*B2 is found at frequencies around 0.05 in Aboriginal Australians and Japanese, while additional rarer alleles have also been observed in almost all populations studied, usually occurring at a combined gene frequency of around 0.05 (Hobart *et al* 1993a).

The continent of Australia has been occupied for more than 50 000 years (Roberts *et al* 1990). The origin of the Australian Aboriginal people is still an issue of great debate among anthropologists (Bellwood 1989) and has given rise to several possible theories. The multiple origin hypothesis was postulated from anthropometric and cranial diversification studies of both living populations and fossils. This hypothesis suggests that either two (Thorne 1980) or three (Birdsell 1977) ancestral groups from the original people of Java and Southeast Asia settled in Australia. Other theories suggest a single group of people from Java settled in Australia, spread out and diversified (Habgood 1985, Brown 1987). Since the initial settlement in Australia the Aboriginal people have been isolated and have evolved into a separate race now called Australoid. Prior to European settlement, around 200 years ago, the Aboriginal people existed as hunter-gatherer tribes and were scattered throughout the entire continent, including Tasmania (Kirk 1981). The estimated size of the Aboriginal population at the time of first European contact was 250 000 to 300 000 (Radcliffe-Brown 1930).

Other native people of the Pacific were classified by European voyagers according to their geographical distribution, thus the terms Melanesians, Micronesians and

Polynesians were adopted. Anthropological, archaeological and linguistic evidence suggests that colonisation of these regions of the Pacific occurred in two waves. The New Guinea highland people are thought to share the same ancestral group as the Aboriginal Australians (Bellwood 1989). This founder population is believed to have come from Java 40 000 to 50 000 years ago and diverged once established in their new homeland. Approximately 4 000 years ago a second wave of people, this time of southern Mongoloid origin, moved from Southeast Asia into the Pacific, spreading out to occupy regions of Micronesia and Polynesia (Bellwood 1989). Genetic studies of indigenous people of the Pacific provide a unique opportunity to examine patterns of migration and the effects of isolation from other human groups (Serjeantson and Hill 1989). Additionally the knowledge of the genetic background of native populations allows identification of possible disease-susceptibility loci.

A study of C2, C4, C6 and factor B, the more polymorphic complement component proteins, have been carried out in samples from the Pacific (Ranford *et al* 1982, Ranford 1989) and these show that the common alleles are the same in all populations studied. Ranford (1989) suggests that these groups may have evolved from a common ancestral origin, however differences in the frequencies of these common genes between the groups shows they have been evolving separately for considerable time. In general, variation in the complement proteins is restricted in Polynesians, Micronesians and Melanesians as C2 is not variable, fewer C4 alleles are found compared to other populations and more than 80 percent of the factor B genes are the common, slow allele. Despite some similarities in genetic make-up of the Polynesians, Melanesians and Micronesians, they are not a homogenous group. Micronesians lack the C4B null allele and they have a unique C6 gene frequency distribution (Ranford 1989).

Isolation of some tribes of Aboriginal Australians has occurred through vast land distances and may be comparable to many of the remote islands of the Pacific (Bellwood 1989). As a result some genetic diversity has developed between them. Differences in C4Q0, C4A4 and C4B2 allele frequencies between some tribes has arisen (Ranford 1989), possibly as a result of migration by genetically different people or by divergent evolution through geographic or cultural isolation.

The genetics of the complement components in the Pacific is not sufficient evidence to reliably conclude the migrational patterns that occurred around the area. However, in combination with other genetic, linguistic, archaeological and anthropometric studies this information may be useful in helping to clarify origins of the people of the Pacific.

Further genetic sudies in the Pacific have utilised the human leucocyte antigen (HLA) system. Conclusions reached in one such review (Serjeantson and Hill 1989) include the common ancestry of Australian and Papua New Guinean highlanders, evident from HLA -B, C linkage. Similarity in HLA distributions of Papua New Guineans and Melanesians suggest a common ancestry. Polynesians and Micronesians have a very limited number of similarities in the HLA system and it appears that Micronesia has experienced genetic admixture not found in other groups of the Pacific.

Early Australian travellers left genetic evidence of their presence in Melanesia, but did not transfer Melanesian genes into Polynesia (Serjeantson and Hill 1989). Polynesia was probably colonised by a small number of settlers, since many HLA alleles are missing from the region. HLA allele frequencies and their linkage show that New Zealand Maoris, Hawaiians and Easter Islanders are of the same Polynesian branch. The possibility of an Amerindian contribution to Eastern Polynesia is possible although it is probably of great antiquity.

The migratory patterns and divergence dates suggested by the HLA system have proven remarkably consistent with those suggested by archaeological, linguistic and cultural studies of the region. This is quite remarkable given that this system is such a small segment of the human genome and it has long been subjected to forces of selection, mutation and drift.

Examination of further polymorphisms in the C6 gene in several human groups and chimpanzees could generate more information about the similarites of different ethnic groups as well as about the evolution of C6. The common protein allele determining site and four polymorphisms identified in Chapter 4 were examined in small groups of Aboriginal Australians (from the Kimberley region), Caucasian-SLE patients, Caucasian-laboratory staff and students, Polynesians (from Rarotonga) and Micronesians (from Nauru) as well as a group of chimpanzees. Regions of the C6 gene were PCR amplified from these samples and subjected to restriction enzyme digest analysis to determine the presence or absence of specific polymorphisms.

5.2 Aims

To examine the occurrence of five DNA polymorphisms in the C6 gene in Aboriginal Australians, Caucasians, Micronesians, Polynesians and chimpanzees.

5.3 Materials and Methods

5.3.1 Samples

The samples examined were Caucasian SLE patients, including the rare-protein typed individuals, Caucasian laboratory staff and students, a group of Polynesians (Rarotongans), Micronesians (Nauruans), Aboriginal Australians (Kimberley region) and a number of chimpanzees. The SLE patients, laboratory staff and students were protein typed for this study (Chapter 3) while the Rarotongans, Nauruans and Kimberley Aboriginals had been typed previously by P Ranford (1989). DNA extraction from the Polynesians and Micronesian samples

was carried out for earlier work while the Aboriginal Australian samples were extracted from hemolysates for the present study. Chimpanzee samples were from animals held under long-term observation in one of several primate colonies of the Laboratory of Slow, Latent and Temperate Virus infection of the National Institute of Health. The majority were captured in the wild in Africa, but a small number may have been born in captivity. DNA from these samples was extracted from hemolysates by A Browne in 1992. Storage of all DNA was at 4^{o}C.

5.3.2 Methods

All samples were examined by PCR-RFLP following discovery of new polymorphisms affecting restriction enzyme sites (Chapter 4). Exons three, 10, 11-12 and 13 were PCR amplified then digested with Dde I, Bsm I, Aci I, Bsr I and Mae II respectively. Restriction enzymes Aci I and Bsr I were both used for exon 11 polymorphisms. Exon three digests were run on 15% polyacrylamide gels, exon 10 on 10% polyacrylamide gels, exon 11-12 - Aci I 2% agarose, exon 11-12 - Bsr I 2% agarose, and exon 13 10% polyacrylamide gels.

A minimum number of ten individuals (20 chromosomes) were chosen for examination, where possible. The probability of not detecting a given polymorphism with this sample size was calculated for several allele frequencies. For allele frequencies of 0.5, 0.25 and 0.1 the probabilities of non-detection are 0.02, 2.3 and 24.4 percent respectively.

5.4 Results

5.4.1 Exon 3

Table 5.1 shows the results of 153 Dde I digests in the different groups examined. Comparison with protein data where available shows absolute correlation between the two tests except for one sample, a variant protein type AB21 discussed in Chapter 4. The chimpanzees had not been protein typed so it was not possible to compare DNA results but the fact that they possessed this polymorphism suggests the exon three substitution is the protein allele determining site in chimpanzees as well as humans. The prevalence of AA homozygotes in chimpanzees suggested a higher A allele frequency in chimpanzees compared with

Table 5.1 Genotypes and allele frequencies of the C6 exon 3, Dde I locus in Aboriginal Australians, Caucasian SLE patients and healthy individuals, Micronesians, Polynesians, and chimpanzees.

Group	N	Genotype Frequencies AA Obs	AA Exp	AB Obs	AB Exp	BB Obs	BB Exp	χ^2	P	Percent Allele Frequencies A	B
Aboriginal Aust.	40	7	8	12	9	1	2	0.803	0.669	65.0	35.0
Caucasians - SLE[a]	104	20	19	23	25	9	8	0.168	0.920	60.6	39.4
Caucasians - healthy	30	2	4	11	8	2	4	0.178	0.411	50.0	50.0
Micronesians	64	1	3	17	13	14	16	1.667	0.435	29.7	70.3
Polynesians	48	6	7	14	12	4	3	0.287	0.866	54.2	45.8
chimpanzees	20	9	9	1	1	0	0	NA	NA	95.0	5.0

[a] Rare-protein typed SLE patients not included

NA = not applicable.

N = number of chromosomes.

humans. The human population groups are all in good agreement with Hardy-Weinberg expectations (Table 5.1). Sample sizes are too small for accurate comparisons of the allele frequencies in the different populations, however the Nauruan group appears to have somewhat different allele frequencies to all other groups. The observed allele frequencies in the different population groups are in general quite different from those of published studies for larger groups (Chapter 3, Table 3.1) of the same population, highlighting the effect of restricted sample size.

5.4.2 Exon 10

Results of PCR-RFLP digests and analysis of exon ten using the restriction enzyme Bsm I are shown in Figure 5.1 and Table 5.2. This polymorphism was found in the heterozygous form in six of the 65 human samples examined and is therefore not uncommon. It was not found in the chimpanzee group, suggesting it arose after divergence of chimpanzees and humans. The sequence change does not cause an amino acid change in the protein and is therefore a silent polymorphism. This study suggests it is a relatively common polymorphism, occurring in at least four different population groups. It was of interest to note the apparent increased frequency of this polymorphism in the Nauruan group examined.

5.4.3 Exon 11

This exon shows four DNA polymorphisms (Chapter 4), possibly representing a hot spot for mutation in this gene. Two of these polymorphisms were chosen for examination in these population groups as it was probable these were polymorphisms not associated with a specific rare-protein type. The results of PCR-RFLP analysis are shown in Table 5.3 and 5.4. Both of these polymorphisms occur relatively frequently in the samples examined.

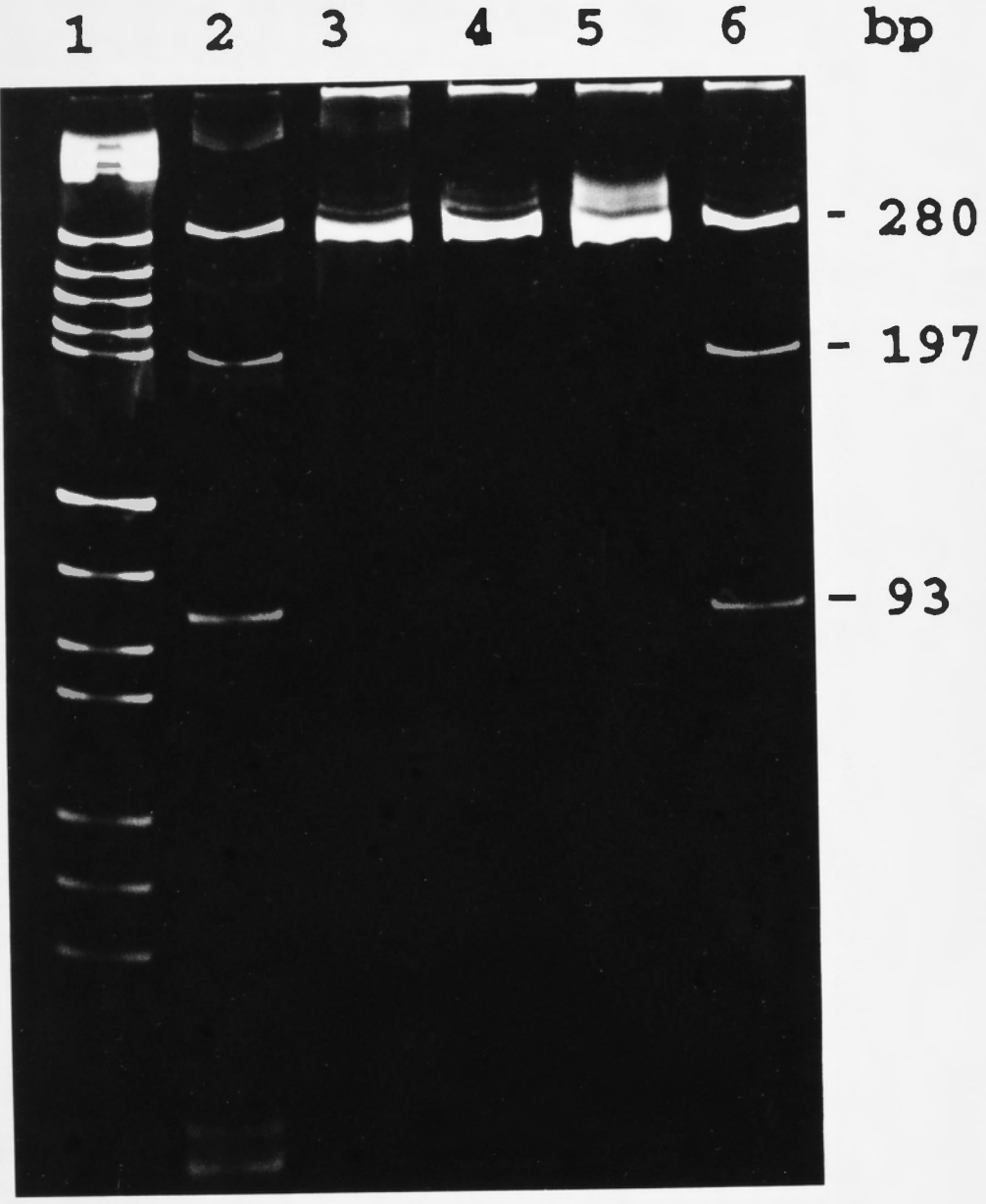
5.4.3.1 Exon 11-12, Aci I polymorphism

Of 44 human samples examined two showed this polymorphism in a homozygous form while five possesed the

Figure 5.1 Amplified restriction fragment length polymorphism typing of C6 exon 10, Bsm I site. **a** Polyacrylamide gel after electrophoresis of uncut 280 bp amplification product (lane 5) and Bsm I fragments from individuals heterozygous for the polymorphism (lanes 2 and 6) and individuals homozygous for the published exon 10 sequence (lanes 3 and 4). The fragment sizes, in bp of a standard DNA (Hae III digested pBR322) are shown in lane 1. **b** Position, sequence, amino acid, restriction site and pI values for the published and new exon 10 alleles.

C6 Exon 10 Bsm I Digest

a



b

Polymorphism at cDNA 1675, amino acid 518, exon 10.159

Phenotype;	Published	New Allele
Codon;	TGT	TGC
Amino acid;	Cys	Cys
Bsm I site;	- (GAATG'CN')	+ (GAATG'CN')
pI (+A, +B);	6.64, 6.71	6.64, 6.71

Table 5.2 Genotypes and allele frequencies of the C6 exon 10, Bsm I locus in Aboriginal Australians, Caucasian SLE patients and healthy individuals, Micronesians, Polynesians, and chimpanzees.

Group	N	Genotype Frequencies			Percent Allele Frequencies	
		DD	DC	CC	D	C
Aboriginal Aust.	20	9	1	0	95.0	5.0
Caucasians - SLE	30	14	1	0	96.7	3.3
Caucasians - healthy	30	15	0	0	100.0	0.0
Micronesians	28	11	3	0	89.3	10.7
Polynesians	22	10	1	0	95.5	4.5
chimpanzees	20	10	0	0	100.0	0.0

C = new exon 10 sequence, **D** = published exon 10 sequence.

N = number of chromosomes

Table 5.3 Genotypes and allele frequencies of the C6 exon 11-12, Aci I locus in Aboriginal Australians, Caucasian SLE patients and healthy individuals, Micronesians, Polynesians, and chimpanzees.

Group	N	Genotype Frequencies			Percent Allele Frequencies	
		LL	LS	SS	L	S
Caucasians - SLE	22	9	2	0	90.9	9.1
Caucasians - healthy	18	6	2	1	77.8	22.2
Micronesians	12	5	1	0	91.7	8.3
Polynesians	16	7	0	1	87.5	12.5
chimpanzees	12	6	0	0	100.0	0.0

S = new exon11-12 sequence, **L** = published exon 11-12 sequence

N = number of chromosomes

change in a heterozygous form. Six chimpanzees were examined and none showed this change. Gene frequencies were variable and probably reflect the inacuracy of small sample size. Interestingly eight human samples (not included in Table 5.3) showed a digest pattern different to the expected band sizes. The uncut PCR product was 619 bp and samples possessing the identified Aci I site were 591 bp, whereas these unusual samples showed only one band of approximately 560 bp (Figure 5.2). Repeated PCR and digest analysis of these samples did not change the different pattern. The undigested PCR product of these samples appeared to be 619 bp on examination prior to digestion, eliminating the possibility of a smaller initial PCR product due perhaps to a shorter intron 11 sequence. This result suggests there is an additional polymorphic Aci I site approximately 30 bp 3' from the identified polymorphic Aci I site at amino acid 544 or 30-60 bp in from the exon 12 end of the PCR product. The identified Aci I polymorphism causes an amino acid change that results in an overall charge change in the protein and yet these samples show no evidence of altered IEF gel patterns.

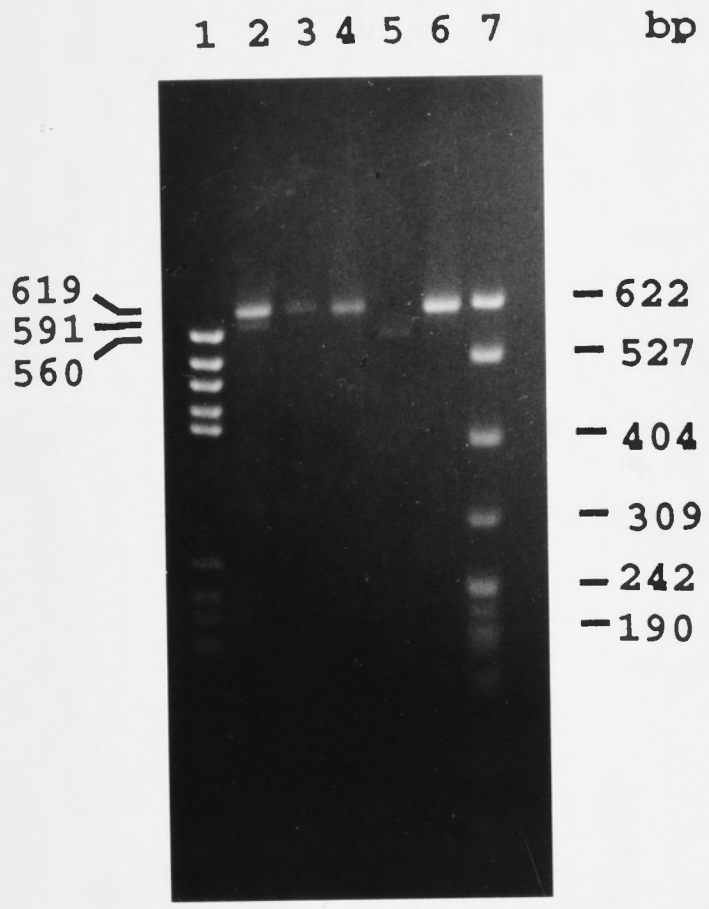
5.4.3.2 Bsr I polymorphism

The polymorphism at amino acid 546, altering a Bsr I restriction enzyme site was found in a homozygous form in two human samples and in the heterozygous form in 12 of a total of 45 samples examined. Gene frequencies for this polymorphism are given in Table 5.4. Sample numbers are small and as a result allele frequencies are quite variable. Only five chimpanzees were examined at this site and none showed this change. Interestingly seven samples examined showed a digest pattern completely different to that seen for homozygous normal, homozygous mutant and heterozygous samples. These samples were PCR amplified and digested several times to rule out the possibility of partial digests. This finding suggests a further polymorphism not found in the sequenced samples which also affects a Bsr I recognition site. It was difficult to identify the exact sizes of the new fragments

Figure 5.2 Amplified restriction fragment length polymorphism typing of C6 exon 11-12, Aci I site. **a** Polyacrylamide gel after electrophoresis of PCR amplified exon 11-12, Aci I digested fragments. Homozygous published samples (lanes 3, 4, 6), a heterozygous sample (lane 1) and a homozygous polymorphic allele sample (lane 5). The fragment sizes, in bp of a DNA standard (Hae III and Msp I digested pBR322) are shown in lanes 1 and 7 respectively. **b** Position, sequence, amino acid, restriction site and pI values for published and new alleles.

C6 Exon 11-12 Aci I Digest

a



b

Polymorphism at cDNA 1752, amino acid 544, exon 11.010

Phenotype;	Published	New Allele
Codon;	GAC	GGC
Amino acid;	Asp	Gly
Aci I site;	- (C'CG'C)	+ (C'CG'C)
pI (+A, +B);	6.64, 6.71	6.71, 6.77

Table 5.4 Genotypes and allele frequencies of the C6 exon 11-12, Bsr I locus in Aboriginal Australians, Caucasian SLE patients and healthy individuals, Micronesians, Polynesians, and chimpanzees.

Group	N	Genotype Frequencies			Percent Allele Frequencies	
		TT	TU	UU	T	U
Caucasians - SLE	24	8	4	0	83.3	16.7
Caucasians - healthy	20	4	4	2	60.0	40.0
Micronesians	10	3	2	0	80.0	20.0
Polynesians	18	7	2	0	88.9	11.1
chimpanzees	10	0	0	5	100.0	0.0

T = new exon 11-12 sequence, **U** = published exon 11-12 sequence

N = number of chromosomes

but estimates of 260 bp and 180 bp were taken (Figure 5.3). All of these unusual samples also possessed the unexplained Aci I pattern.

5.4.4 Exon 13

The polymorphism identified in this exon, at amino acid 680, was initially thought to contribute to the C6*B2 protein allele however the calculated pI for this change did not correspond to expectations based on IEF gel patterns. It was thus decided to examine this site in a number of other samples in order to identify whether the change was a polymorphism found in non-rare protein types or a possible cause of the B2 allele. Of 63 human samples examined two showed this polymorphism in homozygous form and 29 in the heterozygous form (Figure 5.4, Table 5.5). One of the two homozygous samples was the B2 (WP3) phenotyped sample from which the polymorphism was initially identified and the other was a Nauruan sample with normal protein type. Of the 29 heterozygous individuals identified 22 had common protein allele types. The seven heterozygous samples with rare protein types included four of the five rare-protein type SLE patients and represented three different rare alleles (A1, B2 (WP3, Nauru) and A?B21) which suggests this polymorphism is not responsible for any of them. It thus appears that this sequence change is a relatively common polymorphism which has no identifiable affect on IEF protein type. Allele frequencies vary considerably across the different groups examined, probably reflecting sample sizes. This polymorphism was found in a heterozygous form in three chimpanzees suggesting it is an ancient polymorphism.

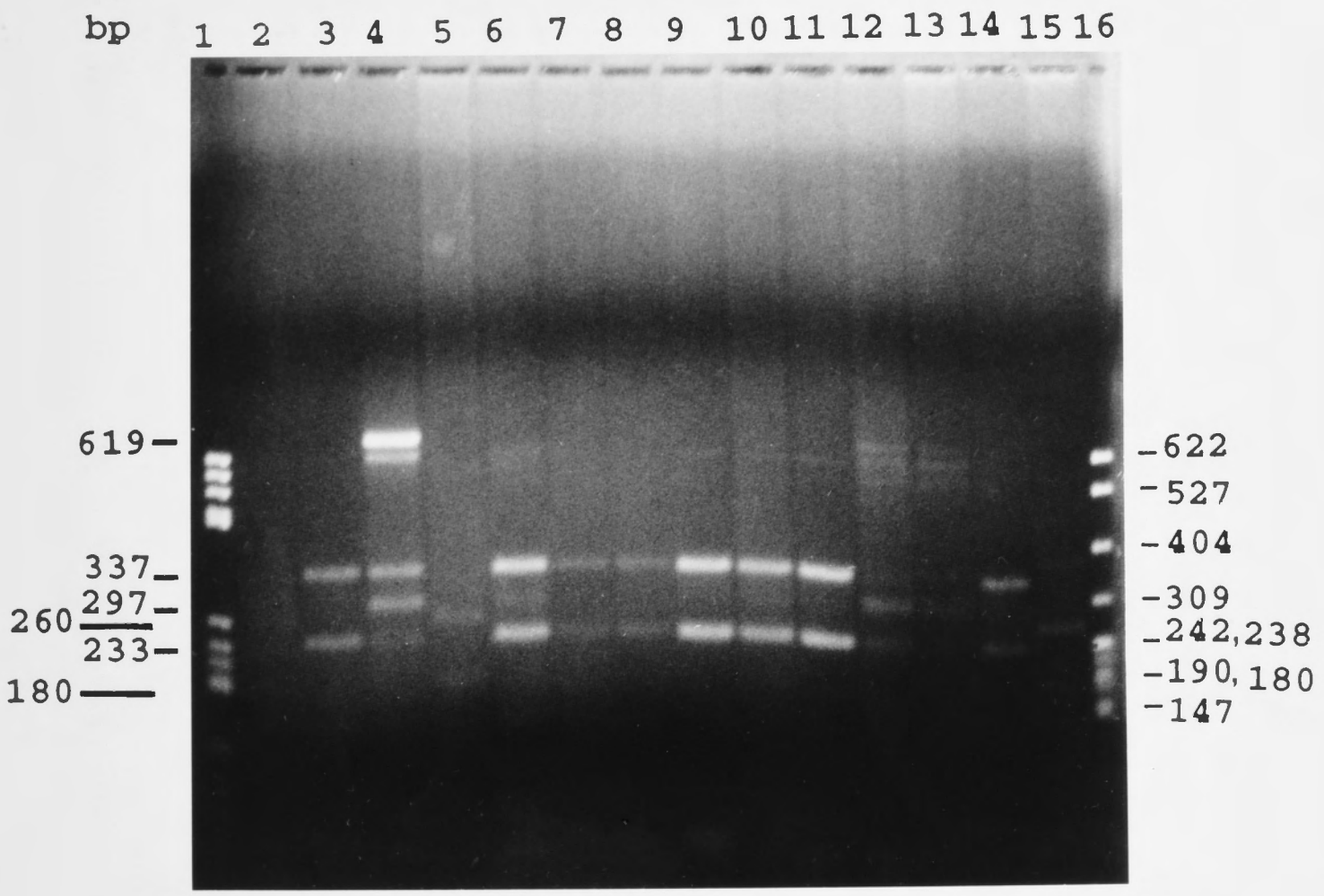
5.5 Discussion

This study of C6 DNA polymorphisms reinforces the known polymorphic nature of complement component six. The C6 protein is well known to be highly polymorphic with at least 19 different alleles identified with isoelectric focusing. The present study shows that the polymorphic nature of the C6 gene is even greater than suggested by protein studies. Of nine identified DNA polymorphisms

Figure 5.3 Amplified restriction fragment length polymorphism typing of exon 11-12, Bsr I site. **a** Polyacrylamide gel after electrophoresis of Bsr I digested fragments of C6 exon 11-12 PCR products. Individuals heterozygous for this polymorphism (lanes 4 and 6), samples homozygous published alleles (lanes 3, 7, 8, 9, 10, 11, 14), samples homozygous for the polymorphism (lanes 12, 13). Two samples with patterns different to all others, representing new alleles are in lanes 5 and 15. The fragment sizes, in bp of a DNA standard (Hae III and Msp I digested pBR322) are shown in lanes 1 and 16 respectively. **b** Position, sequence, amino acid, restriction site and pI values for the confirmed alleles of exon 11, identified with Bsr I.

C6 Exon 11-12 Bsr I Digest

a



b

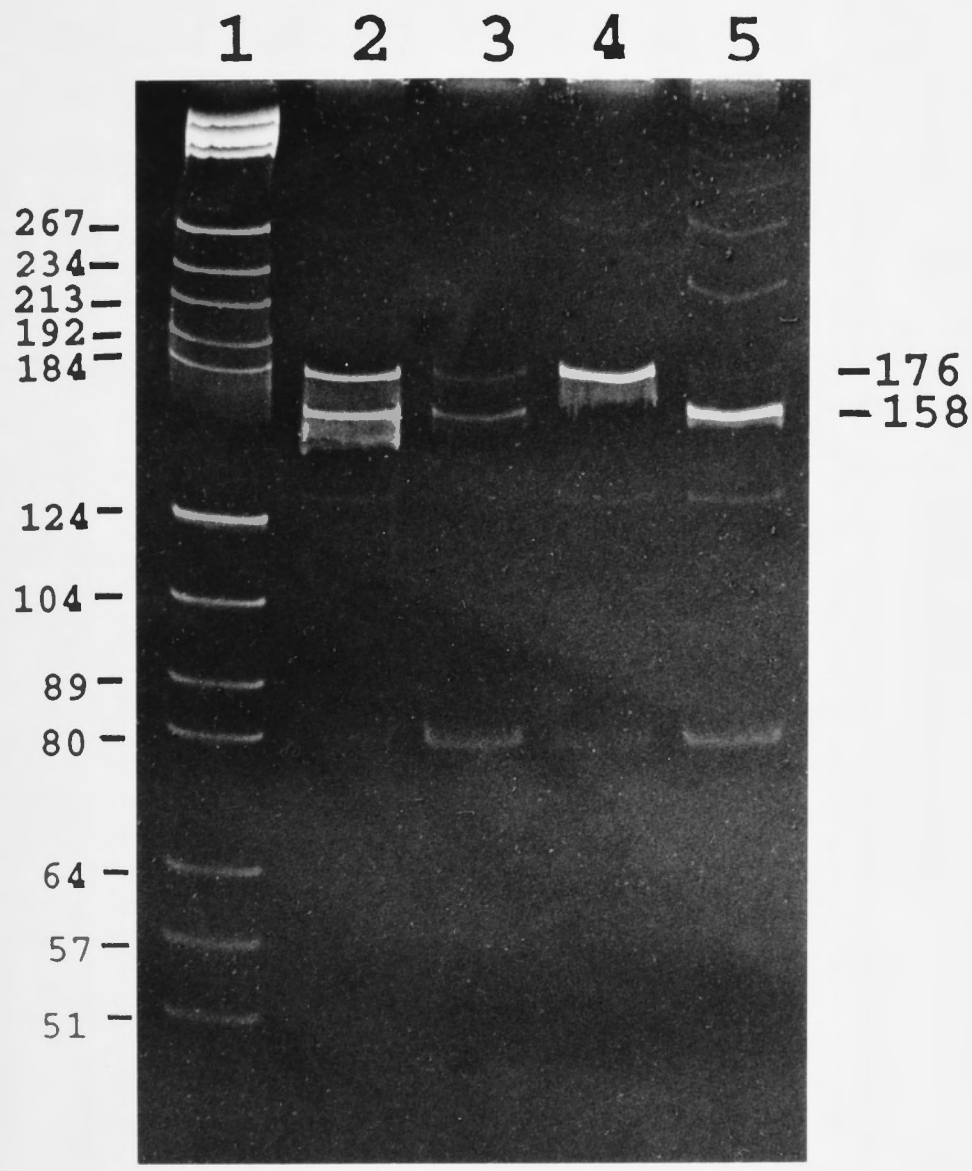
Polymorphism at cDNA 1759, amino acid 546, exon 11.017

Phenotype;	Published	New Allele
Codon;	CAG	CAC
Amino acid;	Gln	His
Bsr I site;	- (ACTG'GN')	+ (ACTG'GN')
pI (+A, +B);	6.64, 6.71	6.67, 6.73

Figure 5.4 Amplified restriction fragment length polymorphism typing of exon 13, Mae II site. **a** Polyacrylamide gel after electrophoresis of uncut 176 bp amplification product (lane 4) and Mae II fragments from samples heterozygous for the exon 13 polymorphism (lanes 2 and 3) and homozygous for the polymorphism (lane 5). The fragment sizes, in bp of a DNA standard (Hae III digested pBR322) are shown in lane 1. **b** Position, sequence, amino acid, restriction site and pI values for the published and new exon 13 alleles.

C6 Exon 13 Mae II Digest

a



b

Polymorphism at cDNA 2159, amino acid 680, exon 13.133

Phenotype;	Published	New Allele
Codon;	C|GG	T|GG
Amino acid;	Arg	Trp
Mae II site;	- (A'CGT)	+ (A'CGT)
pI (+A, +B);	6.64, 6.71	6.57, 6.64

Table 5.5 Genotypes and allele frequencies of the C6 exon 13, Mae II locus in Aboriginal Australians, Caucasian SLE patients and healthy individuals, Micronesians, Polynesians, and chimpanzees.

Group	N	Genotype Frequencies			Percent Allele Frequencies	
		XX	XY	YY	X	Y
Aboriginal Aust.	20	7	3	0	85.0	15.0
Caucasians - SLE	30	6	9	0	70.0	30.0
Caucasians - healthy	30	12	3	0	90.0	10.0
Micronesians	26	5	6	2	61.5	38.5
Polynesians	20	2	8	0	60.0	40.0
chimpanzees	20	7	3	0	85.0	15.0

X = new exon 13 sequence, **Y** = published exon 13 sequence

N = number of chromosomes

only two correlated with an altered protein type, leaving seven others that were not indicated from protein work. This suggests that the C6 gene is three or four times more polymorphic than the protein data indicates. This level of polymorphism in C6 could reflect a lack of natural selection for the gene, in that mutations are tolerated because a functional C6 protein is not essential for survival of the individual. Alternatively these DNA sequence changes may not affect the proteins function. A report by Davies *et al* (1993) describes a high level of polymorphism in the 5' untranslated region of the α-galactosidase A gene detected with SSCP analysis. Few polymorphisms for this gene had previously been reported. This suggests that more detailed examination of gene sequences in general may reveal higher levels of polymorphism than was previously thought and as such what is found in the C6 gene may not be unusual

The population studies were not aimed at accurately determining the frequency of the different base substitutions in the groups examined, for that larger numbers of samples would need to be examined. The purpose of this work was to indicate the presence of these polymorphisms in the different groups and to give an indication of allele frequencies.

Many population studies have examined C6 protein types in different ethnic groups (summary Chapter 3, Table 3.1). Allele frequencies for the common-protein types are well established and relatively similar in the major ethnic groups. Results in the present study show good agreement with published protein data for the Caucasian - SLE group, but other groups appear quite different. This disagreement is probably a reflection of the limited sample size in the present study. Except for the healthy Caucasians, who showed an equal frequency of C6*A and C6*B alleles, other groups were in agreement with published data in which protein allele was most comomon. No published data are available on protein allele frequencies of chimpanzees, but rhesus monkey studies (Hall and Alper

1977) show similar A and B allele frequencies to those of Caucasians. The chimpanzees examined in this study showed a predominance of A alleles, this however disagreed with similar studies on other samples carried out by Tuxworth, Bontrop, Fernie, Hobart, Lachmann and Mouse (personal communications from M Hobart) who found all 30 of their chimpanzee samples possessed the AB phenotype, a finding in conflict with Hardy-Weinberg expectations. Mitochondrial DNA studies of the chimpanzees used in the present study (personal communications C Wise) show remarkable diversity, ruling out the possibility of inbreeding in the group. Thus the predominance of A alleles in this study is likely to be a reflection of restricted sample size.

This study highlights the accuracy of replacing the protein test with a DNA based analysis. Of all common-protein typed samples examined none showed disagreement between the two tests. The DNA method also allows clarification of rare-protein samples as the background allele can be definitively identified. Additionally, it is possible that a DNA mutation could cause a common A allele to appear as a B allele with IEF and vice versa. While this occurrence is probably quite rare, based on the data obtained from this study, it effectively increases the number of possible polymorphisms that occur in the C6 gene.

The polymorphism identified in exon 10 was a silent substitution that has a gene frequency between three and 10 percent in humans. It is therefore a relatively common polymorphism found in at least four different ethnic groups with the polymorphism arising prior to emergence of these racial groups. None of the chimpanzees examined possessed this polymorphism. Of the 14 Micronesian samples examined three possessed this polymorphism on one of their chromosomes. This suggests an allele frequency of around 10 percent in this ethnic group. While this estimation is probably not precise it does suggest Micronesians are more polymorphic at this site than the other groups examined. This result is in agreement with

protein data for this group as Micronesians show a unique protein allele frequency distribution characterised by a higher frequency of rare-protein alleles (Ranford 1989). This finding also coroborates the HLA data which show that Micronesians have only limited similarites with Polynesians. Serjeantson and Hill (1989) suggest that Micronesians have had an independent source of HLA genes that are not found else-where in the Pacific, and therefore a unique distribution of C6 DNA polymorphisms is not unlikely.

The exon 11 Aci I site polymorphism was present on nine human chromosomes of a total of 72 examined, indicating it too is relatively common in humans. It was not found on any of the 12 chimpanzee chromosomes examined. From these data it is not possible to determine whether the negative finding in chimpanzees is due to the restricted sample size or to it being a lineage-specific polymorphism. This question could be answered with further studies utilising a large number of chimpanzees and perhaps other organsisms. The healthy Caucasians showed the greatest frequency of this polymorphism, around 22 percent, markedly contrasting with results for exon 10 where the estimated frequency for the new allele was zero. Micronesians did not show a higher allele frequency for this polymorphism in relation to other groups but this says little about the level of polymorphism in this ethnic group compared to others since the difference in occurrence was essentially a chance effect of finding homozygous versus heterozygous samples. It is worth noting also that the restriction enzyme Aci I is quite difficult to use and requires considerable care when estimating the amount of PCR product to digest. Additionally reaction conditions are non-standard, requiring the addition of one percent gelatin and an alternative buffer to that recommended by the manufacturer. It is possible that inaccuracies occurred since digestion may have been inefficient and identification of the 591 bp band was sometimes difficult.

The unusual 560 bp Aci I digest pattern was found in eight human samples and no chimpanzees. This finding may be the result of a recent mutation affecting an as yet undetermined Aci I site. The recognition sequence for Aci I is CCGC and many sites which differ from this motif by only one base are present in the exon 11-12 PCR product. It is possible that a point mutation occurs at one of these sites and therefore generates a new Aci I recognition sequence. A good candidate for this explanation occurs in exon 12 at amino acid 621-622, position 12.072. The published sequence at this site is CCGG but if it were CCGC in these samples, a new Aci I site would be generated and would give rise to a digestion product of 566 bp, assuming the samples did not possess the confirmed Aci I polymorphism. The likelihood of additional polymorphisms in the exon 11-12 region is not remote based on having found five other sequence changes in this area of the C6 gene. It is however highly unusual that all eight samples possess this "new" unconfirmed Aci I polymorphism in a homozygous form. Further studies would ideally involve cloning and sequencing a number of these unusual samples in this region in order to determine the underlying sequence of these exons and intron.

The exon 11 Bsr I polymorphism was found on 16 of the 45 examined human chromosomes. Two cases of homozygosity were identified with the remaining 12 alleles being found in a heterozygous form. Interestingly the homozygotes and four heterozygotes were found in the healthy Caucasians and the next highest "new" allele frequency occurred in the Caucasian SLE patients, suggesting that this polymorphism may be more prevalent in Caucasoids than Micronesians and Polynesians.

The exon 11 Bsr I polymorphism was not identified on any of the 10 chimpanzee chromosomes, and again this may be a result of small sample size or that it is a species specific polymorphism. A more extensive study could determine its presence or absence in chimpanzees.

Like the Aci I polymorphism, unusual Bsr I digest results were identified in seven samples, all of which also possessed the Aci I unusual pattern (the eigth sample was not tested). These digest results did not relate to any form of the identified polymorphism and thus represented a new sequence change. Two possible explanations for the observed band sizes are; 1) amino acid 546-as published; no Bsr I site, possible polymorphism at amino acid 590; new Bsr I site, polymorphism in intron 154 bp from exon 11, also creating a new Bsr I site, and unchanged Bsr I site at amino acid 626-627 as seen in published sequence, 2) amino acid 546-published sequence; no Bsr I site, possible polymorphism in intron 24 bp from exon 11 creating a new Bsr I site, possible polymorphism in intron 154 bp from exon 11, creating a new Bsr I site, and unchanged Bsr I site at amino acid 626-627 as seen in published sequence. Figure 5.5 is a schematic representation of the two possible explanations of the different patterns identified with the Bsr I digests, their approximate sizes and the polymorphisms needed to explain them. It was interesting that this new polymorphism occurred only in a homozygous form and always with the unconfirmed Aci I polymorphism. It never appeared with either a normal allele or with the previously identified sequence change at amino acid 546. It seems highly unlikely that two further sequence changes could occur in this region, however if option two were the correct explanation both these new mutations would be in the intron and therefore somewhat more likely than further exonic polymorphisms. Additionally in light of having found four polymorphisms in this exon already it is perhaps not impossible that this could occur here. It could be that the sequence changes responsible for the Bsr I and Aci I unconfirmed polymorphisms are not simple point mutations and may then occur at the same site. It is difficult to locate a site in this sequence which could be altered such that all these findings would be explained.

Figure 5.5 Schematic representation of the two possible explanations for the unusual Bsr I digest patterns of exon 11-12.

Option 1

Mutations required; 1) amino acid 590, 11.149 TGC → TGG, new Bsr I site created

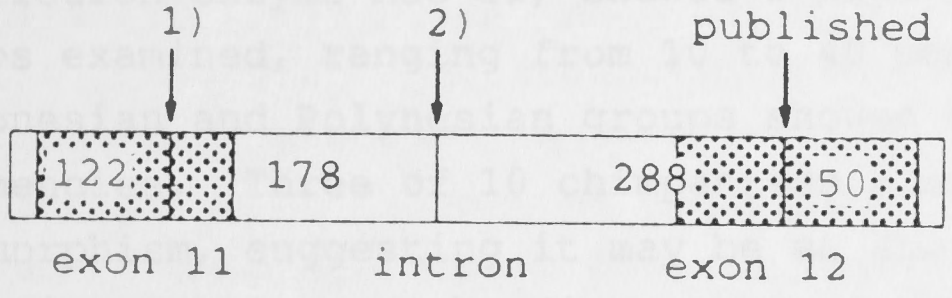
2) intron 11, 154 bp from the end of exon 11, TGC → TGG, new Bsr I site created

Option 2

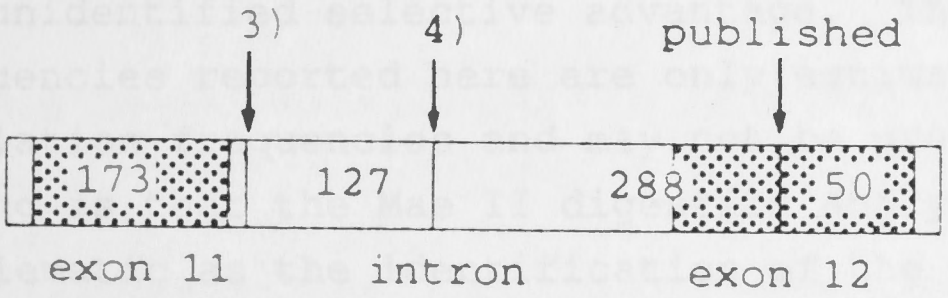
Mutations required; 3) intron 11, 24 bp from the end of exon 11, AAT → ACT, new Bsr I site created

4) intron 11, 154 bp from the end of exon 11, TGC → TGG, new Bsr I site created

Option 1;



Option 2;



The number of samples examined for the two exon 11-12 polymorphisms was much reduced relative to the exon three, 10 and 13 polymorphism studies. This was due to the difficulty in obtaining PCR product for many of these samples. The problem with amplification of this fragment arose because of its larger size relative to the exon three, 10 and 13 fragments and the quality of the template DNA. The age of some of the samples from which DNA was extracted for this study was greater then 10 years and many samples showed considerable degradation. This appeared not to be a problem with the smaller PCR products, but was clearly important when attempting to amplify the 619 bp fragments. Thus it was not possible to obtain a discrete band of PCR product for restriction digestion analysis for many of these samples.

The exon 13 polymorphism, identified with the restriction enzyme Mae II, showed a high frequency in all groups examined, ranging from 10 to 40 percent. The Micronesian and Polynesian groups showed almost equivalent frequencies. Three of 10 chimpanzees also possessed this polymorphism, suggesting it may be an ancient polymorphism established prior to the divergence of chimps and man, some four to six million years ago. Alternatively it may have arisen independently in the two groups due to some as yet unidentified selective advantage. The allele frequencies reported here are only estimates of the population frequencies and may not be precise. Also worth noting is that the Mae II digestion analysis was sometimes problematic as the identification of the published allele sequence and the new allele relied on the enzyme cutting to completion. In homozygous cases it was generally clear that complete digestion was obtained but in some heterozygous cases, several digestion attempts with increasing amounts of enzyme failed to produce bands which were of equal intensity. As such the heterozygote data had to be interpreted and it is possible that inaccuracies occurred.

The frequencies of the exon three and 13 polymorphisms were different in samples from four ethnic groups suggesting divergence between these populations. Differences in allele frequencies between ethnic groups may arise through variations in initial population size, the extent of migration and immigration, climatic or health crises and perhaps cultural factors such as the prevalence of consanguinous marriages. Further studies would utilise large samples to more precisely determine allele frequencies in these groups and thus allow better population comparisons.

From this study of C6 DNA plymorphisms it appears that the exon three, common-protein allele determining site and the exon 13 polymorphism may have arisen prior to divergence of humans and chimpanzees and these have since been maintained through selection. Alternatively it is possible that these polymorphisms arose through convergent evolution, evolving independently in the two species after divergence from a common ancestor. Both theories require selection to be acting on the gene, and as yet it is not clear what advantage either polymorphism would have on the individual. From this study neither theory can be proven and the issue of selection for the C6 gene will be discussed further below and in Chapter 6.

A total of five confirmed (Chapter 4) and two implied polymorphisms were found in the exon 11-12 region of the C6 gene. In proteins critical for survival a cluster of DNA polymorphisms such as this, could suggest that the DNA site does not encode a functionally important protein domain. An exception to this generalisation however is the HLA system, in which greater variability enhances survival as it allows a larger number of antigens to be recognised by the body and hence eliminated. In the case of the C6 protein however it is not clear whether increased variability has any beneficial or detrimental effect on the organism. It is true that many C6 deficient individuals are healthy implying that functional competence of C6 is not critical for survival, however

many deficient individuals experience recurrent infections. These infections may not be life threatening under present day health care, however at the time these polymorphisms were arising this may have been a significant factor. It is possible these polymorphisms conferred some advantage to the individual in the past and have not yet disappeared from the population. An explanation for the maintenance of these polymorphisms is difficult and is an issue which has been well debated. Fisher *et al* (1939) found that chimpanzees and humans shared an apparently neutral polymorphism for tasting PTC (phenyl-thiocarbamide) and this was found in the same frequency in both groups. The conclusion drawn from this was that there must have been strong balancing selection for this "apparently valueless" characteristic to maintain it for such a long time. Since Fishers study there have been many other cases of trans-species polymorphisms (Lawlor *et al* 1988, Mayer *et al* 1988, Gyllensten and Erlich 1989). Many examples of trans-species MHC polymorphisms have been examined at the molecular level and found to be identical suggesting a single molecular ancestor. Takahata and Nei (1990) report that these polymorphisms require selection to explain their existance, since the probability of observing this phenomenon for selectively neutral alleles is very small.

It was not determined in this study whether any of the identified polymorphisms affected protein function, and thus it is difficult to extrapolate about the location of these polymorphims in regard to functionality of the protein domains. This cluster of DNA polymorphisms does suggest analogy with the C4 diversity region and genetic mechanisms such as gene conversion and template mutagenesis may have been involved in C6 evolution. Chapter 6 discusses these theories in further detail

5.6 Conclusions

(1.) The exon three Dde I digest analysis is an accurate, alternative method of C6 protein typing the common C6*A and C6*B alleles.

(2.) The exon 10 polymorphism was found in all four ethnic groups examined, with Micronesians having the highest incidence of this change. The exon 11, Aci I and Bsr I polymorphisms were found in Micronesians, Polynesians and Caucasians, with healthy Caucasians having the highest incidence of both new alleles. The exon 13-Mae II polymorphism was found in all four ethnic groups and chimpanzees.

(3.) Restriction enzymes Aci I and Bsr I identified further sequence changes not found previously, the amount of polymorphism found in the exon 11 to 12 region shows this to be a highly variable region of the C6 gene.

Chapter 6

Chapter 6

General Discussion

This thesis has examined complement component six protein types and restriction fragment length polymorphisms (RFLP) in a group of Caucasian patients with SLE and healthy Caucasian individuals. No association between a specific protein type nor RFLP pattern and SLE was found. Samples from healthy Caucasians showed the same frequency of protein alleles and RFLP alleles as the Caucasian SLE patients. Allele frequencies for the protein and RFLP loci showed good agreement with Hardy-Weinberg expectations in both groups. This study was undertaken after a review of the literature showed there to be some association between SLE and C6 deficiency (Tedesco *et al* 1981, Trapp *et al* 1987). Of the 403 SLE patients examined in this study none appeared to be completely deficient in the C6 protein and the concordance with Hardy-Weinberg expectations showed that partial deficiency of C6 is not a major contributor to inherited predisposition to SLE. This suggests that the reported association between SLE and C6 deficiency may be an ascertainment artifact, a suggestion also made by Wurzner *et al* (1992a) and Sjoholm (1991).

Deficiency of C6 is one of the more common complement component deficiencies, occurring in aproximately one in 60 000 Caucasians (review Morgan and Walport 1991). In a review by Wurzner *et al* (1992a) C6 deficiency is reported to be found particularly in American Blacks (Petersen *et al* 1979), Caucasians and South African Cape coloureds (Orren *et al* 1987) with over 100 cases published. It is likely that patients with SLE are more frequently investigated for complement deficiencies than other patient groups and healthy individuals since associations between SLE and deficiency of C2 and C4 are well established (review Sjoholm 1991, review Densen 1991). As yet there have been few reported large scale studies of terminal complement component deficiencies in autoimmune

patients and healthy controls, thus it is difficult to conclude whether there is truly an increased incidence of C6 deficiency in SLE patients.

It remains unclear whether the terminal sequence affects earlier steps in complement activation. Lachmann et al (1978) report that the complement profile in deficient patients is in general normal for all the other components when individuals are healthy. In contrast Bhakdi *et al* (1988) report that the terminal complement complex exerts a regulatory effect on the formation of the classical and alternative pathway C3 convertases and on the utilisation of C5 by cell bound C5 convertases. Thus a negative feedback function of the MAC may contribute to regulation of the complement system. Individuals with terminal component deficiencies generally do not have impaired complement-dependent opsonisation, however there are exceptions (Boyer *et al* 1975). At present it is difficult to conclude whether absence of a functional MAC due to a deficiency, or non-functional, terminal complement component(s) has any wider effect than simply loss of MAC function. The vast majority of complement deficiencies essentially affect only a single component (Lachmann *et al* 1978). The cases of coexisting deficiencies of: C2 and C9 (Kaneco *et al* 1993); C5 and C4 (Gianella Borradori *et al* 1990); C6 and C2 (Delage *et al* 1979, Orren *et al* 1987) and C7 and C4 (Chapel *et al* 1987, Lopez-Trascasa *et al* 1988, Wuillemin *et al* 1991) are rare and do not suggest that absence of one component causes loss of another. These compound deficiencies appear to be the result of independent mutations (Wurzner *et al* 1992a).

A further explanation for the reported association involves the heterogenous nature of SLE. SLE is an extremely variable disease (Steinberg 1985) and it is becoming apparent that it consists of several subgroups (Arnett 1987, Jacob *et al* 1990). Fronek *et al* (1990) report that a clinical subset of SLE patients (those with nephritis) show an increase in the normally rare DR2 associated DQβ1*0601 allele, and a decrease in DR4

frequency, compared with SLE patients lacking kidney disease and controls. This suggests that there may be different genetic backgrounds and therefore causes of SLE in different subsets of patients. It is possible that a small subset of SLE patients do have associated deficiency of C6 and, given the variability in diagnosis of the disease, this group may have been under-represented in the present study.

This study also examined the association between C6 protein and Taq I RFLP alleles. No absolute correlation between protein and Taq I RFLP alleles was found demonstrating that C6 Taq I RFLP analysis cannot replace the IEF method of C6 allotyping. As a consequence of random genetic drift, selection, gene flow, non-random mating or common ancestry, alleles at separate loci may not associate at random (Lewontin 1988). This association is often thought to be a result of linkage, however non-linked genes may also temporarily or permanently be held in a non-random association. Many attempts have been made over the years, to measure this association, or disequilibrium. It is expected that for two randomly associated loci: A and B, the frequency of the different gametes or haplotypes will be equal. The standard equation to examine this property is;

D = (freq A1B1)(freq A2B2) - (freq A1B2)(freq A2B1) (1.)

where D = 0 implies no association (Lewontin and Kojima 1960). Thus D is often used as a measure of the degree of non-random association. The value of D in this equation is dictated by the allele frequencies and Lewontin (1988) believes this disqualifies D as a general measure of association. Numerous researchers have proposed measures to normalise D for allelic frequencies (Lewontin 1964, Hill 1975, Ohta 1980, Maruyama 1982). Hedrick (1987) examined six different measures of disequilibrium and showed them all to be dependent upon allelic frequencies despite four of them being normalised in some way. Lewontin (1988) concludes that there are no gene frequency independent measures of association, or disequilibrium,

between loci and this idea of association is a very poorly defined concept.

Despite the arguments above many researchers believe that measures of disequilbrium are valid. Numerous different methods for calculating this figure have been proposed and utilised (Weir and Wilson 1986, Haber 1984, Chakravarti *et al* 1984, Maiste and Weir 1992). In the present study three different methods were used and results compared.

In order to examine the possibility of an association between the protein allele determining loci and the Taq I RFLP site it was first necessary to determine which alleles were inherited together. The present study lacked any family data so determining which alleles were inherited together was problematic where samples showed double heterozygosity. In this situation haplotypes are not able to be determined and a problem then arises as to what to do with these samples. Some researchers advocate discarding them (Prof J Edwards, Oxford, M Hobart, Cambridge) which is thought to be reliable but inefficient, while others use the individual allele frequencies to calculate expected distribution of the haplotypes (Thompson *et al* 1988). Further studies utilise a method which examines both possible haplotypes for the double heterozygotes (Prof S Serjeantson, X Gao, Canberra). The present study used all three methods in order to compare the differences. No statistically significant difference was found between the three methods, however discarding the double heterozygotes gave somewhat different haplotype frequencies and the χ^2 value was higher for comparisons between this and the latter two methods.

The simplest method of calculating linkage disequilibrium, (1.) resulted in D values of 0.075, 0.035, and 0.038 (Chapter 3, Table 3.5) for the three haplotype distribution methods. The largest D value was for the method of discarding double heterozygotes and could be

regarded as evidence for a non-random association of these two DNA sites. A second method of calculating linkage was that of Baur and Danilovs (1980). This estimate was calculated using a computer program which incorporated double-counting of heterozygotes. Thus it was only possible to analyse the data with exclusion of double heterozygotes and double counting of heterozygotes. Results for both of these anlyses showed significant linkage disequilibrium for protein type A and Taq I RFLP allele 8.0 as well as for protein type B and Taq I RFLP allele 7.0. Therefore the two methods of dealing with double heterozygotes yield equivalent results for this method of linkage disequilibrium estimation. A further linkage estimation method employed log-linear calculations of the data (Weir and Wilson 1986). The computer program for this model discards double heterozygotes and the results showed a significant deviation from linkage equilibrium.

The present study employed three methods of dealing with double heterozygotes as well as three estimates of an association between two DNA sites. The overall conclusion of these analyses was that these two sites are not associated at random. However, it still remains debateable as to whether any of these estimates of association are truly valid, since some workers believe an allelic frequency-independent measurement is difficult, if not impossible to obtain (Lewontin 1988, Hedrick 1987).

It is thought that C6 is selectively neutral (Hobart *et al* 1975) since gene frequencies for the common alleles are very similar in different racial groups, alleles are always found in Hardy-Weinberg equilibrium and people deficient in this protein are often healthy. Neither the common C6 protein allele determining site (exon three) nor the Taq I RFLP site are thought to affect the function of the protein. Thus should selection for functional C6 protein be operating it would not affect these allele frequencies unless these sites were linked to a functionally critical site. Additionally, there is at

least a 60 kb distance between the two DNA sites and thus recombination events could occur between them. As yet no reports of recombination within the C6 gene or between the C6 and nearby C7 gene have appeared. However until recently (Fernie *et al* 1993) only two polymorphic DNA markers (Coto *et al* 1991a and b) for the C6 gene were available for study and few groups had specifically examined this aspect of the C6 gene. The possibility of recombination events and the apparent neutral selection for these two DNA sites would suggest they should associate at random.

Interestingly Setien *et al* (1993) report a significant association between another C6 RFLP, involving a polymorphic Msp I site (Coto *et al* 1991b), and the C6 Taq I RFLP. Fernie *et al* (in preparation) localise the polymorphic Msp I site to intron three, approximately three kb from exon three. This places the Msp I recognition sequence very close to the C6 protein allele determining site in exon three. Thus Setien and coworkers' (1993) results show equivalent findings to those in the present study, however the validity of measurements of association still remain controversial.

It is possible that the exon three common protein allele determining and Taq I RFLP sites are of very different ages. The C6 protein allelism is known in rhesus monkeys (Hall and Alper 1977), rabbits (Kunstman and Mauff 1980, Goldman *et al* 1982, Rother 1986), rats (Granadose *et al* 1984) and pigs (Shibata *et al* 1993) thus it appears to be an ancient polymorphism. No data are available for the Taq I RFLP in other organisms so it is impossible to conclude whether this polymorphism is recent or of great antiquity. It is possible that the Taq I RFLP site is a recent sequence change and as yet insufficient time has passed for random association of these alleles. The amount of time, or number of generations, needed to produce random assortment of alleles is determined by the rate of recombination and the initial level of disequilibrium between two non-linked sites. Bodmer and

Bodmer (1978) estimate that in a random mating population, if disequilibrium is present it would decline at a rate of (1-r) per generation, where r = the recombination fraction. Thus however close together the loci are, linkage disequilibrium, in the absence of selection, will eventually reach zero, but the rate at which this occurs depends on the recombination fraction. Given the available data for the C6 gene the rate of recombination cannot be accurately estimated but it appears that insufficient time has passed to produce a random association.

Further investigation of the isssue of selection for C6 suggests that non-selection for C6 is not as clear as was indicated above. Results in Chapter 5 show that the C6 protein determining site and a further DNA plymorphism occur in chimpanzees. Additionally, C6 protein allele frequencies similar to those found in Caucasians, are found in rhesus monkeys (Hall and Alper 1977). These findings suggest that some form of selection is either; 1. maintaining these polymorphisms in the populations over great lengths of time, or 2. driving the independent establishment of these polymorphisms in different species. Chimpanzees are thought to have diverged from humans four to six million years ago while rhesus monkeys diverged around 26 million years ago (Bulmer 1991). For polymorphisms to be preserved over such a length of time some selective advantage must arise from them (Fisher et al 1939, Takahata and Nei 1990, Golding 1992). Similarly, for several different species to establish identical polymorphisms independently, some form of selection must be operating.

A review by Wurzner *et al* (1992a) includes a good discussion of possible beneficial effects of terminal complement deficiencies. It is postulated that terminal complement deficiencies may be beneficial in cases of autoimmune disease. The inability to generate the membrane attack complex, and therefore produce cell lysis through this pathway, may prevent destruction of many autologous cells. Wurzner *et al* (1992a) also suggests that sublytic functions of the MAC may play a role in disease pathogenesis, the example given is the affect of MAC-induced release of toxic reactive oxygen metabolites

in the synovial fluid of rheumatoid arthritis patients (Morgan *et al* 1988). Deficiency of any one of the terminal components would prevent the induction of these inflammatory mediators and hence limit the severity of disease.

Individuals with a deficiency in terminal complement components appear to have an increased susceptibility to meningococcal disease (review Wurzner *et al* 1992a), however there is some evidence that terminal component deficient individuals suffer a less severe form of the infection compared with complement sufficient persons (Ross and Densen 1984). The mechanism for this proposed outcome is through a lower degree of tissue damage induced by activation of complement by endotoxin. Further studies (Lehner *et al* 1992) demonstrate that *E. coli* endotoxin release is induced in complement sufficient but not C6 deficient individuals. Interesting findings are also reported for C6 deficiency in the Cape region of South Africa (Orren *et al* 1987). In this population a higher than expected proportion of homozygous C6 deficient individuals was found in families that had experienced recurrent meningococcal infection. The suggested explanation for this is a selective advantage for terminal component deficiency in this population (Orren *et al* 1987).

Studies in rats and rabbits suggest that C6 deficiency may have a protective role in experimental autoimmune myasthenia gravis (Biesecker and Gomez 1989), Heymann nephritis, an experimental counterpart to human membranous glomerulonephritis (Groggel *et al* 1983) and experimental autoimmune thyroiditis (Inoue *et al* 1993).

Thus it appears that some form of selection for C6 deficiency may exist. C6 deficiency is not thought to have an absolute advantage, however there may be some form of heterozygote advantage under some environmental conditions. It is possible to postulate then that some selective force may operate on C6 polymorphisms that alter

protein phenotype. Perhaps particular changes in protein sequence confer some, as yet undetected, advantage to the individual. It is conceivable that DNA polymorphisms resulting in amino acid changes alter the binding efficiency of C6 to C5b or C7. The cluster of polymorphisms, located between exons 10 and 13, found in the present study, are very near the C5b binding domain of the C6 protein. Nakano *et al* (1991) localise the C5b binding domain to the latter two SCRs, encoded for by exons 12 to 14 (Hobart *et al* 1993a). Therefore it is possible that these DNA changes result in altered conformation of the protein in this region consequently changing the C6-C5b binding efficiency. A lower binding efficiency may be an advantage in some situations, for example autoimmune diseases, as this may reduce tissue damage by the MAC. Alternatively these polymorphisms might increase the efficiency of binding and hence facilitate MAC lysis. Under varying environmental conditions it may be advantageous to carry these polymorphisms.

This possible selective advantage may not be presently as strong as in the past, possibly because of better health-care facilities, hygiene and diet. However the past existence of such selection could explain both the trans-species polymorphisms of Chapter 5 as well as the linkage disequilibrium suggested for the protein determining and Taq I RFLP sites.

Screening the 2802 bp coding region of C6 for polymorphisms identified eight sequence changes while a further sequence change was found in non-coding DNA. In total nine DNA sequences changes were identified and two further changes were implied. These 11 changes occurred over a region of 3820 bp, inferring a nucleotide divergence rate of 0.29 percent. Interestingly these changes were not randomly distributed throughout the gene, but were distinctly clustered. Four of these changes occurred in exon 11 and nine of the 11 polymorphisms found were located between exons 10 and 13. Of the eight

sequence changes identified in the C6 coding region only two were silent, with the remaining six resulting in both amino acid and charge changes in the protein. Protein data has shown C6 to be highly polymorphic with a large number of alleles distinguished by their charge. Only two of the detected sequence changes found in the present study related to a protein phenotype, indicating that C6 protein polymorphisms clearly underestimate DNA sequence changes.

Effects of amino acid changes on protein conformation are not well understood and predictions about the result of polymorphisms are not simple. From the present study it appears that some amino acid changes, which also alter the charge of the overall protein molecule, do not greatly affect the conformation and thus migration of a protein through a gel matrix. This study identified four amino acid and charge-altering changes which were not evident from isoelectric focusing. This would seem unusual, however the location of these changes may be more important for conformation of the protein then the charge-change alone.

Li and Sadler (1991) have estimated the amount of nucleotide diversity in humans from studies of published cDNA and genomic sequences that have been carefully checked for sequencing accuracy. The measure of variability was defined as the number of nucleotide differences between two sequences of one gene. In over 75 kb, from 49 genes, a nucleotide diversity ranging from 0.03 to 0.11 percent was found. Li and Saddlers (1991) study also demonstrated that the nucleotide differences were not distributed evenly throughout the gene. Most of the sequence pairs were identical whereas a number of base pairs showed multiple differences. They conclude that this non-random distribution could be partly due to 1) mutation rate variation, 2) sequence errors in the data, 3) differences in sequence length and 4) statistical fluctuations. While this estimate of nucleotide diversity is useful it may not be truly accurate because the genes

examined are perhaps not a random sample since they were sequenced for their biological or medical importance.

The level of polymorphism found in the C6 gene is high, approximately 0.29 percent, compared with Li and Sadlers' (1991) estimate and can not be easily explained by functional advantages. MHC genes show extraordinary levels of polymorphism (review Klein 1986) which many researchers believe is convincingly accounted for by the function of the protein (Parham *et al* 1989). The MHC molecules are peptide-binding molecules which trap a diversity of peptides derived from the degradation of various proteins (Babbitt *et al* 1985, Bjorkman *et al* 1987a and b). An ability to bind a large number of peptides is seen as a selective advantage for the individual, thus the level of polymorphism seen in these genes and its advantage seem clear (Doherty *et al* 1975, Parnham *et al* 1989).

Unlike theories for the MHC, the function of C6 does not easily explain this level of polymorphism because as yet there appears to be no obvious functional advantage to possessing such diversity in DNA sequence. This study presents no evidence to show whether these changes affect the function of the C6 protein and therefore it is difficult to give a reason for their occurrence, or to explain the maintenance of this diversity. It is theoretically possible that the chance effect of random drift during hominid evolution could account for the C6 polymorphisms.

A possible explanation for the amount and clustering of DNA polymorphisms found in this gene is through gene conversion. High levels and non-random distribution of polymorphisms are sometimes considered evidence for such an event (Stephens 1985, Sawyer 1989, Satta 1992). Gene conversion is a controversial topic and its existence in mammalian systems is not universally accepted. It was first recognised as a genetic mechanism in fungi (Lingren 1949) and has since been demonstrated in a number of

eukaryotes. The evidence for this phenomenon has been in abberrant segregation of meiotic products. Instances in which one allele apparently increases in copy number while another decreases, have been observed in *Drosophila*, *Saccharomyces* and *Sordania* (Kitani 1962, Fogel and Mortimer 1969, review Wysocki and Gefter 1989). The mechanism of gene conversion has been the subject of much attention (review Radding 1978). One explanation for gene conversion is that it is the result of DNA repair following intragenic crossing-over (Holliday 1964). Crossing over produces DNA heteroduplexes, these are molecules that contain one strand from one chromatid and the other from the homologous chromatid. As the site of exchange migrates along the gene, it reaches the position where the two alleles differ in their sequence, leaving the non-complementary nucleotides unpaired. The unpaired region is then recognised by proof-reading enzymes which remove one of the nucleotides and replace it with a nucleotide complementary to that on the other strand. Since the nucleotide that is excised is chosen at random, different ratios of wild-type to mutant alleles result (Klein 1986).

The term gene conversion has been used for genetic exchanges in higher eukaryotic organisms in which formation of heteroduplex, or hybrid, DNA is thought to be involved. Strictly, the term should only refer to nonreciprocal exchanges of homologous genetic information. In mammals the four products of meiosis do not stay together and therefore the relationship of the haploid gametes to the original diploid germ cell can not be determined. For this reason,in mammals, it is not possible to identify whether meiotic recombination is reciprocal or nonreciprocal. In mammals the term is often used to describe incompletely-defined genetic exchanges.

The first evidence that gene conversion occurred in mammals was in the fetal globin genes (Slightom *et al* 1980). The conclusion for occurrence of this mechanism was based on two equivalent introns in genes on the same

chromosome having a segment more similar to each other than to their allelic counterpart. Since this first report of gene conversion in mammals many examples of DNA transfer between members of mammalian multigene families have been reported (Mellor *et al* 1983, Pease *et al* 1983, Rechavi *et al* 1983, Gorski and Mach 1986, Liu *et al* 1987, Gally and Edelman 1992).

Findings which are thought to suggest gene conversion are variable, ranging from almost identical segments of DNA between two homologous genes (Gally and Edelman 1992) to high levels and nonuniform distribution of polymorphisms in a particular gene (Stephens 1985, Sawyer 1989, Satta 1992). Evidence in mammals is thus far circumstantial and therefore cannot be conclusively established.

Gene conversion is often associated with duplicated genes, for example C4 (Partanen and Campbell 1989, Braun *et al* 1990) and 21hydroxylase (Urabe *et al* 1990, Kinoshita *et al* 1991, Collier *et al* 1993) as well as multigene families (Griffiths and Watterson 1990, Moore *et al* 1992). Fernie *et al* (in preparation) suggest that the C6 gene arose through early duplication, and there has since been a collapse of the duplicated system. This hypothesis has been proposed as an explanation for the distribution and maintenance of the C6 protein allelism. Evidence for such an event is difficult to establish, however some of the present findings in the C6 gene could be evidence for this. The strongest finding suggestive of this possibility is in the identification of three different bases at one site in exon 11 of one SLE patient. Athough this finding may be a result of PCR error or contamination, the possibility that it is real still exists. If proven real this finding could represent evidence for duplication of this exon, possibly as part of an incomplete collapse of the duplicated system in this individual. Additionally it may be that the entire gene is duplicated and the two genes are too similar to be easily distinguished. This patient also possessed three

proven polymorphisms in exon 11. As discussed above this level of sequence variation over 172 bp is surprisingly high but would appear less unusual if this exon was duplicated. Unfortunately no further time, or DNA from this patient were available and thus further studies could not be pursued.

Additional evidence for gene duplication comes from chimpanzee studies by M Hobart (personal communications). In that study 30 chimpanzee samples were examined at the protein determining site in exon three, and all 30 showed an AB genotype. This result is not in agreement with Hardy-Weinberg expectations and therefore requires explanation. It seems unlikely that 30 samples could have been analysed incorrectly, however contamination of the PCR primers with an AB DNA sample is a simple possibility. Alternatively these findings may indeed be real and gene duplication could explain the result. However further evidence is needed before this hypothesis could be accepted.

The components of the membrane attack system (C6, C7, C8 and C9) are thought to have a common ancestor (DiScipio and Hugli 1989, Setien *et al* 1993, Hobart *et al* 1993a). This hypothesis is based on the similarites at both the protein and DNA level. The similarites in amino acid sequence are shown in Table 6.1 and nucleotide sequence comparisons are in Table 6.2. Examination of the genomic organisation of these components presents further evidence for their common ancestry as the intron/exon boundaries are remarkably well conserved across all components (Hobart *et al* 1993a, Kaufmann *et al* 1993). Comparison of the amino acid sequences for these proteins shows that the region of amino acid sequence which aligns with the C6 exon 11 is more highly conserved across all of the proteins than any other single region. This could indicate that exon 11 has been involved in gene conversion or duplication events among the MAC components. Conversion events and duplications can facilitate the production of polymorphisms as mismatching and unequal

Table 6.1 Amino acid sequence identity of the membrane attack complex proteins. Regions of amino acid sequence align with C6 amino acids encoded for by C6 exons.

	% Identity[a] for Amino acids Region[b]															
Component Comparisons	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**
C6-C7	36	44	32	28	5	36	17	27	45	**54**	22	31	24	7	8	24
C6-C8A	22	20	49	28	6	32	2	40	37	**56**	-	-	-	-	-	-
C6-C8B	24	27	40	25	13	33	7	31	33	**47**	-	-	-	-	-	-
C6-C9	18	33	45	28	16	30	7	20	29	**10**[c]	-	-	-	-	-	-

[a] % identity calculated using the method of Myers and Miller (1988) and PC/Gene package

[b] Using amino acid sequence alignments from Haefliger *et al* (1989)

[c] Only nine amino acids of C9 correspond to this region of the C6 gene.

	Total Amino Acid Sequence		
	This study	Haefliger *et al* (1989)	
	% Identity	% Identity	% Conserved
C6-C7	27.5	33.5	56.1
C6-C8A	28.4	30.1	46.2
C6-C8B	25.6	28.4	48.3
C6-C9	21.3	26.6	47.2

Table 6.2 Nucleotide sequence identity for the membrane attack complex proteins. Regions refer to sections of DNA sequence that align with C6 exons.

Component	% Identity[a] Regions 6	7	8	9	10	11	12	13	Total DNA sequence
C6-C7	59.4	55.2	50.5	49.7	57.3	**61.6**	55.7	52.3	54.7
C6-C8A	50.0	54.4	52.9	63.3	54.2	**59.5**	-	-	63.9
C6-C8B	49.5	57.7	48.0	55.6	55.5	**56.5**	-	-	69.1
C6-C9	51.3	47.2	53.3	55.5	55.5	**67.9**	-	-	61.8

[a] % Identity calculated using the method of Myers and Miller (1988) and the computer package PC/Gene

crossing over events can occur more easily (Petes and Fink 1982). Additionally these findings could suggest that this region of the protein is critical for the common function, or association, of these components. In contrast regions six and eight show much lower levels of identity across the different proteins. This may suggest a lack of duplication/conversion events in these regions and possibly a less critical functional role, or a role specific for each component and therefore not conserved across these proteins.

Comparison of the nucleotide sequences of the genes for these components shows a similar trend for region 11, although this is not as marked as for the amino acid sequence. The percentage identity for region 11 is greater then the percentage identity for the total DNA sequence for C6-C7 and C6-C9 only, indicating that the DNA sequence for this region is not as highly conserved as is the amino acid sequence. This analysis shows that the conservation of the C6 protein and gene is not uniform throughout and suggests that the well conserved areas are functionally important regions for interaction of the proteins. Additionally the similarites of specific areas could suggest that they arose from a common ancestor, through duplication or conversion.

The non-random distribution of polymorphisms in the C6 gene can also be interpreted as evidence for mutation generating mechanisms. Differential rates of mutation across a gene could result in clusters of polymorphisms. Wolfe *et al* (1989) provide evidence for significant variation in mutation rate among regions of the mammalian genome. They show that the rate of some substitutions varies among genes and is correlated with the base composition of genes and their flanking DNA. Also proposed is that these differences in mutation rate arise because mutation patterns vary with the timing of replication of different regions in the germline.

DNA mechanisms involving templated mutagenesis also give rise to non-random distribution of polymorphisms. Ripley (1982) suggests two models for this mechanism both of which involve quasi-palindromic sequences. The basic idea is that palindromic sequences in double-stranded DNA molecules have an inherent property of self-complementarity within each single-strand of DNA. This complementarity permits these sequences to form uniquely paired DNA structures. A classical example of self-complementarity is the formation of hairpin structures. In quasi-palindromic sequences where this complementarity is imperfect, the formation of unusual DNA structures is still possible. The imperfect base-pairing in these structures provides the potential to generate mutations in a templated and therefore predictable manner.

Examination of the DNA sequences surrounding the polymorphisms found in exon 11 of the C6 gene showed there to be a short stretch of palindromic sequence (GCAGTAGACG) around the site for two identified mutations (cDNA; 1752 and 1753, amino acid 544). Ripley (1982) does not indicate a minimal number of bases required for the palindrome, thus it is difficult to assess whether what is seen in the C6 gene is sufficient for this mutational mechanism to occur. Several other DNA sites found to be polymorphic were also examined but no further palindromic sequences were identified. However, several sites (cDNA 1845, Mnl I site and cDNA 1675, Bsm I site) appeared to be located within regions of repetitive sequence suggesting mutations could arise through slippage of paired DNA sequences. It is possible that specific regions of a gene are more likely to contain polymorphisms because of the repetitive or palindromic nature of the surrounding DNA sequence.

Although it has not been proven that any of these mutation-creating DNA mechanisms occur in the C6 gene it is evident that this non-random distribution of DNA polymorphisms is a phenomenon seen in other genes within

the human genome. The significance and mechanisms for this occurrence however, remain to be established.

The present study of the complement component six gene highlights the value of molecular analysis in studies of human genes. Fine level examination of DNA sequences is a very useful tool for understanding the structure-function relationship of proteins, identification of mutations resulting in human diseases as well as studies of human population groups and evolution. The information gained from this study, in conjunction with broader functional analysis, significantly contributes to the understanding of this gene.

Appendix

Appendix

Nucleotide and amino acid sequence of the complement component six gene

The numbering of the cDNA and amino acids corresponds to the numbering system of DiScipio and Hugli (1989) in which the first amino acid (Cysteine) after the 21 amino acid signal peptide sequence is designated 1.

Intron/exon boundary information is from Hobart *et al* (1993a) and intron sequence data is as yet incomplete and is from MJ Hobart (Cambridge, UK).

Key;

Known intron sequence = ACTGTCTGA

Exon sequence = ACTGTCATGC

Amino acid = MetAlaArg

Amino acid numbering is above the line and cDNA sequence numbering is at the end of the line

* = polymorphic sites

Primer position = *ACGTCTGAGTG*

C6 EXON 1

```
TTGA-TTGCCTTAAGGAAGGCAGAAAAA-GCAAATATTTTCTAAAATCACATTTGAATCT
CTGACTCAGGATGACTTGTGAAGGTTTACCTCAAAAAAGCATTTGATAT--GTTATGTAT

                                  -21              -15

                  *               MetAlaArgArgSerValLeuTyrPhe

TCTTTCATTTTAGGGCCTGGAGGCTCTCAAGGCATGGCCAGACGCTCTGTCTTGTACTTC   84

     -10          -5              1              5

IleLeuLeuAsnAlaLeuIleAsnLysGlyGlnAlaCysPheCysAspHisTyrAlaTrp

ATCCTGCTGAATGCTCTGATCAACAAGGGCCAAGCCTGCTTCTGTGATCACTATGCATGG   144

   10             15             20             25

TyrGlnTrpThrSerCysSerLysThrCysAsnSerGlyThrGlnSerArgHisAr(g)

ACTCAGTGGACCAGCTGCTCAAAAACTTGCAATTCTGGAACCCAGAGCAGACACAGGTGG   200

GTGTGAGCTTTGTGGCTTTTCTTTGTGTCCTGGAAGGACTCAAGGATGAAGTCAGCAACA
TCAGGAGTGCAACACTAAGCAATGGGAACATATATATATCAAAGTAAGAACTTACAAATT
TGGATGAAATTTGAAGTCTCTCTCTTATGAGGTTGAAAGGAA-TCTA-GTTTGTTCTTTC
TACAAGAACGAGGGGTTTCAAAGAAA-TTAATTAAGAA-CTG-GCAACACTTTCAGTTGA
A-ATTTAA-T-GCTTAGAA-CACTACCCAAGCTTTTTTCAA-CAGCATACAA--T-GCTT
GCCTATAGTTAGTGAGTTAATTATAGAGAA--TTCTACTCA-TAAA-CGAGTCTAACAGG
ACCTCAG
```

C6 EXON 2

```
C-GAAAATTTTCTTTCTTTTCTATTTATTGTAGTTAATGCTCAATGATACTTTGATTATG

                                                          28

                                                      (Ar)gGln

GAGCAGGATATATGGTGAATACACTCATTCTAAGCAACTTCTTCTTCATGGTTACAGACA   203

     30             35             40             45

 IleValValAspLysTyrTyrGlnGluAsnPheCysGluGlnIleCysSerLysGlnGlu

AATAGTAGTAGATAAGTACTACCAGGAAAACTTTTGTGAACAGATTTGCAGCAAGCAGGA   263

     50             55             60             65

 ThrArgGluCysAsnTrpGlnArgCysProIleAsnCysLeuLeuGlyAspPheGlyPro

GACTAGAGAATGTAACTGGCAAAGATGCCCCATCAACTGCCTCCTGGGAGATTTTGGACC   323

     70             75          79

 TrpSerAspCysAspPheCysIleGluLysGln

ATGGTCAGACTGTGACCCTTGTATTGAAAAACAGGTAGGCAACCATGGGCA-TTCCAGGA   358
ATATGAATCACAATAGAGGGCTCTGA-TCCCTGACAGGCTTAACAGCATAGTAATGCTG
AATTACTTCAA-TTCTGAAAATCAACTGTCTCAGCTTCCATTGAGGGAAAATTCCA-TTA
GCAC--TTGTGGGGGGAAGAGGGGGTTTA--TTATGTATTTTACTGTCTTA
```

C6 EXON 3

CATTAGTAAAAA-A-AAGAAATATTACCATAGGAAAAATCAC-GTTGTGATTGAGGAGCC
CAGTTTTCTTAGGTACA-TTTGACCTTGACCCTGCCTCAGAGTTA*ATATCTTTCTCTTCTT*

80 85 90 95

SerLysValArgSerValLeuArgPheSerGlnPheGlyGlyGlnPheCys

*TCC*TTTCAG<u>TCTAAAGTTAGATCTGTCTTGCGTCCCAGTCA-TTTGGGGGACAGCCATGC</u> 408

* 100 105 110 115

ThrGluProLeuValAlaPheGlnProCysIleProSerLysLeuCysLysIleGluGlu

<u>ACTGAGCCTCTGGTAGCCTTTCAACCATGCATTCCATCTAAGCTCTGCAAAATTGAAGAG</u> 469

120 125 128

AlaAspCysLysAsnLysPheArgCysAspSerG(ly)

<u>GCTGACTGCAAGAATAAATTTCGCTGTGACAGTG</u>GTAATGTATTTT*TGTGAAATGTTCAG* 504
*TCCCC*ACTAAAAATAGCTTGACTAAATCAAAGTT-CTCCAATTGACTAAC-GTAGAA

C6 EXON 4

TTTACGCTCCAGTTTTTGTGTGCTCCCTAAAACTCTATCAAAAACAGTGAAGACTTGCCT
TCAACTTTACCTTGAAATTTAAAGCATTTGAGTTTCAAATTAATAAATACTTTCTAAGAT

130

(ly)ArgCysIleAlaArgLys

GATAATTTGTTGCATTGATTTTGTTTGCTTTTCTTGACTAG<u>*GCCGCTGCATTGCCAGAAA*</u> 523

135 140 145 150

LeuGluCysAsnGlyGluAsnAspCysGlyAspAsnSerAspGluArgAspCysGlyArg

<u>*GT*TAGAATGCAATGGAGAAAATGACTGTGGAGACAATTCAGATGAAAGGGACTGTGGGAG</u> 581

155 160 165 170

ThrLysAlaValCysThrArgLysTyrAsnProIleProSerValGlnLeuMetGlyAsn

<u>GAC-A-GGCAGTATGCACACGGAAGTATAATCCCATCCCTAGTGTA*CAGTTGATGGGCAA*</u> 641

175

Gl(y)

<u>*TGGG*</u>TATGTAACATCTTTTTATCATCTTGGGGAGAACAGGTATCAGAACAAATGAGTCAG 644
AGGGAGC--TGAG-TAC-TCATTGTCCATCTTCCTCATTAAA

C6 EXON 5

```
AGTAATGGAATAGGAGACCTAATAGAGCCAGAGAAGCTGATTTGGC-TTCCAGCTTCCTT
GCATGCCTCACCATGGACTAGGAAAAAAAAAAGCGGAGG-GTGGGGGA-TCCTTCAAGTCA
TT-GTCTTGAGGCAAAAAGGTTTAGAAGAGGA--TCCTTTCATATAA-TAAGGTAAATTT
```

```
                                                          176
                                                  (Gl)yPheHisPhe
TAAAGATTGTTCTACatctgttgaattcctgAAAATTACTTTGTTTCTAGGTTTCATTTT   654
   180             185             190             195
LeuAlaGlyGluProArgGlyGluValLeuAspAsnSerPheThrGlyGlyIleCysLys
CTGGCAGGAGAGCCCAGAGGAGAAGTCCTTGATAACTCTTTCACTGGAGGAATATGTAAA   714

   200             205             210             215
ThrValLysSerSerArgThrSerAsnProTyrArgValProAlaAsnLeuGluAsnVal
ACTGTCAAAAGCAGTAGGACAAGTAATCCATACCGTGTTCCGGCCAATCTGGAAAATGTC   774
   220
GlyPheGlu
GGCTTTGAGgtatgacagcctagcatggTGGCAACTCCAACACCGTCAGTGATTATAAGC   783
TAAATTTCT
```

C6 EXON 6

```
CAGCGGCCGCCG-TCAGCCCGCCCGCCACCGTCAGCACCACTGC-TCAA--TCACAGAC-
TGCTAAAAATAACCTTCATTGTTTATCACCCTGTTAAGTCTCAGCTTAGGTGGCTTTTCC
TTGAGGGAGGGTTTCCTGATATTGCCTCTCAGGGTCTTATATTTCCTGTGGGATTCCACT
GGTCACACAGTACTTACCTC-TTAAGTGTTGGTG-TTATCTTTCTCTCTAACTAAACTAT
TAACCCCTCAGTGCAGGAA-TAACACTATTCTTTGTTCTCTAGCAA--TGTGTCCTCAAT
AAATACTGGCTGAGTGC-TAA-TAAATAA-TCATCATTTC-TTAGATATCCACAAGTCGC
```

222 225

ValGlnThrAla

*TCCC*ACTATTTCCTAACTTATATTTTATTTTACATGGATTTTCTCCAGGTACAAACTGC- 795

230 235 240 245

GluAspAspLeuLysThrAspPheTyrLysAspLeuThrSerLeuGlyHisAsnGluAsn

GAAGATGACTTGAAAACAGATTTCTACAAGGATTTAACTTCTCTTGGACACAATGAAAAT 855

250 255 260 265

GlnGlnGlySerPheSerSerGlnGlyGlySerSerPheSerValProIlePheTyrSer

CAACAAGGCTCATTCTCAAGTCAGGGGGGGAGCTCTTTCAGTGTACCAATTTTTTATTCC 915

270 275 280 285

SerLysArgSerGluAsnIleAsnHisAsnSerAlaPheLysGlnAlaIleGlnAlaSer

TCAAAGAGAAGTGAAAATATCAACCATA-TTCTGCCTTCAAACAAGCCATTCAAGCCTCT 975

288

HisLysLys

CACAAAAAG*GTATCAAAAATGGTTTCC*AACCTTTTATTCTTTCTTGAAAATTAACAGTA- 984

-TCCAGTAATCA-GCAATAAGAAGGGGAGA-TTTTGATATTGTACCAAGTATGACTTCTA

ATACAGATTTAAGTAG

C6 EXON 7

```
TTTTTTGATCACTG-TTTCTTCTTTCTTAGAGAACACGTGCCCCCATTAC-CA-TTTAC-
TTATTGAG-TAATCTTACTGTTTTGCAC-TTGGAC-TTGGTATTGGAAAGCAAAATGAAA
ATTATAATTTTGTACGTGAGAACATGCAATAGAGAGTGAGAATTATTGTAATTCTGTGGT
GGGAATGAAACTAATGATAAATCAATGAC-TTTAAATGCTAGGTACTTCAACCTTTC-TT
```

```
                                          290             295
                                        AspSerSerPheIleArgIleHis
ACCTTTAATCATCTTTTTACTTCTTTCTTCCATTTAGGATTCTAGTTTTATTAGGATCCA  1007
          300             305             310             315
 LysValMetLysValLeuAsnPheIleThrLysAlaLysAspLeuHisLeuSerAspVal
TAAAGTGATGAAAGTCTTAAACTTCACAACGAA-GCTAAAGATCTGC-C-TTTCTGATGT  1067
          320             325             330             335
 PheLeuLysAlaLeuAsnHisLeuProLeuGluTyrAsnSerAlaLeuTyrSerArgIle
CTTTTTGAAAGCACTTAACCATCTGCCTCTAGAATACAACTCTGCTTTGTACAGCCGAAT  1127
          340             345             350             355
 PheAspAspPheGlyThrHisTyrPheThrSerGlySerLeuGlyGlyValTyrAspLeu
ATTCGATGACTTTGGGACTCATTACTTCACCTCTGGCTCCCTGGGAGGCGTGTATGACCT  1187
          360             365
 LeuTyrGlnPheSerSerGluGluLeuLysAsnSerG(ly)
TCTCTATCAGTTTAGCAGTGAGGAACTAAAGAACTCAGGTAACTTTCTTCAGTCCTCACT  1225
CTTTTTAACTGGTATGATAGCATGCAATACATTTTTCTCTTACTCTATCATTTTAATTAT
TCTGCTTTGACTTTATTTAGATCCTGCTTTATTTCCAAAATGATATGAAGAATGATAAAA
AATACACATTCAAATAAATAAAACAAAAATAGAAAACTAAGACCAAAAAATAAAGAAAAA
TAAGTATGCTTAACCACTGTGATTTTAGGCATCATATTTGTCTCTAAATTTCCTAACAAT
CTGGGAACATCATAAGCTTACACAGTAGTGGTTACCCAAAGAGGAAGA-GCA-GCCAATA
TCTCAAGGGAAACCAA-TTT
```

C6 EXON 8

```
ATGGGAAGGGCCAA--TGGCTC-T-TGTGCTTCTGAGGGTCAGTGCTGAAAGGGGAGATG
ATGGGCACTCACGGGACTCCTGTTCAGGATCATCTCTGGGCCCATGCAGGATTGAATCTA
TAGGTAAAACTAACTTGAGCTGTCACTTAAGATTATGGGGACTGGGGAAGAGAGGGGCTA
ATATAAAGTATCTGTAGATCCGAA-TGACCAG-GCACCATGTTCCTCTTTGAATTGTCAA
                                                370       375
                                              (G)lyLeuThrGluGluGluAla
TGGTGC-TTCTCAGTGGTTTCTGTTTGTACTTCATCTCCTAGGTTTAACCGAGGAAGAAG   1243
              380               385               390       395
  LysHisCysValArgIleGluThrLysLysArgValLeuPheAlaLysLysThrLysVal
CCAAACACTGTGTCAGGATTGAAACAAAGAAACGCGTTTTATTTGCTAAGAAAACAAAAG   1303
              400               405               410
  GluHisArgCysThrThrAsnLysLeuSerGluLysHisGluG(ly)
TGGAACATAGGTGCACCACCAACAAGCTGTCAGAGAAACATGAAGGTAACAGTACTCTCT   1348
CTTAGCAGCTTATCCAGTGTGCCCTGAGCCTGGCTGTCATTTACAGACCATATTGGATGT
GACTGTGCTATTTTTAGCTTTTCATTTATTTCATTTCTCTTCCTCAGAACCCTGTGTGTT
GGAATAGGGGATACAAATAAAATAGCATTAGCATTACCTAAAAAAGGAAATTGTTAAGGT
TTTCTTTATCTCATACTTAGGCTGTTTAATGAATAA-TTATTGATTTTGCTCATGGACAG
AA
```

C6 EXON 9

```
TGTAATTGGAACTTTCACCATGACAGTCAATTGACTGGCTATCAGAAGCCATGTGACTTA
TCTTATTCTGGGC-CCCAA-AC-TATGCTAAC-TAAATCAGAAGCAGGGCGGTTGAC-TT
AAAGAGTAAGGGGCAGAGTAGTCAAGGACTGGGCTATATTTGC-TCCTAGTCCTGTCTCT
TAGTAACCATGTGTTCCAGGAC-AATTACTTTCGGCTTCAATGTTTATAATAATACCTAC
TTCTCAGGGTTGTCATGAGAATTATATAAGATGTGTGTAAAGTGCCTGGCATCTACCAGA
TGCTTAATAAATGGTAGCTGTTAGTACTGTTATTATTATTTCTATTATTAAACTCCCTAC
AGGGAGGAGGAGTGTTCTCTTGCTTTTTCACTTAGCTTTCTTAATGGAGCTACATACTGT
ACCATCTACTTCTCCCTTCTGTACAAATAGGTTTTTGTTTGTTTGTTTTAGAATAGGAAT
```

```
                                                411       415
                                                SerPheIleGlnGly
TTGCA-TAAAATTAAATGGGC-TTTTTATTGAAACACCTTGCAGGTTCATTTATACAGGG  1364
          420       425       430       435
 AlaGluLysSerIleSerLeuIleArgGlyGlyArgSerGluTyrGlyAlaAlaLeuAla
AGCAGAGAAATCCATATCCCTGATTCGAGGTGGAAGGAGTGAATATGGAGCAGCTTTGGC  1424
          440       445       450       455
 TrpGluLysGlySerSerGlyLeuGluGluLysThrPheSerGluTrpLeuGluSerVal
ATGGGAGAAAGGGAGCTCTGGTCTGGAGGAGAAGACATTTTCTGAGTGGTTAGAATCAGT  1484
          460       465
 LysGluAsnProAlaValIleAspPheGlu
GAAGGAAAATCCTGCTGTGATTGACTTTGAGGTAAGAGTAAAAATTATTCCAGGTAAGAG  1515
ATCTAAAGTAGTTCTCTAGCAGCCCA--TTGCACATCACACAATTTATTTTCCATTTTCT
TCTCAGACAATGAAAGAAATGCCTTTCTTTAGTATTCATGCACAATGTACCTTTTTTTAA
ATGTTAAATTGCAT
```

C6 EXON 10

```
TCCTGGTCACTGCAAGCATACTGGCACTAAGAAGGTCA---CAGTCTGGAGGTTTTACTG
AGGTAATGGTTAAAAGGTGAGCACGTGATTTA-ATTAGCCTATTAGTTTGTTACGTGTAA
GTTCTTCCATTCATCCTACTGGCAAGTTTATTTCTTCATTTTCTCCCACTCCACA-TTTC
TCATCTAACTCCAACAAAAGCCAAAAC-GTCCTGGAAAAAAAGAACATGGAAAGCTTCTC
TTATCCTCTCTAAATTTTTTGCTGCTTGAATCAAATTTAATGAGGTTTCAGTAGAGCAGG
AAAGAGTTGCCTAACTCTTGGTCACCTATCTCAGTGCTTTGATGCATGCATAGAATACTT
AATAAATAACATGATATATAACCAAGTAAGCAACACCTCCCATGCTTCCTTCACAGTAAA
ATTGTAGGGAAGGTAAGTGCCATATACATCACACCTGTCCACAGGTAGAAGGCTTTTGCC
TACTGGTTCAACAGGTGCAATACTATATTCTGCAATAATCCCTTTGAAGTGTAAAATTTT
CAGGCTTTCCCTCTCATTTCAGAGAACTGGGCAGTAATGGCAACCTATTACCAAAGGGGA
```

```
                                        466         470
                                        LeuAlaProIleValAspLeuValArg
TGTGACTGGACCTCCTCTCTCCCATTTTCTCCTAGCTTGCCCCCATCGTGGACTTGGTAA   1539
  475             480             485             490
  AsnIleProCysAlaValThrLysArgAsnAsnLeuArgLysAlaLeuGlnGluTyrAla
GAAACATCCCCTGTGCAGTGACAAAACGGAACAACCTCAGGAAAGCTTTGCAAGAGTATG   1599
  495             500             505             510
  AlaLysPheAspProCysGlnCysAlaProCysProAsnAsnGlyArgProThrLeuSer
CAGCCAAGTTCGATCCTTGCCAGTGTGCTCCATGCCCTAATAATGGCCGACCCACCCTCT   1659
  515         *   520             525             530
  GlyThrGluCysLeuCysValCysGlnSerGlyThrTyrGlyGluAsnCysGluLysGln
CAGGGACTGAATGTCTGTGTGTGTGTCAGAGTGGCACCTATGGTGAGAACTGTGAGAAAC   1719
  535             540
  SerProAspTyrLysSerA(sn)
AGTCTCCAGATTATAAATCCAGTAAGTATCAGGAATCTATTGTGAGGTAGATAAGTTTTC   1742
```

C6 EXONS 11 and 12

CCCTCCAAAGAGTATTCTAAGTTGGTCAATTAAAAAGAAACAAAACTTCTATTAGCAACC
TCCACCTTGTACAGGCTCAGAGGGAAGATAAACCTGCAAAAAGTGTGAGTCTCAGCTGTA
ACCTACCAACTATGTGAGCTTGGGAAAGTTACCCAGCTTCCCTATGCCTCAATCTGTTCA
TCCATAAAATGAGGATAACACCAGTATCTATCTAAGGAATATGAATATATTATATGCGTA
TTTATATATCAAGAATACATATACAATGCTCATAACATTAAGCCACAACCAAAGTTTTCA
TACAGGACTAAGTATGTTCATAGTTACCTCAAATCCTCCTTAGAAACAGGGTAAGGCATG
GGAGTAAGAATGTATACTACTTCCTTTAAAAGTGTAATTTAATATGCATTCTGTTAAGAA
GATGTTTATATTTATACACATATGAGTGCACATTTTTAAAAGGCCTCCAAAGCCAAGAAT
ACAGAGGTTTCTTAACAGTTGAAAATAATTCATAATGATTGCCATTTGTTGCCATTTTAT
TTAGGCTTATGGGAACAAAGTCTAAAAGGAAATCAGGCAATGTGTCAGGCCCCTTGCACA
GGTAATTTCTTTTAATCCTTAGAAGAACCCTATCATATCCCACAAAGGAAGTTACTGAGC
TAAAAAGTGAGGAAACCAGGTTGGCCTCACTTCCAAACCCAAATTCAGCTGCACCATGAT

542 ** * 550

(A)snAlaValAspGlyGlnTrpGlyCysTrpSer

CCATGGGAGT*GAGTCTTCCTTTGTTCC*AGATGCAGTAGACGGACAGTGGGGTTGTTGGTC 1772

555 560 565 570

SerTrpSerThrCysAspAlaThrTyrLysArgSerArgThrArgGluCysAsnAsnPro

TTCCTGGAGTACCTGTGATGCTACTTATAAGAGATCGAGAACCCGAGAATGCAATAATCC 1832

* 580 585 590

AlaProGlnArgGlyGlyLysArgCysGluGlyGluLysArgGlnGluGluAspCysThr

TGCCCCCCAACGAGGAGGGAAACGCTGTGAGGGGGAGAAGCGACAAGAGGAAGACTGCAC 1892

595 598

PheSerIleMetGluAsnAs(n)

ATTTTCAATCATGGAAAACAA*GTAAGGCCGGGTAAGGGAAAAGAATGGGTCATTTTGCCC* 1914
TAACTCTCAATAACATTGAGAGTTCAAATGGTTATTAGTCTACTGCTAAGTAAAGAAATA
ATTAGAACTTTACCTATGCCAACACAGAATATTAATGAGCCACTGGGTCTACTGCCGTGC
CTTTTTCTGTGTGTAATGTTTTCTATTTAATACATATATGTATGCAATGTGTACACATGT
ATACACATAAGCATATACACACATGTAAATGTGTGTACA*TACATACATATACACACATAT*

600 605 610

(As)nGlyGlnProCysIleAsnAspAspGluGluMetLysGluVal

TTATGTTTCCCTTTTAGTGGACAACCATGTATCAATGATGATGAAGAAATGAAAGAGGTC 1956

615 620 625 * 632

AspLeuProGluIleGluAlaAspSerGlyCysProGlnProValProProGluAsnGly

GATCTTCCTGAGATAGAAGCAGATTCCGGGTGTCCTCAGCCAGTTCCTCCAGAAAATGGA 2016

635

PheIleArg

TTTATCCGGGT*CAGCATCCTAAACCTCA*CTTTTGGTTT-TAGTTTTAAAAACCTCGAAAA 2025
GTATA

C6 EXONS 13 and 14

```
GCCTTAAAGAGTTTAGAGAGTGTATTCATAGCATTAATAAGTTTCTGGTTTCTGCTTTAA

          640       645       650

    AsnGluLysGlnLeuTyrLeuValGlyGluGlnValGluIleSerCysLeuThrGly

TTAAGAATGAAAAGCAACTATACTTGGTTGGAGAAGATGTTGAAATTTCATGCCTTACTG   2079

  655       660       665       670

  PheGluThrValGlyTyrGlnTyrPheArgCysLeuProAspGlyThrTrpArgGlnGly

GCTTTGAAACTGTTGGATACCAGTACTTCAGATGCTTACCAGACGGGACCTGGAGACAAG   2139

  675       *

  AspValGluCysGlnA(rg)

GGGATGTGGAATGCCAACGTGAGAACTGGGCAAACAATTTTGCTTGACTACTTTACAGAA   2159
TGAATGTGCCCAATATAAAAACAGTAAAATCACATCCTTAGATCAATACATTTTACTAGT
TTTTAGTAACTCTGGTGAAAGATTTTGTGAAGCAGAAAATAAATAGAAGGGCTCATTTGC
TAAAACTGGATTTCCAGATTATCTCCATATTTCAGTGTCCTGCTCCTACCTATTACTGTC
ACACACATAGGACTGTGAAGTATGCACTTCAAAAGAAATCCAGGAAGTCTGCAAGCTGTA
TCTAACCACAGACATCTGCTTTGCCCAATGTTCAAA-TGACGCAAAATGCTGAGTCTCAA
GAAGATGATATTTACACTCTTTATGTAACAGAAATGGGAAGTTGTGTAAGATGGGATTTT
GTGTAAGACGTAGTTTTCGCCAAGACACTACTCCTTGGTTCATGTTCACATTCTTTACTG
AGAAAGAGTGAATGAGCTTTAGGCTCTGCACATATCTGTGCTAGAGTAAAGTGAAAAACA
AGATTCAAATAAGGCTTGAATGGGAGTCCAAAGACTTATTGTGGGAGACCATGAAGCAGA
GTTGAGACCAGTGTCAATTAGAATGCCCATCACTCTGCATCAGTCATAGTGTGGTAACTT
GCATTTGGAGAAATCCATTAGGGATAAGAGAAAAGAAAGTGTTTCTTGAAAGGAAAAGGT
CCCAATGATTTCATAAAAGTTGGATGTGTCTGGAGCCAGTTCTTTAGCAAGGGA-TCTCA
GATAATGAAACTAAAGGGAATTTTTCAATGAAGGGTCTGAGGGAAAGCCCAGGATTCTTA
TCCCCCAGGTTTTCTGATCAAAAGGAGGCACCAAAGCTGAACAGATGCAGTAAATCTCAT
CAAAAATCACCTCTTTGGTTGCACAAAATTTGCCTAAATTTGTTGTTTATTCTGCTCCAC

                    681       685

                  (A)rgThrGluCysIleLysProValValGln

TAAACACATATGAAATGTTGCTCTCTTGCAGGGACGGAGTGCATCAAGCCAGTTGTGCAG   2187

690       695       700       705

GluValLeuThrIleThrProPheGlnArgLeuTyrArgIleGlyGluSerIleGluLeu

GAAGTCCTGACAATTACACCATTTCAGAGATTGTATAGAATTGGTGAATCCATTGAGCTA   2247

710       715       720       725

ThrCysPheLysGlyPheValValAlaGlyProSerArgTyrThrCysGlnGlyAsnSer

ACTTGCCCCAAAGGCTTTGTTGTTGCTGGGCCATCAAGGTACACATGCCAGGGGAATTCC   2307

730       735       740

TrpThrProProIleSerAsnSerLeuThrCysGluLysA(sp)

TGGACACCACCCATTTCAAACTCTCTCACCTGTGAAAAAGGTGAGTAGCAGCTCAGTGTT   2347
CTGGGGTCTGATAAGGTGGTGCACCCTGAGATTGGTAGTATAGCCTCTATAGCCATTATA
TTTCTTAAGCACTGAGAGGCTAAAAAAAGA-GCAAGTTTAGACAGTGTGGGTCAAACCGGA
ACTATGCAATAGCTCTGCTAATTCCCTAAGCTCATTGCCATCCTGCGTCCTTTATTCTCT
TCTGGGCTTTTATAACGTGTGAAAGAAAGGGAGAAAGTTAACCTTTTGAATAACCTTAAA
TAAAATTATCAGTAAAGCGATCTCATGGAAATTAGTGAGAACTTCAGGTTTAAGTATTGT
GTCTTGGTTTCACTAGACTCCTGATCATAATAATAAATTACAAGGAATGTAAGTGTCAGT
GGAATGAATGTATGTGGAAGCCATAGACTTCAGAAAGTTAGGATCTTTAAGACACTGTAG
GCAGTATCCTCATAA-
```

C6 EXONS 15 and 16

```
ATTTAGGTTGGAAAATGACACCTCATGTGCACTGGAAAGGACCTTTCAGGTGGAATATTT
AATGACAATGTATTTAGAGAAGCCAAAATGAATGATCTAACTCCAAGAGATAAATATGAA
TCCTTTACCACTGCCTCTTCTCTGAATTCAAGCTTATCACTTAGAATCCATGCTGTGCAC

                                    745                 750

                                (A)spThrLeuThrLysLeuLysGlyHisCys

GTTCTTTTCTCCTTTTTTCTTTTGTTTCAGATACTCTAACAAAATTAAAAGGCCATTGT   2376

      755                 760                 765                 770

GlnLeuGlyGlnLysGlnSerGlySerGluCysIleCysMetSerProGluGluAspCys

CAGCTGGGACAGAAACAATCAGGATCTGAATGCATTTGTATGTCTCCAGAAGAAGACTGT   2436


Se(r)
ACGTAAGAGATACCCTACAGACTGTGTCTGGAA-TTGGGAAAACCAGTCTAGTGTAACCA   2438
AATTTGATGGAAGAGCTGAGCTTTTGAAGCACAGACATGAAATAATGAATTTTCCACATA
ACTGATATTCCTGTGCTCTGCTGAATTCCTCCTTCCCCATTGGGTAGCAAGTTTCTTAAT
TATCTCTCAGCCCAGTTCCCTGTTTAAAATGAGATTGTTCATTATAACAGTCTCATTTTG
GTTTTAAAGACTTAACTTCTGCACATGTAAAAACTTTAAACTCAGTTATATCAACACTTT
CCTGTTCCTCTCACTCAAGCCCAGGGCTTGGAGTGGGGTGGGGAAAACAGGCAAAAGAAA
ACTGGTTGGTGAGTGAACACACTACATTGGCTACGTGCCCCCTGTTTTTTTCTGATCTTC

                                    775                 780

                                (Se)rHisHisSerGluAspLeuCysValPhe

ATATAGAAATGCTTGCTTTCCTCTCTTGGGCAGCCATCATTCAGAAGATCTCTGTGTGTT   2466

      785                 790                 795                 800

 AspThrAspSerAsnAspTyrPheThrSerProAlaCysLysPheLeuAlaGluLysCys

TGACACAGACTCCAACGATTACTTTACTTCACCCGCTTGTAAGTTTTTGGCTGAGAAATG   2526

      805                 810                 815                 820

 LeuAsnAsnGlnGlnLeuHisPheLeuHisIleGlySerCysGlnAspGlyArgGlnLeu

TTTAAATAATCAGCAACTCCATTTTCTACATATTGGTTCCTGCCAAGACGGCCGCCAGTT   2586

      825                 830                 835                 840

 GluTrpGlyLeuGluArgThrArgLeuSerSerAsnSerThrLysLysGluSerCysGly

AGAATGGGGTCTTGAAAGGACAAGACTTTCATCCAACAGCACAAAGAAAGAATCCTGTGG   2646

      845                 850           854

 TyrAspThrCysTyrAspTrpGluLysCysSerA(la)

CTATGACACCTGCTATGACTGGGAAAAATGTTCAGGTAAGTTCCAATGGTGACCATACTA   2679
AATGCTCTCACTTAACCTTCATCTCAGCTAGTAC-TCATCTTTCTTAAAAATAACCTATC
CTTAATCAAAATAATTCCTTTTCATTGTA
```

C6 EXON 17

```
CGCC-CCCCTGGGG-TGATTTAGATAAAGCCAA-CCCTAATGT-CTGTGGTAA-TC-C--
                          855             860             865
                    (A)laSerThrSerLysCysValCysLeuLeuProProGln
TGTCTCTCTCTCTCTC-TTACAGCCTCCACTTCCAAATGTGTCTGCCTATTGCCCCCACA   2717
          870             875             880             885
 CysPheLysGlyGlyAsnGlnLeuTyrCysValLysMetGlySerSerThrSerGluLys
GTGCTTCAAGGGTGGAAACCAACTCTACTGTGTCAAAATGGGATCATCAACAAGTGAGAA   2777
          890             895             900             905
 ThrLeuAsnIleCysGluValGlyThrIleArgCysAlaAsnArgLysMetGluIleLeu
AACATTGAACATCTGTGAAGTGGGAACTATAAGATGTGCAAACAGGAAGATGGAAATACT   2837
          910       913
 HisProGlyLysCysLeuAlaSTOP
GCATCCTGGAAAGTGTTTGGCCTAGCACAATTACTGCTAGGCCCAGCACAATGAACAGAT   2897
TTACCATCCCGAAGAACCAACTCCTACAAATGAGAATTCTTGCACAAACAGCAGACTGGC   2957
ATGCTCAAAGTTACTGACAAAAATTATTTTCTGTTAGTTTGAGATCATTATTCTCCCCTG   3017
ACTCTCCTGTTTGGGCATGTCTTATTCAGTTCCAGCTCATGACGCCCTGTAGCATACCCC   3077
TAGGTACCAACTTCCACAGCAGTCTCGTAAATTCTCCTGTTCACATTGTACAAAAATAAT   3137
GTGACTTCTGAGGCCCTTATGTAGCCTGTGACATTAAGCATTCTCGCAATTAGAAATAAG   3197
AATAAAACCCATAATTTTCTTCAATGAGTTAATAAACAGAAATCTCCAGAACCTCTGAAA   3257
CACATTCTTGAAGCCCAGCTTTCATATCTTCATTCAACAAATAATTTCTGAGTGTGTATA   3317
CAGGATGTCAAGTACTGACCAAAGTCCTGAGAACTCGGCAGATAATAAAACAGACAAAAG   3377
CCTTTGCCTTCATGAAGCATACATTCATTCAGGGGTAGACACACAAAAAATGAAATAAAC   3427
AGGTAAATATGTAGCA-GTTTGATGGTGATAAT-G--TTGGAGAAAATACAACACGGAAG   3444
GGTGAGTCCCATCCTTAATTATCATCTTA-GCAAGGAAATA-TCCATAACAATAAACATC
GCTGTAAGGATGATT
```

References

Abbott C, West L, Povey S, Jeremiah S, Murad Z, DiScipio R, Fey G (1989) The gene for human complement component C9 mapped to chromosome 5 by polymerase chain reaction. *Genomics* **4**:606-609

Alper CA, Boenisch T, Watson L (1972) Genetic polymorphism in human glycine-rich beta-glycoprotein. *J Exp Med* **135**:68-80

Alper CA, Hobart MJ, Lachmann PJ (1975) Polymorphism of the sixth component of complement. In: Isoelectric focusing, Ed: Aburthnott JP, Beely JA, Butterworths publishing, London

Alper CA, Marcus D, Raum D, Petersen BH, Spira TJ (1983) Genetic polymorphism in C8 beta-chains. Evidence for two unlinked genetic loci for the eighth component of human complement (C8). *J Clin Invest* **72**:1526-1531

Alsenz J, Becherer JD, Esparza I, Daoudaki ME, Avila D, Oppermann S, Lambris JD (1989) Structure and function analysis of C3 from different species (Abstract). *Compl Inflamm* **6**:307-308

Anderson L, Sigurdardottir S, Borsch C, Gustafson K (1991) Evolution of *MHC* polymorphism: extensive sharing of polymorphic sequence motifs between human and bovine DRB alleles. *Immunogenetics* **33**:188-193

Arden B, Wakeland EK, Klein J (1980) Structural comparisons of serologivcally indistinguishable H-2K encoded antigens for inbred and wild mice. *J Immunol* **125**:2424-2438

Arden B, Klein J (1982) Biochemical comparison of major histocompatibility complex molecules from different subspecies of Mus musculus: evidence for trans specific evolution of alleles. *Proc Natl Acad Sci USA* **79**:2342-2346

Arlaud GJ, Gagnon J, Villiers CL, Colomb MG (1986) Molecular characterization of the catalytic domains of human complement serine protease C1r. *Biochemistry* **25**:5177-5182

Arnett FC, Shulman LE (1976) Studies in familial systemic lupus erythematosus. *Medicine Baltimore* **55**:313-322

Arnett FC (1987) Familial SLE, the HLA system and the genetics of lupus erythematosus. In: Lupus Erythematosus, Ed; Wallace DJ, Dubois EL, Lea and Rehiger, Philadelphia pp161-184

Avila D, Lambris JD (1990) Isolation and characterization of the third complement component of axolotl (Ambystoma mexicanum). *Comp Biochem Physiol B* **95:**839-845

Awdeh ZL, Alper CA (1980) Inherited structural polymorphism of the fourth component of human complement. *Proc Natl Acad Sci U S A* **77:**3576-3580

Babbitt BP, Allen PM, Matsueda G, Haber E, Unanue ER (1985) Binding of immunogenic peptides to Ia histocompatibility molecules. *Nature* **317:**359-361

Baur MP, Danilovs JA (1980) Population analysis of HLA A, B, C and other genetic markers, In: Histocompatibility Testing, Ed: Terasaki PI, UCLA tissue typing lab, Los Angeles pp955-957

Belt KT, Carroll MC, Porter RR (1984) The structural basis of the multiple forms of human complement component C4. *Cell* **36:**907-914

Belt KT, Yu CY, Carroll MC, Porter RR (1985) Polymorphism of human complement component C4. *Immunogenetics* **21:**173-180

Bellwood PS (1989) The colonization of the Pacific: some current hypothesis In: The colonization of the Pacific: A genetic trail, Ed; Hill AVS, Serjeantson SW, Oxford Science Press, Oxford pp1-59

Bellwood PS (1991) The Austronesian dispersal and the origin of languages. *Scientific American* **265**:88-93

Bentley DR (1988) Structural superfamilies of the complement system. *Exp Clin Immunogenet* **5:**69-80

Bhakdi S, Maillet F, Muhly M, Kazatchkine MD (1988) The cytolytic C5b-9 complement complex: feedback inhibition of complement activation. *Proc Natl Acad Sci USA* **85**:1912-1916

Biesecker G, Gomez CM (1989) Inhibition of acute passive transfer experimental autoimmune myasthenia gravis with Fab antibody to complement C6. *J Immunol* **142:**2654-2659

Birdsell JB (1977) The recalibration of a paradigm for the first peopling of greater Australia. In: Sunda and Sahul Ed; Allen J, Golson J, Jones R, Academic Press, London pp113-167

Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC (1987a) Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* **329:**506-512

Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC (1987b) The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature* **329:**512-518

Blake MS, Johnston KH, Russell Jones GJ, Gotschlich EC (1984) A rapid, sensitive method for detection of alkaline phosphatase-conjugated anti-antibody on Western blots. *Anal Biochem* **136:**175-179

Block SR, Winfield JB, Lockshin MD, D'Angelo WA, Christian CL (1975) Studies of twins with systemic lupus erythematosus. A review of the literature and presentation of 12 additional sets. *Am J Med* **59:**533-552

Bodmer WF, Bodmer JG (1978) Evolution and function of the HLA system. *Brit Med Bull* **34:**309-316

Botto M, Fong KY, So AK, Koch C, Walport MJ (1990) Molecular basis of polymorphisms of human complement component C3. *J Exp Med* **172:**1011-1017

Boyer JT, Gall EP, Norman ME, Nilsson UR, Zimmerman TS (1975) Hereditary deficiency of the seventh component of complement. *J Clin Invest* **56:**905-913

Braun L, Schneider PM, Giles CM, Bertrams J, Rittner C (1990) Null alleles of human complement C4. Evidence for pseudogenes at the C4A locus and gene conversion at the C4B locus. *J Exp Med* **171**:129-140

Brown P (1987) Pleistocene and Holocene size reduction: the Australian human evidence. *Archaeology in Oceania* **22**:41-46

Bulmer M (1991) Use of generalised least squares in reconstructing phylogenies from sequence data. *Mol Biol Evol* **8**:863-883

Campbell AK, Morgan BP (1985) Monoclonal antibodies demonstrate protecion of polymorphonuclear leucocytes against complement attack. *Nature* **317**:164-166

Campbell RD, Morley BJ, Sargent CA, Janjua NJ (1985) Molecular basis for allelic variation at the factor B locus. *Complement* ***2***:14-15

Campbell RD (1987) The molecular genetics and polymorphism of C2 and factor B. *Br Med Bull* **43**:37-49

Campbell RD, Law SK, Reid KB, Sim RB (1988) Structure, organization, and regulation of the complement genes. *Annu Rev Immunol* **6**:161-195

Campbell RD, Dunham I, Kendall E, Sargent CA (1990) Polymorphism of the human complement component C4. *Exp Clin Immunogenet* **7**:69-84

Caras IW, Davitz MA, Rhee L, Weddell G, Martin DW, Jr., Nussenzweig V (1987) Cloning of decay-accelerating factor suggests novel use of splicing to generate two proteins. *Nature* **325**:545-549

Carney DF, Haviland DL, Noack D, Wetsel RA, Vik DP, Tack BF (1991) Structural aspects of the human C5 gene. Intron/exon organization, 5'-flanking region features, and characterization of two truncated cDNA clones. *J Biol Chem* **266**:18786-18791

Carroll MC, Campbell RD, Bentley DR, Porter RR (1984) A molecular map of the human major histocompatibility complex class III region linking complement genes C4, C2 and factor B. *Nature* **307**:237-241

Carroll MC, Katzman P, Alicot EM, Koller BH, Geraghty DE, Orr HT, Strominger JL, Spies T (1987) Linkage map of the human major histocompatibility complex including the tumor necrosis factor genes. *Proc Natl Acad Sci USA* **84**:8535-8539

Cavalli-Sforza LL, Bodmer WF (1971) The genetics of human populations. Freeman and Sons publishers, San Francisco

Chakravarti DN, Campbell RD, Gagnon J (1983) Amino acid sequence of a polymorphic segment from fragment C4d of human complement component C4. *FEBS Lett* **154**:387-390

Chakravarti A, Buetow KH, Antonarakis SE, Waber PG, Boehm CD, Kazazian HH (1984) Nonuniform recombination within the human beta-globin gene cluster. *Am J Hum Genet* **36**:1239-1258

Chakravarti DN, Campbell RD, Porter RR (1987) The chemical structure of the C4d fragment of the human complement component C4. *Mol Immunol* **24**:1187-1197

Chakravarti DN, Chakravarti B, Parra CA, Muller Eberhard HJ (1989) Structural homology of complement protein C6 with other channel-forming proteins of complement. *Proc Natl Acad Sci U S A* **86**:2799-2803

Chapel HM, Peto TE, Luzzi GA, Thompson RA, Fielder AH, Batchelor JR (1987) Combined familial C7 and C4B deficiency in an adult with meningococcal disease. *Clin Exp Immunol* **67**:55-58

Christiansen FT, Dawkins RL, Uko G, McCluskey J, Kay PH, Zilko PJ (1983) Complement allotyping in S. *Aust N Z J Med* **13**:483-488

Chung LP, Bentley DR, Reid KB (1985) Molecular cloning and characterization of the cDNA coding for C4b-binding protein, a regulatory protein of the classical pathway of the human complement system. *Biochem J* **230:**133-141

Collier S, Tassabehji M, Sinnott P, Strachan T (1993) A de novo pathological point mutation at the 21hydroxylase locus: implications for gene conversion in the human genome [published erratum appears in *Nat Genet* 1993 May; 4(1): 101]. *Nat Genet* **3**:260-265

Cooper NR, Nemerow GR (1983) Complement, viruses and virus infected cells. *Springer Semin Immunopathol* **6**:327-347

Cooper NR (1988) Laboratory investigation of complement proteins and complement receptors. In: Baillieres Clinical Immunology and Alllergy Eds: Kazatchkine MD Bailliere Tindall, London pp 263-293

Corbo RM, Tartaglia M, Scacchi R, Rickards O, Dash S, Sharma SK, Ramdeo IN, Sangatramani TC, Vyas M, Modiano G (1991) A survey of six genetic markers on the populations of Punjab and Rajasthan (India). *Gene Geogr* **5:**113-121

Coto E, Dominguez O, Martinez Naves E, Urra JM, Gutierrez V, Lopez Larrea C (1990a) TaqI polymorphism at the human complement component C9 gene. *Nucleic Acids Res* **18:**5581

Coto E, Martinez Naves E, Dominguez O, Urra JM, Rodriguez V, Lopez Larrea C (1990b) TaqI polymorphism in the complement component C7 gene. *Nucleic Acids Res* **18:**1929

Coto E, Martinez Naves E, Dominguez O, DiScipio RG, Urra JM, Lopez Larrea C (1991a) DNA polymorphisms and linkage relationship of the human complement component C6, C7, and C9 genes. *Immunogenetics* **33:**184-187

Coto E, Dominguez O, Martinez-Naves E, Setien F, Gutierrez V, Lopez-Larrea C (1991b) MspI polymorphism at the human complement component C6 gene (C6). *Nucl Acids Res* **19**:194

Cotton RG, Rodrigues NR, Campbell RD (1988) Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. *Proc Natl Acad Sci U S A* **85:**4397-4401

Dankert JR, Esser AF (1985) Proteolytic modification of human complement protein C9: loss of poly(C9) and circular lesion formation without impairment of function. *Proc Natl Acad Sci U S A* **82:**2128-2132

Davies JP, Winchester BG, Malcolm S (1993) Sequence variation in the first exon of α-galactosidase A. *J Med Genet* **30**:658-663

de Bruijn MH, Fey GH (1985) Human complement component C3: cDNA coding sequence and derived primary structure. *Proc Natl Acad Sci U S A* **82:**708-712

DeHoratius RJ, Messner RP (1975) Lymphocytotoxic antibodies in family members of patients with systemic lupus erythematosus. *J Clin Invest* **55:**1254-1258

DeHoratius RJ, Pillarisetty R, Messner RP, Talal N (1975) Anti-nucleic acid antibodies in systemic lupus erythematosus patients and their families. Incidence and correlation with lymphocytotoxic antibodies. *J Clin Invest* **56:**1149-1154

Delage JM, Lehner Netsch G, Lafleur R, Simard J, Brun G, Prochazka E (1979) Simultaneous occurrence of hereditary C6 and C2 deficiency in a French-Canadian family. *Immunology* **37:**419-428

Densen P (1991) Complement deficiencies and meningococcal disease. *Clin Exp Immunol* **86 Suppl 1:**57-62

Dewald G, Nothen MM, Cichon S (1993) Polymorphism of human complement component C6: an amino acid substitution (Glu/Ala) within the second thrombospondin repeat differentiates between the two common allotypes C6 A and C6 B. *Biochem Biophys Res Commun* **194:**458-464

DiScipio RG, Gagnon J (1982) Characterization of human complement components C6 and C7. *Mol Immunol* **19:**1425-1431

DiScipio RG, Gehring MR, Podack ER, Kan CC, Hugli TE, Fey GH (1984) Nucleotide sequence of cDNA and derived amino acid sequence of human complement component C9. *Proc Natl Acad Sci U S A* **81:**7298-7302

DiScipio RG, Chakravarti DN, Muller Eberhard HJ, Fey GH (1988) The structure of human complement component C7 and the C5b-7 complex. *J Biol Chem* **263:**549-560

DiScipio RG, Hugli TE (1989) The molecular architecture of human complement component C6. *J Biol Chem* **264:**16197-16206

Doherty PC, Zinkernagel RM (1975) A biological role for the major histocompatibility antigens. *Lancet* **1:** 1406

Dunckley H, Gatenby PA, Hawkins B, Naito S, Serjeantson SW (1987) Deficiency of C4A is a genetic determinant of systemic lupus erythematosus in three ethnic groups. *J Immunogenet* **14:**209-218

Eldridge PR, Hobart MJ, Lachmann PJ (1983) The genetics of the sixth and seventh components of complement in the dog: polymorphism, linkage, locus duplication, and silent alleles. *Biochem Genet* **21:**81-91

Ellison RT, Mason SR, Kohler PF, Curd JG, Reller LB (1986) Meningococcemia and acquired complement deficiency. Association in patients with hepatic failure. *Arch Intern Med* **146:**1539-1540

Engleman EP, Shearn MA (1983) Arthritis and allied rheumatic disorders. In: Current medical diagnosis and treatment, Ed; Knapp MA, Chatton MJ

Farries TC, Steuer KL, Atkinson JP (1990) Evolutionary implications of a new bypass activation pathway of the complement system. *Immunol Today* **11:**78-80

Farries TC, Atkinson JP (1991) Evolution of the complement system. *Immunol Today* **12:**295-300

Feinberg AP, Vogelstein B (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* **132:**6-13

Feinberg AP, Vogelstein B (1984) "A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity". Addendum. *Anal Biochem* **137:**266-267

Feldman HA, Schreiber R (1977) Toxoplasmosis dangerous in complement deficiency [letter]. *N Engl J Med* **297:**1403

Fernie BA, Delbridge G, Hobart MJ (1993) Correlation of a Glu/Ala substitution at position 98 with the complement C6 A/B phenotypes. *Hum Mol Genet* **2:**591-592

Fernie BA, Hobart MJ, Delbridge GJ, Potter PC, Orren A, Lachmann PJ (1994) C6 haplotypes: association of a Dde I site polymorphism to complement deficiency genes and the Msp I restriction fragment length polymorphism (RFLP). *Clin Exp Immunol* **95**:351-356

Fernie BA, Wurzner, Unsworth DJ, Tuxworth RI, Orren A, Hobart MJ, Lachmann PJ (in preparation) DNA polymorphisms of the C6 and C7 genes.

Fielder AH, Walport MJ, Batchelor JR, Rynes RI, Black CM, Dodi IA, Hughes GR (1983) Family study of the major histocompatibility complex in patients with systemic lupus erythematosus: importance of null alleles of C4A and C4B in determining disease susceptibility. *Br Med J Clin Res Ed* **286:**425-428

Figueroa F, Gunther E, Klein J (1988) MHC polymorphism pre-dating speciation. *Nature* **335**:265-267

Figueroa JE, Densen P (1991) Infectious diseases associated with complement deficiencies. *Clin Microbiol Rev* **4:**359-395

Fisher RA, Ford EB, Huxley J (1939) Taste testing the Anthropoid Apes. *Nature* **144**:750

Fogel S, Mortimer RK (1969) Informational transfer in meiotic gene conversion. *Proc Natl Acad Sci U S A* **62:**96-103

Fong KY, Botto M, Walport MJ, So AK (1990) Genomic organization of human complement component C3. *Genomics* **7:**579-586

Francke U, Pellegrino MA (1977) Assignment of the major histocompatibility complex to a region of the short arm of human chromosome 6. *Proc Natl Acad Sci U S A* **74:**1147-1151

Fraser BM, Poon MC, Hoar DI (1992) Identification of factor IX mutations in haemophilia B: application of polymerase chain reaction and single strand conformation analysis. *Hum Genet* **88:**426-430

Fronek Z, Timmerman LA, Alper CA, Hahn BH, Kalunian K, Peterlin BM, McDevitt HO (1990) Major histocompatibility complex genes and susceptibility to systemic lupus erythematosus. *Arthritis Rheum* **33**:1542-1553

Fujisaku A, Harley JB, Frank MB, Gruner BA, Frazier B, Holers VM (1989) Genomic organization and polymorphisms of the human C3d/Epstein-Barr virus receptor. *J Biol Chem* **264:**2118-2125

Fujita T, Gigli I, Nussenzweig V (1978) Human C4-binding protein. II. Role in proteolysis of C4b by C3b -inactivator. *J Exp Med* **148:**1044-1051

Fujita T, Nussenzweig V (1979) The role of C4-binding protein and beta 1H in proteolysis of C4b and C3b. *J Exp Med* **150:**267-276

Gaidano G, Ballerini P, Gong JZ, Inghirami G, Neri A, Newcomb EW, Magrath IT, Knowles DM, Dalla Favera R (1991) p53 mutations in human lymphoid malignancies: association with Burkitt lymphoma and chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A* **88:**5413-5417

Gally JA, Edelman GM (1992) Evidence for gene conversion in genes for cell-adhesion molecules. *Proc Natl Acad Sci U S A* **89:**3276-3279

Gianella Borradori A, Borradori L, Schneider PM, Gautier E, Spath PJ (1990) Combined complete C5 and partial C4 deficiency in humans: clinical consequences and complement-mediated functions in vitro. *Clin Immunol Immunopathol* **55:**41-55

Gigli I, Fujita T, Nussenzweig V (1979) Modulation of the classical pathway C3 convertase by plasma proteins C4 binding protein and C3b inactivator. *Proc Natl Acad Sci U S A* **76:**6596-6600

Glass D, Raum D, Gibson D, Stillman JS, Schur PH (1976) Inherited deficiency of the second component of complement. Rheumatic disease associations. *J Clin Invest* **58:**853-861

Glass D, Raum D, Balavitch D, Kagan E, Rabson A, Schur PH, Alper CA (1978) Inherited deficiency of the sixth component of complement: a silent or null gene. *J Immunol* **120:**538-541

Glavac D, Dean M (1993) Optimisation of the single-strand conformation polymorphism (SSCP) technique for detection of point mutations. *Hum Mutat* **2:**404-414

Golding B (1992) The prospects for polymorphisms shared between species. *Heredity Edinburgh* **68:**263-276

Goldman MB, Cohen C, Stronski K, Bangalore S, Goldman JN (1982) Genetic control of C6 polymorphism and C6 deficiency in rabbits. *J Immunol* **128:**43-48

Gorski J, Mach B (1986) Polymorphism of human Ia antigens: gene conversion between two DR beta loci results in a new HLA-D/DR specificity. *Nature* **322:**67-70

Gottardi E, Losekoot M, Fodde R, Saglio G, Camaschella C, Bernini LF (1992) Rapid identification by denaturing gradient gel electrophoresis of mutations in the gamma-globin gene promoters in non-deletion type HPFH. *Br J Haematol* **80:**533-538

Granados J, Cramer DV, Caputo JB, Marcus D, Alper CA (1984) Genetic polymorphism of the sixth component (C6) of rat complement. *J Immunol* **133**:405-407

Griffiths RC, Watterson GA (1990) The number of alleles in multigene families. *Theor Popul Biol* **37**:110-123

Groggel GC, Adler S, Rennke HG, Couser WG, Salant DJ (1983) Role of the terminal complement pathway in experimental membranous nephropathy in the rabbit. *J Clin Invest* **72**:1948-1957

Grossberger D, Marcuz A, Du Pasquier L, Lambris JD (1989) Conservation of structural and functional domains in complement component C3 of Xenopus and mammals. *Proc Natl Acad Sci U S A* **86**:1323-1327

Gyllensten UB, Erlich HA (1989) Ancient roots for polymorphism at the HLA-DQα locus in primates. *Proc Natl Acad Science USA* **86**:9986-9990

Haber M (1984) Log linear models for linked loci. *Biometrics* **40**:189-198

Habgood PJ (1985) The origin of the Australian aborigines: an alternative approach and view. In:Hominid evolution: past, present and future Ed: Tobias P, Liss AR, New York pp367-380

Haefliger JA, Peitsch MC, Jenne DE, Tschopp J (1991) Structural and functional characterization of complement C8 gamma, a member of the lipocalin protein family. *Mol Immunol* **28**:123-131

Haefliger JA, Tschopp J, Nardelli D, Wahli W, Kocher HP, Tosi M, Stanley KK (1987) Complementary DNA cloning of complement C8 beta and its sequence homology to C9. *Biochemistry* **26**:3551-3556

Haefliger JA, Tschopp J, Vial N, Jenne DE (1989) Complete primary structure and functional characterization of the sixth component of the human complement system. Identification of the C5b-binding domain in complement C6. *J Biol Chem* **264**:18041-18051

Hall JE, Jr, Alper CA (1977) Genetic polymorphism of the sixth component of complement (C6) in the rhesus monkey. *J Immunol* **119:**253-255

Hauptmann G, Uring-Lambert B, Vegneduzzi-Lamouche N, Mascart-Lemone F (1987) RFLP studies of 3 complete C4 deficient patients. *Complement* **4**:166-169

Hauptmann G (1989) Frequency of complement deficiencies in man, disease associations and chromosome assignment of complement genes and linkage groups. A summary of the data from the literature. *Complement Inflamm* **6:**74-80

Hedrick PW (1987) Gametic disequilibrium measures: proceed with caution. *Genetics* **117:**331-341

Herrmann D, Sodetz JM, Rittner C, Schneider PM (1989) DNA polymorphism of the human complement C8 beta gene: formal genetics and intragenic localization. *Immunogenetics* **30:**291-295

Hill WG (1974) Estimation of linkage disequilibrium in randomly mating populations. *Heredity Edinburgh* **33:**229-239

Hill WG (1975) Linkage disequilibrium among multiple neutral alleles produced by mutation in finite population. *Theor Popul Biol* **8:**117-126

Hobart MJ, Lachmann PJ, Alper CA (1975) Polymorphism of human C6. In: Protides of the Biological Fluids, Ed: Peeters H, Pergamon Press, New York pp575-580

Hobart MJ, Lachmann PJ (1976) Allotypes of complement components in man. *Transplant Rev* **32:**26-42

Hobart MJ, Cook PJL, Lachmann PJ (1977) Linkage studies with C6. *J Immunogen* **4**:423-428

Hobart MJ, Joysey V, Lachmann PJ (1978) Inherited structural variation and linkage relationships of C7. *J Immunogenet* **5:**157-163

Hobart MJ, vaz Guedes MA, Lachmann PJ (1981) Polymorphism of human C5. *Ann Hum Genet* **45:**1-4

Hobart MJ (1990) C7 reference typing and nomenclature recommendations. *Complement Inflamm* **7**:240-242

Hobart MJ, Fernie B, DiScipio RG (1993a) Structure of the human C6 gene. *Biochemistry* **32**:6198-6205

Hobart MJ, Fernie BA, DiScipio RG, Lachmann PJ (1993b) A physical map of the C6 and C7 complement component gene region on chromosome 5p13. *Hum Mol Genet* **2**:1035-1036

Hobart MJ, Fernie B, DiScipio RG (in preparation) Structure of the C7 gene.

Holers VW, Cole JL, Lublin DM, Seya T, Atkinson JP (1985) Human C3b- and C4b-regulatory proteins. A new multi -gene family *Immunol Today* **6**:188-192

Hongyo T, Buzard GS, Calvert RJ, Weghorst CM (1993) 'Cold SSCP': a simple, rapid and non-radioactive method for optimized single-strand conformation polymorphism analyses. *Nucleic Acids Res* **21**:3637-3642

Howard OM, Rao AG, Sodetz JM (1987) Complementary DNA and derived amino acid sequence of the beta subunit of human complement protein C8: identification of a close structural and ancestral relationship to the alpha subunit and C9. *Biochemistry* **26**:3565-3570

Hu J, Du R (1992) [Genetic polymorphism of complement component six (C6) in five Han subpopulations]. *I Chuan Hsueh Pao* **19**:481-485

Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**:167-170

Imagawa DK, Barbour SE, Morgan BP, Wright TM, Shin MS, Ramm LF (1987) Role of complement C9 and calcium in the generation of arachiodonic acid and its metabolites from rat polymorphonuclear leucocytes. *Mol Immunol* **24**:1263-1271

Inoue K, Niesen N, Biesecker G, Milgrom F, Albini B (1993) Role of late complement components in experimental autoimmune thyroiditis. *Clin Immunol Immunopathol* **66**:1-10

Ishii Y, Zhu ZB, Macon KJ, Volanakis JE (1993) Structure of the human C2 gene. *J Immunol* **151**:170-174

Jacob CO, Fronek Z, Lewis GD, Koo M, Hansen JA, McDevitt HO (1990) Heritable major histocompatibility complex class II-associated differences in production of tumor necrosis factor alpha: relevance to genetic predisposition to systemic lupus erythematosus. *Proc Natl Acad Sci U S A* **87**:1233-1237

Jazwinska EC, Dunckley H, Propert DN, Gatenby PA, Serjeantson SW (1988) Gm typing by immunoglobulin heavy-chain gene RFLP analysis. *Am J Hum Genet* **43**:175-181

Jensen JA, Festa E, Smith DS, Cayer M (1981) The complement system of the nurse shark: hemolytic and comparative characteristics. *Science* **214**:566-569

Jeremiah SJ, Abbott CM, Murad Z, Povey S, Thomas HJ, Solomon E, DiScipio RG, Fey GH (1990) The assignment of the genes coding for human complement components C6 and C7 to chromosome 5. *Ann Hum Genet* **54**:141-147

Johnson CA, Densen P, Hurford RK, Jr., Colten HR, Wetsel RA (1993) Type I human complement C2 deficiency. A 28-base pair gene deletion causes skipping of exon 6 during RNA splicing. *J Biol Chem* **268**:2268

Kai C, Yoshikawa Y, Yamanouchi K, Okada H (1983) Isolation and identification of the third component of complement of Japanese quails. *J Immunol* **130**:2814-2820

Kan YW, Dozy AM (1978) Antenatal diagnosis of sickle-cell anaemia by D.N.A. analysis of amniotic-fluid cells. *Lancet* **2**:910-912

Kaneco Y, Tsukamoto A, Uwatoko S, Kurokawa K, Maeda H, Kuwata S, Tokunaga K, Juji T (1993) Chronic neutropenia associated with C2 and C9 deficiency. *Int Arch Allergy Immunol* **102**:304-306

Kaplan D (1984) The onset of disease in twins and siblings with systemic lupus erythematosus. *J Rheumatol* **11**:648-652

Kaufman KM, Snider JV, Spurr NK, Schwartz CE, Sodetz JM (1989) Chromosomal assignment of genes encoding the alpha, beta, and gamma subunits of human complement protein C8: identification of a close physical linkage between the alpha and the beta loci. *Genomics* **5**:475-480

Kaufmann T, Rittner C, Schneider PM (1993) The human complement component C8B gene: structure and phylogenetic relationship. *Hum Genet* **92**:69-75

Kelly RB, Cozzarelli NR, Deutscher MP, Lehman IR, Kornberg A (1970) Enzymatic synthesis of deoxyribonucleic acid. XXXII. Replication of duplex deoxyribonucleic acid by polymerase at a single strand break. *J Biol Chem* **245**:39-45

Kerr MA (1980) The human complement system: assembly of the classical pathway C3 convertase. *Biochem J* **189**:173-181

Kinoshita E, Matsumoto T, Kondoh T, Yoshimoto M, Niikawa N, Tsuji Y (1991) Deletion pattern in the 21 -hydroxylase gene detected by polymerase chain reaction. *Acta Paediatr Jpn* **33**:1-5

Kirk RL, Ranford PR, Court J, Zimmet P (1980) Distribution of complement C'2 and C'6 types in Australian cases of diabetes mellitus. *Med J Aust* **2**:614-615

Kirk RL (1981) Aboriginal man adapting, Clarendon Press, Oxford

Kishimoto TK, O'Connor K, Lee A, Roberts TM, Springer TA (1987) Cloning of the beta subunit of the leukocyte adhesion proteins: homology to an extracellular matrix receptor defines a novel supergene family. *Cell* **48**:681-690

Klein J (1980) Generation of diversity at MHC loci: implications for T-cell recptor repertoires. *Immunol* **80**:239-253

Klein J (1986) Natural history of the major histocompatibility complex. John Wiley and Sons, New York

Klein J (1987) Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. *Hum Immunol* **19**:155-162

Klickstein LB, Wong WW, Smith JA, Weis JH, Wilson JG, Fearon DT (1987) Human C3b/C4b receptor (CR1). Demonstration of long homologous repeating domains that are composed of the short consensus repeats characteristics of C3/C4 binding proteins. *J Exp Med* **165**:1095-1112

Klob WP, Muller Eberhard HJ (1976) The membrane attack mechanism of complement: the three polypeptide chain structure of the eigth component (C8). *J Exp Med* **143**:1131-1139

Koch C, Behrendt N (1986) A novel polymorphism of human complement component C3 detected by means of a monoclonal antibody. *Immunogenetics* **23**:322-325

Kohonen-Corish MRJ, Serjeantson SW (1986) HLA-DRb gene DNA polymorphisms revealed by Taq I correlate with HLA-DR specificities. *Hum Immunol* **14**:263-271

Komatsu N, Kido A, Oya M (1989) C7 polymorphism in Japanese: new allele C7*8. *Hum Hered* **39**:49-51

Koski CL, Ramm LE, Hammer CH, Mayer MM, Shin ML (1983) Cytolysis of nucleated cells by complement: cell death displays multi-hit characteristics. *Proc Natl Acad Sci USA* **80**:3816-3819

Kristensen T, Wetsel RA, Tack BF (1986) Structural analysis of human complement protein H: homology with C4b binding protein, beta 2-glycoprotein I, and the Ba fragment of B2. *J Immunol* **136**:3407-3411

Kristensen T, D'Eustachio P, Ogata RT, Chung LP, Reid KB, Tack BF (1987) The superfamily of C3b/C4b-binding proteins. *Fed Proc* **46**:2463-2469

Kucher AN, Puzyrev VP, Ivanova OF, Khu TsTu, Siuiu-TsTu, Du ZhF (1993) [Genetic polymorphism of the sixth (C6) and seventh (C7) components of complement in Russian rural inhabitants of the Tomsk region]. *Genetika* **29**:1889-1894

Kuhnl P, Kreckel P (1980) C6 phenotypes and allele frequencies in a German population. *Immunobiology* **158**:50-54

Kunstmann G, Mauff G, Pulverer G (1980) C6 polymorphism and rare alleles in Western Germany. *Immunobiology* **158**:55-59

Kunstmann G, Mauff G (1980) Genetic polymorphism of rabbit C6. *Immunobiology* **158**:30-33

Lachmann PJ, Hobart MJ, Woo P (1978) Combined genetic deficiency of C6 and C7 in aman. *Clin Exp Immunol* **33**:193-203

Lachmann PJ, Hobart MJ, Woo P (1979) Combined genetic deficiency of C6 and C7 in man. *Clin Exp Immunol* **33**:193-203

Lachmann PJ, Walport MJ (1987) In: Autoimmunity and autoimmune disease, Ed: Whelan J, John Wiley and Sons pp149-171

Laursen I, Koch C (1989) Purification of chicken C3 and a structural and functional characterization. *Scand J Immunol* **30**:529-538

Law SK, Gagnon J (1985) The primary structure of the fourth component of human complement (C4)-C-terminal peptides. *Biosci Rep* **5**:913-921

Law SK, Gagnon J, Hildreth JE, Wells CE, Willis AC, Wong AJ (1987) The primary structure of the beta-subunit of the cell surface adhesion glycoproteins LFA-1, CR3 and p150,95 and its relationship to the fibronectin receptor. *EMBO J* **6**:915-919

Lawlor DA, Ward FE, Ennis PD, Jackson AP, Parham P (1988) HLA-A and B polymorphisms predate the divergence of humans and chimpanzees. *Nature* **335**:268-271

Leddy JP, Frank MM, Gaither T, Baum J, Klemperer MR (1974) Hereditary deficiency of the sixth component of complement in man I. Immunochemical, biologic and family studies. *J Clin Invest* **53**:544-553.

Lehner PJ, Davies KA, Walport MJ, Cope AP, Wurzner R, Orren A, Morgan BP, Cohen J (1992) Meningococcal septicaemia in a C6-deficient patient and effects of plasma transfusion on lipopolysaccharide release. *Lancet* **340**:1379-1381

Lewontin RC, Kojima K (1960) The evolutionary dynamics of complex polymorphisms. *Evolution* **14**:458-472

Lewontin RC (1964) The interaction of selection and linkage I. General considerations: heterotic models. *Genetics* **49**:49-67

Lewontin RC (1988) On measures of gametic disequilibrium. *Genetics* **120**:849-852

Leytus SP, Kurachi K, Sakariassen KS, Davie EW (1986) Nucleotide sequence of the cDNA coding for human complement C1r. *Biochemistry* **25**:4855-4863

Li WH, Saddler LA (1991) Low nucleotide diversity in man. *Genetics* **129**:513-523

Lichtenheld MG, Olsen KJ, Lu P, Lowrey DM, Hameed A, Hengartner H, Podack ER (1988) Structure and function of human perforin. *Nature* **335**:448-451

Lim D, Gewurz A, Lint TF, Ghaze M, Sepheri B, Gewurz H (1976) Absence of the sixth component of complement in a patient with repeated episodes of meningococcal meningitis. *J Pediatr* **89**:42-47

Lingren (1949) The yeast cell; Its genetics and cytology. Educational Publishers, St Louis

Litman GW, Kehoe JM (1978) *Comp Immunol* **5**:205-228

Liu TJ, Liu L, Marzluff WF (1987) Mouse histone H2A and H2B genes: four functional genes and a pseudogene undergoing gene conversion with a closely linked functional gene. *Nucleic Acids Res* **15**:3023-3039

Lockshin MD (1991) Therapy for systemic lupus erythematosus. *New Eng J Med* **324**:189-191

Loos M, Heinz HP (1986) Component deficiencies. 1. The first component: C1q, C1r, C1s. *Prog Allergy* **39**:212-231

Lopez-Trascasa M, Vicario JL, Martinez-Laso J, Gil Aguado A, Pascual-Salcedo D, Arnaiz A, Fontan G (1988) C7 and C4B deficiency in a patient with rheumatoid arthritis. *Complement* **4**:187-192

Lu B, Du CS (1988) [Genetic polymorphism of complement component C6 in four Chinese populations in Guangdong province: three new variants]. *I Chuan Hsueh Pao* **15**:455-461

Lublin DM, Lemons RS, Le Beau MM, Holers VM, Tykocinski ML, Medof ME, Atkinson JP (1987) The gene encoding decay-accelerating factor (DAF) is located in the complement-regulatory locus on the long arm of chromosome 1. *J Exp Med* **165**:1731-1736

Maiste PJ, Weir BS (1992) Estimating linkage disequilibrium from conditional data [letter]. *Am J Hum Genet* **50**:1139-1140

Mancuso DJ, Tuley EA, Westfield LA, Worrall NK, Shelton Inloes BB, Sorace JM, Alevy YG, Sadler JE (1989) Structure of the gene for human von Willebrand factor. *J Biol Chem* **264**:19514-19527

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning A laboratory manual. Cold Spring Harbour, New York, Cold Spring Harbour Laboratory

Marazziti D, Eggerton G, Fey GH, Stanley KK (1988) Relationships between the gene and protein structure in human complement component C9. *Biochem* **27**:6529-6534

Marlin SD, Morton CC, Anderson DC, Springer TA (1986) LFA -1 immunodeficiency disease. Definition of the genetic defect and chromosomal mapping of alpha and beta subunits of the lymphocyte function-associated antigen 1 (LFA-1) by complementation in hybrid cells. *J Exp Med* **164:**855-867

Maruyama T (1982) Stochastic integrals and their application to population genetics. In: Molecular Evolution, protein polymorphisms and the neutral theory, Ed: Kimura M. Japan Scientific Societies Press, Tokyo pp151-166

Mauff G, Alper CA, Awdeh Z, Batchelor JR, Bertrams J, Bruun Petersen G, Dawkins RL, Demant P, Edwards J, Grosse Wilde H, Hauptmann G, Klouda P, Lamm L, Mollenhauer E, Nerl C, Olaisen B, O'Neill G, Rittner C, Roos MH, Skanes V, Teisberg P, Wells L (1983) Statement on the nomenclature of human C4 allotypes. *Immunobiology* **164:**184-191

Mayer WE, Jonker M, Klein D, Ivanyi P, Van Seventer G, Klein J (1988) Nucleotide sequences of chimpanzee MHC class I alleles: evidence for trans-species mode of evolution. *EMBO* **7**:2765-2774

McAdam RA, Goundis D, Reid KB (1988) A homozygous point mutation results in a stop codon in the C1q B-chain of a C1q-deficient individual. *Immunogenetics* **27:**259-264

McConnell TJ, Talbot WS, McIndoe RA, Wakeland EK (1988) The origin of MHC class II gene polymorphism within the genus *Mus*. *Nature* **332**:651-654

Meddings J, Grennan DM (1980) The prevalence of systemic lupus erythematosus (SLE) in Dunedin. *N Z Med J* **91:**205-206

Medicus RG, Esser AF, Fernandez HN, Muller Eberhard HJ (1980) Native and activated properdin: interconvertibility and identity of amino- and carboxy-terminal sequences. *J Immunol* **124:**602-606

Medicus RG, Gotze O, Muller Eberhard HJ (1976) The serine protease nature of the C3 and C5 convertases of the classical and alternative complement pathways. *Scand J Immunol* **5:**1049-1055

Medof ME, Lublin DM, Holers VM, Ayers DJ, Getty RR, Leykam JF, Atkinson JP, Tykocinski ML (1987) Cloning and characterisation of cDNAs encoding the complete sequence of decay-accelerating factor of human complement. *Proc Natl Acad Sci U S A* **84:**2007-2011

Mellor AL, Weiss EH, Ramachandran K, Flavell RA (1983) A potential donor gene for the bm1 gene conversion event in the C57BL mouse. *Nature* **306:**792-795

Merril CR, Goldman D, Sedman SA, Ebert MH (1981) Ultrasensitive stain for proteins in polyacrylamide gels shows regional variation in cerebrospinal fluid proteins. *Science* **211:**1437-1438

Michaud J, Brody LC, Steel G, Fontaine G, Martin LS, Valle D, Mitchell G (1992) Strand-separating conformational polymorphism analysis: efficacy of detection of point mutations in the human ornithine delta -aminotransferase gene. *Genomics* **13:**389-394

Miller ME, Nilsson UR (1970) A familial deficiency of the phagocytosis-enhancing activity of serum related to a dysfunction of the fifth component of complement (C5). *N Engl J Med* **282:**354-358

Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* **16:**1215

Monahan JB, Sodetz JM (1981) Role of the β-subunit in the interaction of the eigth component of human complement with the membrane-bound cytolytic complex. *J Biol Chem* **256**:3258:3262

Moore LA, Tidyman WE, Arrizubieta MJ, Bandman E (1992) Gene conversion within the skeletal myosin multigene family. *J Mol Biol* **223**:383-387

Morgan BP, Campbell AK, Luzio JP, Hallett MB (1984) Recovery of polymorphonuclear leucocytes from complement attack. *Biochem Soc Trans* **12**:779-784

Morgan BP, Campbell AK (1985) The recovery of human polymorphonuclear leucocytes from sublytic complement attack is mediated by changes in intracellular free calcium. *Biochem J* **231**:205-208

Morgan BP, Dankert JR, Esser AF (1987) Recovery of human neutrophils from complement attack:removal of the membrane attack complex by endocytosis and exocytosis. *J Immunol* **138**:246-253

Morgan BP, Daniels RH, Watts MJ, Williams BD (1988) In vivo and in vitro evidence of cell recovery from complement attack in rheumatoid synovium. *Clin Exp Immunol* **73**:467-472

Morgan BP, Vora JP, Bennett AJ, Thomas JP, Matthews N (1989) A case of hereditary combined deficiency of complement components C6 and C7 in man. *Clin Exp Immunol* **75**:396-401

Morgan BP, Walport MJ (1991) Complement deficiency and disease. *Immunol Today* **12**:301-306

Muller Eberhard HJ (1986) The membrane attack complex of complement. *Annu Rev Immunol* **4**:503-528

Myers EM, Miller W (1988) *CABIOS* **4**:11-17

Myers RM, Larin Z, Maniatis T (1985) Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA:DNA duplexes. *Science* **230**:1242-1246

Nagasawa S, Stroud RM (1977) Cleavage of C2 by C1s into the antigenically distinct fragments C2a and C2b: demonstration of binding of C2b to C4b. *Proc Natl Acad Sci U S A* **74**:2998-3001

Nagata M, Hara T, Aoki T, Mizuno Y, Akeda H, Inaba S, Tsumoto K, Ueda K (1989) Inherited deficiency of ninth component of complement: an increased risk of meningococcal meningitis. *J Pediatr* **114**:260-264

Nakamura S, Ooue O, Abe K (1984) Genetic polymorphism of the seventh component of complement in a Japanese population. *Hum Genet* **66**:279-281

Nakamura S, Ooue O, Akiyama K, Abe K (1984) Genetic polymorphism of complement C6 and haplotype analysis between C6 and C7 in a Japanese population. *Hum Genet* **68:**138-141

Nakano Y, Hashimoto M, Choi NH, Sugita Y, Tobe T, Mazda T, Tomita M (1991) Functional and structural domains of the sixth component (C6) of human complement. *Chem Pharm Bull Tokyo* **39:**432-436

Nasmyth KA (1982) The regulation of yeast mating-type chromatin structure by S. *Cell* **30:**567-578

Ng SC, Rao AG, Howard OM, Sodetz JM (1987) The eighth component of human complement: evidence that it is an oligomeric serum protein assembled from products of three different genes. *Biochemistry* **26:**5229-5233

Nilsson UR, Miller ME, Wyman S (1974) A functional abnormality of the fifth component of complement (C5) from human serum of individuals with a familial opsonic defect. *J Immunol* **112:**1164-1176

Nishimukai H, Kitamura H, Kishida T, Tamaki Y (1985) C6 polymorphism in Japanese: typing by agarose gel isoelectric focusing-immunofixation. *Hum Hered* **35:**30-33

Nishimukai H, Tamaki Y (1986) Genetic polymorphism of the seventh component of complement: a new variant. *Vox Sang* **51:**60-62

Nonaka M, Natsuume Sakai S, Takahashi M (1981) The complement system in rainbow trout (Salmo gairdneri). II. Purification and characterization of the fifth component (C5). *J Immunol* **126:**1495-1498

Nonaka M, Iwaki M, Nakai C, Nozaki M, Kaidoh T, Natsuume Sakai S, Takahashi M (1984a) Purification of a major serum protein of rainbow trout (Salmo gairdneri) homologous to the third component of mammalian complement. *J Biol Chem* **259:**6327-6333

Nonaka M, Fujii T, Kaidoh T, Natsuume Sakai S, Yamaguchi N, Takahashi M (1984b) Purification of a lamprey complement protein homologous to the third component

of the mammalian complement system. *J Immunol* **133:**3242-3249

Ohta T (1980) Linkage disequilibrium between amino acid sites in immunoglobulin genes and other multigene families. *Genet Res* **36:**181-197

Olving JH, Olaisen B, Teisberg P, Gedde-Dahl T, Thorsby E (1977) Nonlinkage between C6 and chromosome 6 markers. *Hum Genet* **37**:125-129

Olving JH, Teisberg P, Olaisen B (1980) Polymorphism of the sixth component of complement (C6) in Norwegian Lapps. *Hum Hered* **30:**211-214

Orita M, Suzuki Y, Sekiya T, Hayashi K (1989) Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* **5:**874-879

Orita M, Sekiya T, Hayashi K (1990) DNA sequence polymorphisms in Alu repeats. *Genomics* **8:**271-278

Orren A, Hobart MJ, Nash HR, Lachmann PJ (1985) Close linkage between mouse genes determining the two forms of complement component C6 and component C7, and cis action of a C6 regulatory gene. *Immunogenetics* **21:**591-599

Orren A, Potter PC, Cooper RC, du Toit E (1987) Deficiency of the sixth component of complement and susceptibility to Neisseria meningitidis infections: studies in 10 families and five isolated cases. *Immunology* **62:**249-253

Orren A, Wurzner R, Potter PC, Fernie BA, Coetzee S, Morgan BP, Lachmann PJ (1992) Properties of a low molecular weight complement component C6 found in human subjects with subtotal C6 deficiency. *Immunology* **75:**10-16

Palsdottir A, Fossdal R, Jensson O, Arnason A (1987) Bgl II polymorphic site upstream to the human complement component C4A gene. *Nucleic Acids Res* **15:**2395

Parham P, Lawlor DA, Lomen CE, Ennis PD (1989) Diversity and diversification of HLA-A,B,C alleles. *J Immunol* **142:**3937-3950

Park KS, Yanagisawa M, Tokunaga K, Omoto K (1988) Genetic polymorphisms of complement components C6 and C7 in Korean. *Jinrui Idengaku Zasshi* **33:**325-331

Partanen J, Koskimies S (1986) Human MHC class III genes, Bf and C4. Polymorphism, complotypes and association with MHC class I genes in the Finnish population. *Hum Hered* **36:**269-275

Partanen J, Campbell RD (1989) Restriction fragment analysis of non-deleted complement C4 null genes suggests point mutations in C4 null alleles, but gene conversion in C4B null alleles. *Immunogenetics* **30:**520-523

Passas CM, Wong RL, Peterson M, Testa MA, Rothfield NF (1985) A comparison of the specificity of the 1971 and 1982 American Rheumatism Association criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* **28:**620-623

Patthy L (1985) Evolution of the proteases of blood coagulation and fibrinolysis by assembly from modules. *Cell* **41:**657-663

Patthy L (1987) Intron-dependent evolution: preferred types of exons and introns. *FEBS Lett* **214:**1-7

Pease LR, Schulze DH, Pfaffenbach GM, Nathenson SG (1983) Spontaneous H-2 mutants provide evidence that a copy mechanism analogous to gene conversion generates polymorphism in the major histocompatibility complex. *Proc Natl Acad Sci U S A* **80:**242-246

Pervaiz S, Brew K (1987) Homology and structure-function correlations between alpha 1-acid glycoprotein and serum retinol-binding protein and its relatives. *FASEB J* **1:**209-214

Petersen BH, Lee TJ, Snyderman R, Brooks GF (1979) Neisseria meningitidis and Neisseria gonorrhoeae bacteremia associated with C6, C7, or C8 deficiency. *Ann Intern Med* **90:**917-920

Petes T, Fink GR (1982) Gene conversion between repeated genes. *Nature* **300**:216-217

Petrella EC, Bredehost R, Vogel CW (1989) Purification of cobra C3: Initial characterisation and comparison to cobra venom factor (Abstract). *Compl Inflamm* **6**:386-387

Podack ER, Kolb WP, Esser AF, Muller Eberhard HJ (1979) Structural similarities between C6 and C7 of human complement. *J Immunol* **123:**1071-1077

Podack ER, Preissner KT, Muller Eberhard HJ (1984) Inhibition of C9 polymerization within the SC5b-9 complex of complement by S-protein. *Acta Pathol Microbiol Immunol Scand Suppl* **284:**89-96

Potter PC, Frasch CE, van der Sande WJ, Cooper RC, Patel Y, Orren A (1990) Prophylaxis against Neisseria meningitidis infections and antibody responses in patients with deficiency of the sixth component of complement. *J Infect Dis* **161:**932-937

Prentice HL, Schneider PM, Strominger JL (1986) C4B gene polymorphism detected in a human cosmid clone. *Immunogenetics* **23:**274-276

Radcliffe-Brown AR (1930) The aboriginal population: former numbers and distribution of the Australian aborigines. In:Official Yearbook of the Commonwealth of Australia, number 23, Government printers, Melbourne

Radding CM (1978) Genetic recombination: strand transfer and mismatch repair. *Annu Rev Biochem* **47:**847-880

Ramm LE, Whitlow MB, Koski CL, Shin ML, Mayer MM (1983) Elimination of complement channels from the plasma membranes of U937, a nucleated mammalian cell line J Immunol 131:1411-1419

Ranford PR, Kirk RL, Zimmet P (1982) Distribution of complement factors Bf, C2 and C6 in the Western Pacific. *Acta Anthropgenetica* **6**:23-32

Ranford PR (1989) Genetic variants of the serum complement components. In: The colonization of the Pacific A genetic trail Ed: Hill AVS, Serjeantson SW, Oxford Science publications, Oxford pp174-193

Rao AG, Howard OM, Ng SC, Whitehead AS, Colten HR, Sodetz JM (1987) Complementary DNA and derived amino acid sequence of the alpha subunit of human complement protein C8: evidence for the existence of a separate alpha subunit messenger RNA. *Biochemistry* **26**:3556-3564

Raum D, Spence MA, Balavitch D, Tideman S, Merritt AD, Taggart RT, Petersen BH, Day NK, Alper CA (1979) Genetic control of the eighth component of complement. *J Clin Invest* **64**:858-865

Raum D, Awdeh Z, Anderson J, Strong L, Granados J, Teran L, Giblett E, Yunis EJ, Alper CA (1984) Human C4 haplotypes with duplicated C4A or C4B. *Am J Hum Genet* **36**:72-79

Rechavi G, Ram D, Glazer L, Zakut R, Givol D (1983) Evolutionary aspects of immunoglobulin heavy chain variable region (VH) gene subgroups. *Proc Natl Acad Sci U S A* **80**:855-859

Reed KC, Mann DA (1985) Rapid transfer of DNA from agarose gels to nylon membranes. *Nucleic Acids Res* **13**:7207-7221

Reid KBM (1985) Molecular cloning and characterisation of the complementary DNA and gene encoding for the B-chain of subcomponent C1q of the human complement system. *Biochem J* **231**:729-735

Reid KBM, Bentley DR, Campbell RD, Chung LP, Sim RB, Kristensen T, Tack BF (1986) Complement system proteins which interact with C3b or C4b: A superfamily of structurally related proteins. *Immunol Today* **7**:230-235

Reid KBM (1988) The complement system In: Molecular Immunology Eds: Hames BD, Glover DM IRL Press, Oxford pp 189-241

Reid KBM (1989) Deficiency of the first component of human complement. *Immunodef Rev* **1**:242-260

Reinitz E, Lawrence M, Diamond B, Keiser H, Alper C (1986) Arthritis and antinuclear antibodies (ANA) with inherited deficiency of the sixth component of complement (C6). *Ann Rheum Dis* **45**:431-434

Rey Campos J, Rubinstein P, Rodriguez de Cordoba S (1987) Decay-accelerating factor. Genetic polymorphism and linkage to the RCA (regulator of complement activation) gene cluster in humans. *J Exp Med* **166**:246-252

Richardson B, Scheinbart L, Strahler J, Gross L, Hanash S, Johnson M (1990) Evidence for impaired T cell DNA methylation in systemic lupus erythematosus and rheumatoid arthritis. *Arthritis Rheum* **33**:1665-1673

Ripley LS (1982) Model for the participation of quasi -palindromic DNA sequences in frameshift mutation. *Proc Natl Acad Sci U S A* **79**:4128-4132

Rittner cg, Dewald E, Berghoff E, Mollenhauer E (1979) C6 polymophismus der sechsten komplementkomponente. *Z Rechstmed* **83**:17-25

Rittner C, Giles CM, Roos MH, Demant P, Mollenhauer E (1984a) Genetics of human C4 polymorphism: detection and segregation of rare and duplicated haplotypes. *Immunogenetics* **19**:321-333

Rittner C, Hargesheimer W, Mollenhauer E (1984b) Population and formal genetics of the human C81(alpha-gamma) polymorphism. *Hum Genet* **67**:166-169

Rittner C, Stradman-Bellinghausen (1993) Human complement C81 (C8A) polymorphism: detection and segregation of new variants. *Hum Genet* **92**:413-416

Roberts RG, JOnes R, Smith MA (1990) Thermoluminesence dating of a 50 000 year old human occupation site in northern Australia. *Nature* **345**:153-156

Rodriguez de Cordoba S, Dykman TR, Ginsberg Fellner F, Ercilla G, Aqua M, Atkinson JP, Rubinstein P (1984) Evidence for linkage between the loci coding for the binding protein for the fourth component of human complement (C4BP) and for the C3b/C4b receptor. *Proc Natl Acad Sci U S A* **81**:7890-7892

Rodriguez de Cordoba S, Lublin DM, Rubinstein P, Atkinson JP (1985) Human genes for three complement components that regulate the activation of C3 are tightly linked. *J Exp Med* **161**:1189-1195

Rogde S, Mevag B, Teisberg P, Gedde-Dahl Jr T, Tedesco F, Olaisen B (1985) Genetic polymorphism of complement component C8. *Hum Genet* **70**:211-216

Rogde S, Mevag B, Teisberg P, Olaisen B, Sodetz J (1989a) A BamH1 RFLP of the C8B gene. *Nucleic Acids Res* **17**:6760

Rogne S, Myklebost O, Stanley K, Geurts van Kessel A (1989b) The gene for human complement C9 is on chromosome 5. *Genomics* **5**:149-152

Rogne S, Mevag B, Gedde Dahl T, Myklebost O (1990) Multiple RFLPs of human complement component nine (C9) detected by TaqI. *Nucleic Acids Res* **18**:3112

Rogde S, Olaisen B, Teisberg P, Sodetz J (1991a) A BamHI RFLP of the C8A gene. *Nucleic Acids Res* **19**:3762

Rogne S, Myklebost O, Olving JH, Kyrkjebo HT, Jonassen R, Olaisen B, Gedde Dahl T, Jr. (1991b) The human genes for complement components 6 (C6) and 9 (C9) are closely linked on chromosome 5. *J Med Genet* **28**:587-590

Rosenfeld SI, Baum J, Steigbigel RT, Leddy JP (1976) Hereditary deficiency of the fifth component of complement in man. II. Biological properties of C5-deficient human serum. *J Clin Invest* **57**:1635-1643

Rosenfeld SI, Weitkamp LR, Ward F (1977) Hereditary C5 deficiency in man: genetic linkage studies. *J Immunol* **119:**604-608

Ross GD, Jensen JA (1973) The first component (C1n) of the complement system of the nurse shark (Ginglymostoma cirratum). I. Hemolytic characteristics of partially purified C1n. *J Immunol* **110:**175-182

Ross SC, Densen P (1984) Complement deficiency states and infection: epidemiology, pathogenesis and consequences of neisserial and other infections in an immune deficiency. *Medicine Baltimore* **63:**243-273

Rother K (1986) Rabbits deficient in C6. *Prog Allergy* **39:**192-201

Rothfield N (1985) Clinical features of systemic lupus erythematosus. In: Textbook of Rheumatology, 2nd edn, Ed: Kelley WN, Harris ED, Ruddy S, Sledge CB Saunders,WB , pp1070-1077

Sagai T, Sakaizumi M, Miyashita N, Bonhomme F, Petras ML, Nielsen JT, Shiroishi T, Moriwaki K (1989) New evidence for trans-species evolution of the H-2 class I polymorphism. *Immunogentics* **30**:89-98

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N (1985) Enzymatic amplification of beta -globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230:**1350-1354

Sakano T, Hamasaki T, Mori M, Ohta T, Ueda K, Ishigame K, Kawamura M, Takeda N (1988) C7 deficiency and persistent haematuria. *Eur J Pediatr* **147:**516-517

Sarkar G, Yoon HS, Sommer SS (1992) Screening for mutations by RNA single-strand conformation polymorphism (rSSCP): comparison with DNA-SSCP. *Nucleic Acids Res* **20:**871-878

Satta Y (1992) Balancing selection at HLA loci. In: Population paleo genetics, Ed: Takahota N, Japan Scientific Society Press, Tokyo pp129-149

Sawyer S (1989) Statistical tests for detecting gene conversion. *Mol Biol Evol* **6**:526-538

Scacchi R, Corbo RM, Cossu G, Mureddu L, Mulas G, Pascone R (1992) Distribution of ORM1, C6, C7 and APO C-II allele frequencies in populations from mainland Italy and Sardinia. *Hum Hered* **42**:309-315

Schendel DJ, O'Neill GJ, Wank R (1984) MHC-linked class III genes. Analysis of C4 gene frequencies, complotypes and associations with distinct HLA haplotypes in German Caucasians. *Immunogenetics* **20**:23-31

Schneider PM, Carroll MC, Alper CA, Rittner C, Whitehead AS, Yunis EJ, Colten HR (1986) Polymorphism of the human complement C4 and steroid 21-hydroxylase genes. Restriction fragment length polymorphisms revealing structural deletions, homoduplications, and size variants. *J Clin Invest* **78**:650-657

Schur PH (1978) Genetics of complement deficiencies associated with lupus-like syndromes. *Arthritis Rheum* **21**:S153-S160

Schur PH (1982) Complement and lupus erythematosus. *Arthritis Rheum* **25**:793-798

Segurado OG, Arnaiz Villena AA, Iglesias Casarrubios P, Martinez Laso J, Vicario JL, Fontan G, Lopez Trascasa M (1992) Combined total deficiency of C7 and C4B with systemic lupus erythematosus (SLE). *Clin Exp Immunol* **87**:410-414

Sellar GC, Goundis D, McAdam RA, Soloman E, Reid KBM (1987) *Complement* **4**:225.

Sellar GC, Blake DJ, Reid KB (1991) Characterization and organization of the genes encoding the A-, B- and C -chains of human complement subcomponent C1q. The complete derived amino acid sequence of human C1q. *Biochem J* **274**:481-490

Serjeantson SW, Hill AVS (1989) The colonization of the Pacific: the genetic evidence. In: The colonization of the Pacific: a genetic trail, Ed: Hill AVS, Serjeantson SW, Oxford University press, Oxford pp286-294

Setien F, Alvarez V, Coto E, DiScipio RG, Lopez Larrea C (1993) A physical map of the human complement component C6, C7, and C9 genes. *Immunogenetics* **38**:341-344

Sheffield VC, Beck JS, Kwitek AE, Sandstrom DW, Stone EM (1993) The sensitivity of single-strand conformation polymorphism analysis for the detection of single base substitutions. *Genomics* **16**:325-332

Sheffield VC, Cox DR, Lerman LS, Myers RM (1989) Attachment of a 40-base-pair G + C-rich sequence (GC -clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. *Proc Natl Acad Sci U S A* **86**:232-236

Shibata T, Akita T, Abe T (1993) Genetic polymorphism of the sixth component of complement (C6) in the pig. *Anim Genet* **24**:97-100

Shinkai Y, Takio K, Okumura K (1988) Homology of perforin to the ninth component of complement (C9). *Nature* **334**:525-527

Sigurdardottir S, Borsch C, Gustafsson K, Andersson L (1992) Gene duplications and sequence polymorphism of bovine class II DQB genes. *Immunogenetics* **35**: 205 -213

Sim R B, Arlaud GJ, Colomb MG (1979) C1 inhibitor -dependent dissociation of human complement component C1 bound to immune complexes. *Biochem J* **179**:449-457

Sitomer G, Stroud RM, Mayer MM (1966) Reversible adsorption of C'2 by EAC'4: role of Mg2+, enumeration of competent SAC'4, two-step nature of C'2a fixation and estimation of its efficiency. *Immunochemistry* **3**:57-69

Sjoholm AG (1991) Inherited complement deficiency states and disease. *Complement Inflamm* **8**:341-346

Slightom JL, Blechl AE, Smithies O (1980) Human fetal G gamma- and A gamma-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell* **21:**627-638

Smith CA, Pangburn MK, Vogel CW, Muller Eberhard HJ (1984) Molecular architecture of human properdin, a positive regulator of the alternative pathway of complement. *J B iol Chem* **259:**4582-4588

Sottrup Jensen L, Stepanik TM, Kristensen T, Lonblad PB, Jones CM, Wierzbicki DM, Magnusson S, Domdey H, Wetsel RA, Lundwall A, et al (1985) Common evolutionary origin of alpha 2-macroglobulin and complement components C3 and C4. *Proc Natl Acad Sci U S A* **82:**9-13

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* **98:**503-517

Springer TA, Dustin ML, Kishimoto TK, Marlin SD (1987) The lymphocyte function-associated LFA-1, CD2, and LFA-3 molecules: cell adhesion receptors of the immune system. *Annu Rev Immunol* **5:**223-252

Stanley KK, Kocher HP, Luzio JP, Jackson P, Tschopp J (1985) The sequence and topology of human complement component C9. *EMBO J* **4:**375-382

Steckel EW, York RG, Monahan JB, Sodetz JM (1980) The eighth component of human complement. Purification and physicochemical characterization of its unusual subunit structure. *J Biol Chem* **255:**11997-12005

Steinberg AD (1985) Systemic Lupus Erythematosus. In: Cecil Textbook of Medicine, Ed; Wyngaarden JB, Smith LH, Philadelphia: pp1924-1932

Stephens JC (1985) Statistical methods of DNA sequence analysis: detection of intragenic recombination or gene conversion. *Mol Biol Evol* **2:**539-556

Suggs SV, Hirose T, Miyake T, Kawashima EH, Johnson MJ, Hakura K, Wallace RB (1981) In: Developmental biology using purified genes, Ed: Brow D, Academic Press, New York

Takahata N (1990) A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. *Proc Natl Acad Sci USA* **87**:2419-2423

Takahata N, Nei M (1990) Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics* **124**:967-978

Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield NF, Schaller JG, Talal N, Winchester RJ (1982) The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* **25**:1271-1277

Tedesco F, Silvani CM, Agelli M, Giovanetti AM, Bombardieri S (1981) A lupus-like syndrome in a patient with deficiency of the sixth component of complement. *Arthritis Rheum* **24**:1438-1440

Tedesco F, Densen P, Villa MA, Petersen BH, Sirchia G (1983) Two types of dysfunctional eighth component of complement (C8) molecules in C8 deficiency in man. Reconstitution of normal C8 from the mixture of two abnormal C8 molecules. *J Clin Invest* **71**:183-191

Theriault A, Boyd E, Whaley K, Sodetz JM, Connor JM (1992) Regional chromosomal assignment of genes encoding the alpha and beta subunits of human complement protein C8 to 1p32. *Hum Genet* **88**:703-704

Thompson EA, Deeb S, Walker D, Motulsky AG (1988) The detection of linkage disequilibrium between closely linked markers: RFLPs at the AI-CIII apolipoprotein genes. *Am J Hum Genet* **42**:113-124

Thorne AG (1980) The arrival of man in Australia In: The Cambridge encyclopedia of archaeology, Ed: Sheratt A, Cambridge University Press pp96-100

Tindall KR, Kunkel TA (1988) Fidelity of DNA synthesis by the Thermus aquaticus DNA polymerase. *Biochemistry* **27**:6008-6013

Tokunaga K, Yukiyma Y, Omoto K (1983) Polymorphism of the complement component C6 in Japanese. *J Immunogenet* **10**:419-424

Tokunaga K, Yamamura N, Omoto K (1984) An immunoblotting technique for complement C6 typing: three new variants. *Jinrui Idengaku Zasshi* **29**:415-419

Tokunaga K, Omoto K, Akaza T, Akiyama N, Amemiya H, Naito S, Sasazuki T, Satoh H, Juji T (1985) Haplotype study on C4 polymorphism in Japanese. Associations with MHC alleles, complotypes, and HLA-complement haplotypes. *Immunogenetics* **22**:359-365

Tokunaga K, Dewald G, Omoto K, Juji T (1986) Family study on the polymorphisms of the sixth and seventh components (C6 and C7) of human complement: linkage and haplotype analyses. *Am J Hum Genet* **39**:414-419

Tokunaga K, Jin F, Mauff G, Nishimukai H, Simeoni E, Suzuki K, Weber W (1990) C6 reference typing report. *Complement Inflamm* **7**:234-239

Tosi M, Duponchel C, Meo T, Julier C (1987) Complete cDNA sequence of human complement Cls and close physical linkage of the homologous genes Cls and Clr. *Biochemistry* **26**:8516-8524

Trapp RG, Mooney E, Coleman TH, Forristal J, Herman JH (1987) Hereditary complement (C6) deficiency associated with systemic lupus erythematosus, Sjogren's syndrome and hyperthyroidism. *J Rheumatol* **14**:1030-1033

Triggs Raine BL, Akerman BR, Clarke JT, Gravel RA (1991) Sequence of DNA flanking the exons of the HEXA gene, and identification of mutations in Tay-Sachs disease. *Am J Hum Genet* **49**:1041-1054

Tsunenari S, Kibayashi K, Higashi T, Pang H, Ding M, Jia J (1992) Genetic polymorphism of complement component C6, C7 and C8(1) in Chinese Han population in northeast China. *Nippon Hoigaku Zasshi* **46**:469-473

Urabe K, Kimura A, Harad F, Iwanaga T, Sasazuki T (1990) Gene conversion in steroid 21-hydroxylase genes. *Am J Hum Genet* **46**:1178:1186

Uring Lambert B, Goetz J, Tongio MM, Mayer S, Hauptmann G (1984) C4 haplotypes with duplications at the C4A or C4B loci: frequency and associations with BF, C2, and HLA-A, B, C, DR alleles. *Tissue Antigens* **24**:70-72

Valet G, Cooper NR (1974) Isolation and characterization of the proenzyme form of the C1s subunit of the first complement component. *J Immunol* **112**:339-350

van Oorschot RA, Birmingham V, Porter PA, Kammerer CM, Vandeberg JL (1993) Linkage between complement components 6 and 7 and glutamic pyruvate transaminase in the marsupial Monodelphis domestica. *Biochem Genet* **31**:215-222

Vaz-Guedes MA, Hobart MJ, Lachmann PJ (1978) Absence of variation in human C5. *J Immunogenetics* **5**:279-282

Vogt W, Dames W, Schmidt G, Dieminger L (1977) Complement activation by the properdin system: formation of a stoichiometric. C3 cleaving complex of properdin factor B with C36. *Immunochemistry* **14**:201-205

Vogt W, Schmidt G, Von Buttlar B, Dieminger L (1978) A new function of the activated third component of complement: binding to C5, an essential step for C5 activation. *Immunology* **34**:29-40

Washio K, Tokunaga K, Omoto K, Misawa S (1986) Human C7 polymorphism: classification and association analysis with C6. *Jinrui Idengaku Zasshi* **31**:345-352

Weiler JM, Daha MR, Austen KF, Fearon DT (1976) Control of the amplification convertase of complement by the plasma protein beta1H. *Proc Natl Acad Sci U S A* **73**:3268-3272

Weir BS, Wilson SR (1986) Log-linear models for linked loci. *Biometrics* **42**:665-670

Weis JH, Morton CC, Bruns GA, Weis JJ, Klickstein LB, Wong WW, Fearon DT (1987) A complement receptor locus: genes encoding C3b/C4b receptor and C3d/Epstein-Barr virus receptor map to 1q32. *J Immunol* **138**:312-315

Weis JJ, Toothaker LE, Smith JA, Weis JH, Fearon DT (1988) Structure of the human B lymphocyte receptor for C3d and the Epstein-Barr virus and relatedness to other members of the family of C3/C4 binding proteins [published erratum appears in J Exp Med 1988 Nov 1;168(5):1953-4]. *J Exp Med* **167**:1047-1066

Wetsel RA, Lemons RS, Le Beau MM, Barnum SR, Noack D, Tack BF (1988) Molecular analysis of human complement component C5: localization of the structural gene to chromosome 9. *Biochemistry* **27**:1474-1482

Whaley K, Ruddy S (1976) Modulation of the alternative complement pathway by bIH globulin *J Exp Med* **144**:1147-1163

Whitehead AS, Solomon E, Chambers S, Bodmer WF, Povey S, Fey G (1982) Assignment of the structural gene for the third component of human complement to chromosome 19. *Proc Natl Acad Sci U S A* **79**:5021-5025

Whitehouse DB (1982) Genetic variation of the sixth component of complement (C6) in the Manx shearwater (Puffinus p. puffinus). *Anim Blood Groups Biochem Genet* **13**:179-187

Whitehouse DB (1984) Genetic polymorphism and linkage of the sixth and seventh complement components (C6 and C7) in the common marmoset. *Biochem Genet* **22**:51-63

Whitehouse DB, Putt W (1983) Immunological detection of the sixth complement component (C6) following flat bed polyacrylamide gel isoelectric focusing and electrophoretic transfer to nitrocellulose filters. *Ann Hum Genet* **47 Pt 1**:1-8

Whitney JB, Copland GT, Skow LC, Russell ES (1979) Resolution of products of the duplicated hemoglobin alpha-chain loci by isoelectric focusing. *Proc Natl Acad Sci U S A* **76**:867-871

Wisnieski JJ, Naff GB, Pensky J, Sorin SB (1985) Terminal complement component deficiencies and rheumatic disease: development of a rheumatic syndrome and anticomplementary activity in a patient with complete C6 deficiency. *Ann Rheum Dis* **44**:716-722

Wolfe KH, Sharp PM, Li WH (1989) Mutation rates differ among regions of the mammalian genome. *Nature* **337**:283-285

Wuillemin WA, Spath PJ, Uring Lambert B, Straub PW (1991) Clinical manifestations in humans of combined C7 and C4 deficiency associated with low levels of C2, C8, and C9. *Complement Inflamm* **8**:70-79

Wurzner R, Nitze R, Gotze O (1990) C7*9, a new frequent C7 allele detected by an allotype-specific monoclonal antibody. *Complement Inflamm 7:290-297*

Wurzner R, Orren A, Potter P, Morgan BP, Ponard D, Spath P, Brai M, Schulze M, Happe L, Gotze O (1991) Functionally active complement proteins C6 and C7 detected in C6-and C7-deficient individuals. *Clin Exp Immunol* **83**:430-437

Wurzner R, Orren A, Lachmann PJ (1992a) Inherited deficiencies of the terminal components of human complement. *Immunodefic Rev* **3**:123-147

Wurzner R, Hobart MJ, Orren A, Tokunaga K, Nitze R, Gotze O, Lachmann PJ (1992b) A novel protein polymorphism of human complement C7 detected by a monoclonal antibody. *Immunogenetics* **35**:398-402

Wurzner R, Mewar D, Lachmann PJ (1993) The dysmorphic but functionally active C6 of subtotal deficient subjects is carboxyterminally truncated (Abstract) *Mol Immunol* **30**:63

Wysocki LJ, Gefter ML (1989) Gene conversion and the generation of antibody diversity. *Annu Rev Biochem* **58**:509-531

Yamamoto T, Davis CG, Brown MS, Schneider WJ, Casey ML, Goldstein JL, Russell DW (1984) The human LDL receptor: a cysteine-rich protein with multiple Alu sequences in its mRNA. *Cell* **39**:27-38

York LJ, Marshall WH, Huang SN (1986) Polymorphism of the seventh complement component, C7, in Chinese. *Hum Hered* **36**:261-262

Yu CY, Belt KT, Giles CM, Campbell RD, Porter RR (1986) Structural basis of the polymorphism of human complement components C4A and C4B: gene size, reactivity and antigenicity. *EMBO J* **5**:2873-2881

Yu CY, Campbell RD (1987) Definitive RFLPs to distinguish between the human complement C4A/C4B isotypes and the major Rodgers/Chido determinants: application to the study of C4 null alleles. *Immunogenetics* **25**:383-390

Yukiyama Y, Tokunaga K, Yoshida K, Miyamoto T (1985) [Studies of genetic polymorphism of complement C2, B and C6 in systemic lupus erythematosus]. *Ryumachi* **25**:274-277

Zeitz HJ, Miller GW, Lint TF, Ali MA, Gewurz H (1981) Deficiency of C7 with systemic lupus erythematosus: solubilization of immune complexes in complement -deficient sera. *Arthritis Rheum* **24**:87-93

Zeng ZM, Tokunaga K, Omoto K, Du CS (1986) Genetic polymorphisms of complement C6 and C7 in two Chinese populations. *Jinrui Idengaku Zasshi* **31**:263-271

Zhang Z, Du RF (1990) [Studies on polymorphism of C6 in Han nationality of China]. *I Chuan Hsueh Pao* **17**:243 -248

Ziccardi RJ, Cooper NR (1979) Active disassembly of the first complement component, C-1, by C-1 inactivator. *J Immunol* **123**:788-792