Molecular Basis of C4 Protein Deficiency in Aboriginal Australians, and a Molecular C4 Allotyping Technique

by

Joanne Lee Banyer

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

November 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 November 1990 to November 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.

Joanne Banyer
Acknowledgements

There are a number of people I would like to acknowledge, for without their contributions, this thesis would not have been possible.

I would like to thank Dr Paul Gatenby, Dr Anthony Veale, Dr Kuldeep Bhatia, Dr Bart Currie and Dr Tim Burt for providing Aboriginal Australian, PNG and Caucasian samples for analysis. Consent for scientific analysis of Aboriginal Australian blood specimens was given by Kimberley region participants to M. Blake (JCSMR, Australian National University) and R. Spargo (Community Health Services, Derby), and by Cape York Penninsula participants to A. Veale.

Thanks also go to Dr Heather Dunckley, Dr Pam Ranford, Dr Liz Jazwinska and Dr Xiaojiang Gao for their contributions of C4 allotyping and HLA typing. Thanks go to Ingrid Jacobson for her help with computing, and the Western Australian C4 reference standard typing laboratory for confirmation of particular donors C4 allotypes. Thanks to the Biomolecular Resource Facility for automated sequencing analysis, and the photography unit of JCSMR for the figures displayed in this thesis.

Thanks to my supervisor Prof. Sue Serjeantson for her kindness, guidance, support and encouragement throughout the past four years. A special thanks to Dr Liz Jazwinska for initially supervising this research, for seeing it to its completion, and for offering me the next and equally challenging task of my scientific career.

A very special thanks to Dr Colleen Olive for her friendship, advice, support and encouragement throughout the past four years. Also, for the time and effort given to the final stages of this thesis. Thanks goes to Mike Harrison, a new found friend, for his support, understanding and assistance through the most difficult stages of writing this thesis. Alot to give to someone he knows little of. Most importantly my thanks and gratitude go to the members of my family, for without their love, encouragement, support and faith in me, this thesis would have been a most difficult task to achieve.
Abstract

Since the initial discovery of the association between $C4$ protein deficiency and autoimmune disease, much effort has been devoted to the characterisation of $C4$ protein and its molecular composition (Dawkins et al., 1983). From the outset, researchers have found the $C4$ protein difficult to work with as it is unstable and degrades rapidly. However, with the advent of molecular techniques, studies of the $C4$ gene have accelerated, uncovering a very complicated and large duplicated gene region.

This thesis reviews the literature on complement component $C4$, and reports on three interrelated research areas involving protein and DNA analysis of complement $C4$. Contributions to the molecular characterisation of the $C4$ complement gene region and population genetics are discussed in relation to current knowledge, and direction for further research is provided.

This study set out to establish whether indigenous people of PNG and Aboriginal Australians have shared a common gene pool. By analysing the distribution of polymorphic $C4$ alleles these studies found the Aboriginal Australian and PNG populations to be distinctly different. Analysis of the distribution of $C4$ alleles in the two ethnic groups indicated that the two Aboriginal populations showed significant similarities but the three geographically distinct PNG populations were significantly different. Studies here suggested these ethnic populations are more reliably distinguished from each other by differences in the frequencies of the most frequently occurring $C4$ alleles, than by the incidence of rare alleles.

The high frequency of $C4$ null alleles in northern located Darwin Aboriginal's reported by Ranford et al. (1987), is not an isolated incident as this study also found the highest frequencies of $C4$ null alleles in northern located Aboriginal Australian's compared to any ethnic group studied. This could be associated with the high occurrence of the autoimmune disease Systemic Lupus Erythematosus (SLE) in Aboriginal Australians, and strong
susceptibility of Aboriginal Australians to this disease (Anstey et al., 1993).

Accurate identification of $C4$ alleles including $C4$ null alleles is essential both for investigating gene-linked disease associations, as well as for advancing future population genetics studies. Studies here demonstrate that $C4$ protein allotyping does not accurately identify the $C4$ genotype of an individual. This was highlighted by the lack of Hardy-Weinberg equilibrium with respect to $C4$ allele frequencies of the East Cape York Aboriginal population as well as the identification of a number of alleles by molecular analysis which could not be identified by protein typing. The difficulties associated with determining $C4$ allotypes by $C4$ protein gel electrophoresis include instability of the $C4$ protein, high degree of technical expertise for interpretation of allotypes, and limitations of this technique in distinguishing overlapping alleles.

It was established in these studies that the polymorphic $C4d$ region of $C4$ genes may be used by molecular means to identify and type $C4$ alleles. With the advent of molecular biological techniques such as protein and DNA sequencing, and RFLP analysis of the $C4$ gene region, information required for development of a molecular-based typing protocol has become available. Some of the previously described $C4A$ and $C4B$ alleles have been found to have unique combinations of polymorphic amino acid residues distributed within a short stretch of DNA, otherwise known as the $C4d$ region. This region of the gene is highly polymorphic encoding amino acid residues known to be involved in immune-complex binding, and also encodes antigenic determinants. Therefore the $C4d$ region was considered an ideal candidate for the development of a molecular-based allotyping protocol. The molecular-based $C4$ typing technique developed here is based on PCR-RFLP analysis of the $C4d$ region. Protein and DNA sequence information, $C4$ protein allotyping data, and computer analysis which estimates the predicted isoelectric value of a protein were used to develop this protocol. Application of the molecular $C4$ typing protocol enabled the identification of $C4$ alleles which co-migrate by $C4$ protein gel electrophoresis as well as a number of rare and novel $C4$ alleles.
The molecular basis of C4 null alleles has been extensively studied in Caucasians, however, very little is known about the genetic basis of C4 null alleles in other ethnic populations. C4 null alleles from Aboriginal Australians were investigated here in order to identify the genetic basis of these alleles compared to those in Caucasians. Many Caucasian C4 null alleles result from large deletions. However, this study found that the C4B null allele in Aboriginal Australians results from either duplication of C4A genes on a single haplotype, or possession of a hybrid C4A/C4B type of novel allele, which was denoted as C4A *CAN1. These findings have considerable significance in relation to the high frequency of C4 null alleles found in Aboriginal Australian populations. Many of these "null alleles" may indeed be functional alleles which have been misinterpreted by C4 protein typing. The molecular basis of C4A null alleles remains unresolved, however, this study indicates that the defect is likely to be due to post-transcriptional processing of the allele.

This thesis describes a detailed investigation into the molecular basis of C4 alleles including rare and null alleles. This information together with the vast amount of molecular data available on C4 enabled the development of a molecular-based C4 typing protocol. This protocol is considered to be a more reliable and accurate typing method than the traditional protein typing method, and can be easily adopted in any laboratory for routine allotyping.
Abstracts of papers presented at conferences


Oral Presentation- "Investigation of complement component $C4$ alleles in Australian Aboriginal populations"

Poster- in collaboration with Dr Kuldeep Bhatia. "Limited genic diversity in the highlands of Papua New Guinea extends to all three classes of the MHC loci".

Australasian Society of Immunology, 1993.

Oral Presentation- "Molecular genetic basis of complement component $C4$ deficiency in Aboriginal Australians".


Oral Presentation- "Development of a molecular $C4$ allotyping technique based on PCR-RFLP".
Dedication

With love for my mother and father, to whom I owe so much.
Table of Contents

Statement ii
Acknowledgements iii
Abstract iv
Abstracts of papers presented at conferences vii
Dedication viii
Table of Contents ix
List of Figures xvii
List of Tables xix
Abbreviations xxi

Chapter 1: Introduction 1

1.1 Introduction 1

1.2 The role of complement in the human immune response 2
1.2.1 Classical and alternative complement pathways 4
1.2.2 Terminal pathways 5
1.2.3 Inflammatory response 6
1.2.4 Regulatory proteins 6
1.2.5 Complement receptors 7
1.2.6 Additional roles of complement 8

1.3 C4 deficiency and systemic lupus erythematosus (SLE) 8

1.4 C4 deficiency and other disease states 11

1.5 C4 protein and origin of its synthesis 12

1.6 Detection of C4 protein 14

1.7 C4 protein as a polymorphic marker for population studies 16

1.8 Complement C4 genetics 17
1.8.1 Gene localisation 17
1.8.2 Structure of the C4 gene locus and surrounding gene region 18
1.8.3 Exon and intron structure of C4 20
1.8.4 C4 gene size variation 21
1.8.5 Molecular basis of C4A and C4B 21
1.8.6 Functional differences between \textit{C4A} and \textit{C4B} isotypes

1.8.7 Chido and Rodgers antigenic determinants

1.8.8 \textit{C4} polymorphism and method of detection

1.9 \textit{C4} analysis in other animals- purpose and findings

1.10 Evolution of the \textit{C4} gene region

1.10.1 Multiple copies of a \textit{C4} module

1.10.2 \textit{C4A} and \textit{C4B} isotypic residues and \textit{C4} gene size variation

1.10.3 Evolution of \textit{C4} polymorphism

1.10.3.1 Gene conversion mechanism

1.10.3.2 Gene conversion in mammals

1.10.3.3 Gene conversion and the \textit{C4/CYP21} module

1.10.4 Gene duplication and evolution of the \textit{C4} gene region

1.11 Molecular analysis of \textit{C4} null alleles

1.11.1 \textit{C4} deletions

1.11.2 \textit{C4A} gene mutation resulting from a small insertion

1.11.3 Isoexpression, homoeexpression and hybrid \textit{C4} null alleles

1.11.4 Association of \textit{C4} null alleles with HLA haplotypes

1.12 Summary of review material most relevant to research studies undertaken for this thesis

1.13 Hypothesis of studies for this thesis

\textbf{Chapter 2: Materials and Methods}

2.1 Materials

2.2 Methods

2.3 Study populations

2.4 Caucasian donors

2.5 \textit{C4} protein allotyping

2.5.1 Gel preparation

2.5.2 Dialysis

2.5.3 Electrophoresis
2.5.4 Immunoprecipitation and staining 53
2.5.5 Carboxypeptidase B treatment 53
2.6 Statistical analysis 54
2.7 Genomic DNA extraction 54
2.8 DNA digestion 54
2.9 Gel electrophoresis 55
2.10 Gel staining with EtBr and silver 55
2.11 Southern blotting 57
2.12 Radioactive labelling of DNA probes and hybridisation 57
  2.12.1 Nick translation of C4 pAT-A 57
  2.12.2 Hybridisation of nick translated probes 57
  2.12.3 Labelling 5' end of DNA oligonucleotides 58
  2.12.4 Hybridisation of 5' end-labelled oligonucleotides 58
2.13 Total RNA extraction 59
  2.13.1 RNA preparation 59
  2.13.2 Estimation of RNA concentration and protein contamination 59
  2.13.3 Electrophoresis of RNA 60
2.14 Northern blots and hybridisation 60
2.15 First strand cDNA synthesis 61
2.16 PCR amplification 61
2.17 Electro-elution of PCR fragments from agarose gels 62
2.18 Cloning of C4 specific DNA fragments 62
  2.18.1 Sticky-end cloning 62
  2.18.1.1 Preparation of PCR DNA 62
  2.18.1.2 Preparation of M13 cloning vectors 63
  2.18.2 Blunt-end cloning 63
  2.18.2.1 Preparation of PCR DNA 63
  2.18.2.2 Preparation of M13 cloning vectors 63
  2.18.3 Ligation of PCR insert and vector DNA 64
  2.18.4 Preparation of competent E.Coli cells 64
  2.18.5 Transformation of competent cells 64
  2.18.6 Selection of recombinant plaques 65
  2.18.7 Direct Gel Electrophoresis (DIGE), and hybridisation selection of recombinant clones 65
2.19 Automated DNA sequencing of recombinant clones 66
2.19.1 Preparation of single stranded DNA template

2.19.2 Estimation of concentration of ssDNA template

2.19.3 Single cycle sequencing

2.20 Computer analysis of derived $C4$ protein sequences to estimate their isoelectric point, denoted as 'PI' (Predicted Isoelectric Point)

Chapter 3: Investigation of complement $C4$ alleles in Aboriginal Australian and Papua New Guinea (PNG) populations

3.1 Introduction

3.2 Aim

3.3 Materials and Methods

3.3.1 Aboriginal Australian and PNG population donors

3.3.2 $C4$ protein allotyping

3.3.3 Estimation of $C4$ allelic frequencies

3.3.4 Statistical analysis

3.4 Results

3.4.1 $C4$ allotypes identified by protein gel electrophoresis and immunofixation

3.4.2 $C4$ allelic polymorphism

3.4.3 Testing the goodness of fit of the $C4$ allele frequencies to the Hardy-Weinberg Equilibrium

3.4.4 $C4$ gene frequency analysis

3.4.4.1 $C4A$ gene frequency analysis

3.4.4.2 $C4B$ gene frequency analysis

3.4.4.3 Significant differences in $C4$ gene frequencies between different Aboriginal Australian and PNG populations

3.4.5 Distribution of $C4A$ and $C4B$ allotypes within Aboriginal Australian populations

3.4.6 Distribution of $C4A$ and $C4B$ allotypes within three PNG populations
3.4.7 Significant differences of C4 allotype frequencies between different Aboriginal Australian and PNG populations 94

3.4.8 C4A and C4B linkage disequilibrium relationships in Aboriginal Australian and PNG populations 94

3.4.9 Determination of the genetic distance between Aboriginal Australian and PNG populations 97

3.5 Discussion 97

Chapter 4: Molecular-based C4 allotyping protocol 108

4.1 Introduction 108

4.2 Aim 110

4.3 Materials and Methods 110

4.3.1 C4 nomenclature 110

4.3.2 Samples for C4 PCR-RFLP allotyping 110

4.3.3 C4 protein gel electrophoresis and immunofixation 110

4.3.4 Genomic DNA extraction 111

4.3.5 Taq I restriction analysis of genomic DNA, and C4 specific hybridisation 111

4.3.6 Primer design for PCR amplification of polymorphic C4 gene regions 111

4.3.7 PCR amplification of nucleotide bases encoding polymorphic amino acid residues 112

4.3.8 PCR-RFLP analysis of nucleotide bases encoding C4 polymorphic amino acid residues 112

4.3.9 Cloning and sequencing of nucleotide bases encoding amino acid residues #286 to #330 in short C4 alleles 117

4.3.10 Allele specific oligonucleotide (ASO) hybridisation to nucleotides encoding polymorphic amino acid residue #1182 117

4.3.10.1 Samples for ASO typing 117

4.3.10.2 PCR amplification of nucleotides encoding amino acid residue #1182 and Southern blotting 119
4.3.10.3  End-labelling of allele-specific oligonucleotides (ASO), and hybridisation to membranes

4.3.11  Sequencing of C4A*3 and C4A*4 alleles to confirm ASO hybridisation to nucleotides encoding amino acid residue #1182

4.3.11.1  Donor samples

4.3.11.2  PCR amplification of the gene region surrounding nucleotide bases which encode amino acid residue #1182

4.3.11.3  Cloning of PCR products, and automated sequencing

4.3.12  Predicted Isoelectric Point (PI) of derived protein sequences for each C4 allele

4.4  Results

4.4.1  Design of the C4 PCR-RFLP molecular allotyping protocol

4.4.2  Amino acid residues encoded by polymorphic nucleotide bases as determined by PCR-RFLP analysis

4.4.2.1  Amino acid residue #328

4.4.2.2  Amino acid residue #306

4.4.2.3  Amino acid residue #458

4.4.2.4  Amino acid residue #707

4.4.2.5  Amino acid residue #1054

4.4.2.6  Amino acid residue #1101-#1106

4.4.2.7  Amino acid residue #1157

4.4.2.8  Amino acid residue #1182

4.4.2.9  Amino acid residue #1186

4.4.2.10  Amino acid residue #1188 and #1191

4.4.2.11  Amino acid residue #1267

4.4.3  C4 allele-specific polymorphism

4.4.3.1  C4A alleles

4.4.3.2  C4B alleles

4.4.3.3  Long and short C4 genes

4.5  Discussion
Chapter 5. Molecular basis of C4 null alleles in Aboriginal Australians

5.1 Introduction

5.2 Aim

5.3 Materials and Methods

5.3.1 C4 null allele donors for genetic analysis

5.3.2 C4 protein gel electrophoresis and immunofixation

5.3.3 Genomic DNA extraction

5.3.4 Taq I restriction analysis of genomic DNA, and C4-specific hybridisation

5.3.5 Gel electrophoresis, Southern blotting and hybridisation of PCR amplified products containing C4 isotypic nucleotide bases

5.3.6 C4A- and C4B-specific PCR, and PCR-RFLP analysis of C4 polymorphic regions

5.3.7 Screening by PCR-RFLP analysis, for a 2 bp insertion in exon 29 which creates a C4A Q0 allele

5.3.8 Total RNA extraction from Aboriginal donors H11, H14, and H19. Preparation of cDNA, and C4 null allele specific PCR analysis

5.3.9 Predicted Isoelectric Point (PI) of derived protein sequences for each C4 allele

5.4 Results

5.4.1 Detection of C4 null alleles by protein gel electrophoresis and immunofixation

5.4.2 Taq I RFLP analysis of the C4 gene region

5.4.3 C4A-, and C4B- specific oligonucleotide hybridisation analysis of Aboriginal donors' having C4 null alleles

5.4.4 C4A and C4B isotypic PCR amplification of C4 *Q0 alleles

5.4.5 PCR-RFLP analysis of nucleotide bases encoding the C4A and C4B isotypic amino acid residues in C4B null allele donors

5.4.6 PCR-RFLP analysis of nucleotide bases encoding C4 allelic polymorphic amino acid residues in C4B *Q0 alleles of Aboriginal donors
5.4.7 PCR-RFLP analysis of nucleotide bases encoding polymorphic amino acid residues of C4A null alleles of Aboriginal donors 188
5.4.8 Aboriginal C4B *Q0 alleles identified by PCR-RFLP analysis, and their predicted isoelectric points (PI) 189
5.4.9 Aboriginal C4A *Q0 alleles identified by PCR-RFLP analysis, and their predicted isoelectric points (PI) 191
5.4.10 PCR-RFLP analysis of Aboriginal donors having C4A Q0 alleles to screen for a 2 bp insertion into exon 29 192
5.4.11 Transcription analysis of C4 *Q0 alleles in Aboriginal donors H11, H14, and H19 194
5.4.12 Association of the C4A *CAN1 allele with the C4B Q0 allotype in ten related Groote Island Aboriginal donors, and 50 non-related SLE Caucasian patients by PCR-RFLP analysis 199
5.4.13 Frequency of the C4A *CAN1 allele in Aboriginal Australians from Cape York 201
5.5 Discussion 201

Chapter 6 General Discussion 208

References 218
List of Figures

Figure 1.1 Classical and alternative complement activation pathways 3
Figure 1.2 C4 gene region structure and polymorphic amino acid residues situated in each region of the pro-C4 protein molecule 19
Figure 1.3 Model for deletion and duplication within the C4 gene region 32

Figure 2.1 DNA molecular weight standards 56

Figure 3.1 C4 allotypes in East and West Cape York Aboriginal Australian populations 74
Figure 3.2 C4A allotypes in coastal, inland and highland regions of PNG 75
Figure 3.3 C4B allotypes in coastal, inland and highland regions of PNG 76
Figure 3.4 Genetic distance dendrogram based on C4A and C4B gene frequencies 98

Figure 4.1 Schematic representation of the molecular based C4 typing protocol based on PCR-RFLP 114
Figure 4.2 PCR-RFLP banding patterns of nucleotide bases encoding polymorphic amino acid residues #306 and #328 from short (16 kb) and long (23 kb)/short (16 kb) C4 PCR fragments 124
Figure 4.3 Sequencing results of short C4 genes from nucleotide bases encoding residues #286 to #330, which extends across the variable sized intron 9 130
Figure 4.4 PCR-RFLP banding patterns of nucleotide bases encoding C4 allele polymorphic amino acid residues #458, #707, #1054 and #1101-#1106 134
Figure 4.5 PCR-RFLP banding patterns of nucleotide bases encoding polymorphic amino acid residues #1157, #1186, #1188, #1191, and #1267 in C4 PCR fragments 139
Figure 4.6  $C4B$ 12 and $C4B$ 22 allotypes identified by protein gel electrophoresis 142

Figure 4.7  $C4$ allotypes identified by protein gel electrophoresis 143

Figure 4.8  $C4$ allotypes identified by protein gel electrophoresis 144

Figure 5.1  $C4$ allotypes of $C4$ Q0 Aboriginal donors determined by protein gel electrophoresis and immunofixation 174

Figure 5.2  $Taq I$ RFLP and $C4$-specific probe hybridisation analysis of the $C4$ gene region for $C4$ Q0 Aboriginal Australian donors 175

Figure 5.3  Hybridisation of $C4A$- and $C4B$-specific oligonucleotides to Aboriginal donors with $C4$ Q0 alleles 180

Figure 5.4  $C4B$ isotypic PCR amplification of Aboriginal donors with $C4B$ Q0 alleles 181

Figure 5.5  PCR-RFLP banding patterns of nucleotide bases encoding polymorphic amino acid residues #1054 and #1101-#1106 in $C4$ Q0 alleles 182

Figure 5.6  PCR-RFLP banding patterns of nucleotide bases encoding polymorphic amino acid residues #1157, #1186, #1188, #1191 and #1267, in $C4$ Q0 alleles 183

Figure 5.7  PCR-RFLP banding patterns from the analysis of nucleotide bases encoding polymorphic amino acid residues #306 and #328 in $C4$ Q0 allele donors 186

Figure 5.8  PCR-RFLP analysis of $C4A$ Q0 allele donors, in search of a 2 bp insertion in exon 29 193

Figure 5.9  $C4B$-specific PCR amplification of nucleotides encoding amino acid residues #1091 to #1191 from cDNA samples of Aboriginal donors H11, H14 and H19 195

Figure 5.10  PCR-RFLP analysis of nucleotide bases encoding residues #1188 and #1191, in cDNA samples from H11, H14 and H19 196

Figure 5.11  $C4A$-specific amplification of cDNA samples from $C4A$ Q0 allele donor H19 197
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Murine $C4$ investigations and findings</td>
<td>27</td>
</tr>
<tr>
<td>1.2</td>
<td>$C4$ studies of mammals</td>
<td>28</td>
</tr>
<tr>
<td>1.3</td>
<td>$C4$ gene features of humans and other primates</td>
<td>29</td>
</tr>
<tr>
<td>3.1</td>
<td>Observed and expected numbers of $C4A$ allotypes in East and West Cape York Aboriginal Australians</td>
<td>77</td>
</tr>
<tr>
<td>3.2</td>
<td>Observed and expected numbers of $C4A$ alleles in three PNG populations</td>
<td>78</td>
</tr>
<tr>
<td>3.3</td>
<td>Observed and expected numbers of $C4B$ alleles in East and West Cape York Aboriginal Australians</td>
<td>80</td>
</tr>
<tr>
<td>3.4</td>
<td>Observed and expected numbers of $C4B$ alleles in three PNG populations</td>
<td>81</td>
</tr>
<tr>
<td>3.5</td>
<td>$C4A$ gene frequencies of two Aboriginal and three PNG populations</td>
<td>82</td>
</tr>
<tr>
<td>3.6</td>
<td>$C4B$ gene frequencies of two Aboriginal and three PNG populations</td>
<td>83</td>
</tr>
<tr>
<td>3.7</td>
<td>Statistically significant differences of $C4$ gene frequencies between Aboriginal Australian and PNG populations</td>
<td>89</td>
</tr>
<tr>
<td>3.8</td>
<td>Distribution of $C4A$ and $C4B$ allotypes in Aboriginal Australian populations</td>
<td>91</td>
</tr>
<tr>
<td>3.9</td>
<td>Numbers of $C4A$ and $C4B$ allotypes in three PNG populations</td>
<td>92</td>
</tr>
<tr>
<td>3.10</td>
<td>Distribution of $C4A$ and $C4B$ allotypes expressed as percentages in three PNG populations</td>
<td>93</td>
</tr>
<tr>
<td>3.11</td>
<td>Statistically significant differences of $C4$ allotype frequencies between different Aboriginal Australian and PNG populations</td>
<td>95</td>
</tr>
<tr>
<td>3.12</td>
<td>$C4A$ and $C4B$ allele linkage disequilibrium relationships in Aboriginal Australian and PNG populations</td>
<td>96</td>
</tr>
<tr>
<td>4.1</td>
<td>$C4$ primers for PCR amplification of polymorphic amino acid residues</td>
<td>113</td>
</tr>
</tbody>
</table>
Table 4.2  *C4* polymorphic amino acid residues, their isotypic-specificity and *C4* gene size specificity  

Table 4.3  PCR conditions for each set of *C4* primers  

Table 4.4  Experimental details for PCR-RFLP analysis of *C4* nucleotide bases encoding polymorphic amino acid residues  

Table 4.5  β-chain, exon 9 polymorphic amino acid residues in Aboriginal Australians  

Table 4.6  β-chain, exon 9 polymorphic amino acid residues in Caucasians  

Table 4.7  Taq I RFLP analysis of the *C4* gene region for Aboriginal Australians, and Caucasians  

Table 4.8  PCR-RFLP results for *C4* polymorphic amino acids of Aboriginal Australians  

Table 4.9  PCR-RFLP results for *C4* polymorphic amino acid residues of Caucasians  

Table 4.10  Polymorphic amino acid residues for specific *C4A* alleles, and their predicted isoelectric points  

Table 4.11  Polymorphic amino acid residues for specific *C4B* alleles, and their predicted isoelectric points  

Table 5.1  *C4*- specific primers for PCR analysis of cDNA from *C4* *Q0* allele donors  

Table 5.2  PCR conditions for each pair of *C4*- specific primers in cDNA synthesis  

Table 5.3  *C4* allotypes according to protein electrophoresis, HLA *DR, DQB* and *DQA* genotypes and *C4 Taq I* RFLP results for *C4* null allele Aboriginal Australian donors  

Table 5.4  *C4* polymorphic amino acids of Aboriginal Australian donors as derived from analysis of nucleotide sequence by PCR-RFLP analysis  

Table 5.5  β-chain, exon 9 polymorphic amino acid residues in Aboriginal Australians  

Table 5.6  *C4A* *Q0* allele amino acid residue polymorphism, derived from nucleotide base PCR-RFLP analysis  

Table 5.7  PCR-RFLP results for nucleotides encoding *C4* polymorphic amino acids of Groote Island Aborigines
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASO</td>
<td>allele-specific oligonucleotide</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine 5'-triphosphate</td>
</tr>
<tr>
<td>bp</td>
<td>base pairs</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
</tr>
<tr>
<td>cps</td>
<td>counts per second</td>
</tr>
<tr>
<td>dATP</td>
<td>2'-deoxyadenosine 5'-triphosphate</td>
</tr>
<tr>
<td>dCTP</td>
<td>2'-deoxycytosine 5'-triphosphate</td>
</tr>
<tr>
<td>dd</td>
<td>double distilled</td>
</tr>
<tr>
<td>ddNTP</td>
<td>dideoxynucleotide triphosphate</td>
</tr>
<tr>
<td>DEP</td>
<td>diethyl pyrocarbonate</td>
</tr>
<tr>
<td>dGTP</td>
<td>2'-deoxyguanosine 5'-triphosphate</td>
</tr>
<tr>
<td>DIGE</td>
<td>direct gel electrophoresis</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethylsulfoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxyribonucleotide triphosphate</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>dTTP</td>
<td>2'-deoxythymidine 5'-triphosphate</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediamine tetra-acetic acid</td>
</tr>
<tr>
<td>HLA</td>
<td>human leukocyte antigen</td>
</tr>
<tr>
<td>IDDM</td>
<td>insulin dependent diabetes mellitus</td>
</tr>
<tr>
<td>IgG</td>
<td>immunoglobulin G</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropylthio-β-galactosidase</td>
</tr>
<tr>
<td>kb</td>
<td>kilobase pairs</td>
</tr>
<tr>
<td>kDa</td>
<td>kilodaltons</td>
</tr>
<tr>
<td>LDL</td>
<td>low density lipoprotein</td>
</tr>
<tr>
<td>MAC</td>
<td>membrane attack complex</td>
</tr>
<tr>
<td>MHC</td>
<td>major histocompatibility complex</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
</tr>
<tr>
<td>PFGE</td>
<td>pulse-field gel electrophoresis</td>
</tr>
<tr>
<td>RFLP</td>
<td>restriction fragment length polymorphism</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulphate</td>
</tr>
<tr>
<td>SLE</td>
<td>systemic lupus erythematosis</td>
</tr>
<tr>
<td>SSC</td>
<td>tri sodium citrate</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Name</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------</td>
</tr>
<tr>
<td>SSCP</td>
<td>single-stranded conformation polymorphism</td>
</tr>
<tr>
<td>TAE</td>
<td>tris-acetate EDTA</td>
</tr>
<tr>
<td>TBE</td>
<td>tris-borohydrate EDTA</td>
</tr>
<tr>
<td>TE</td>
<td>tris-EDTA</td>
</tr>
<tr>
<td>Tris</td>
<td>2-amino-2-(hydroxymethyl)-1,3-propanediol</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>X-gal</td>
<td>5-bromo-4-chloro-3-indoyl-β-D-galactosidase</td>
</tr>
</tbody>
</table>
Chapter 1
General Introduction

1.1 Introduction

Complement component C4 protein, as part of the complement pathway, has an important role in immune response. Deficiencies in C4 protein are implicated in the pathogenesis of some autoimmune diseases (Dawkins et al., 1983; Atkinson, 1988; Robb et al., 1988). Knowledge of the molecular basis of C4 is necessary to advance studies in the role of C4 complement, diagnosis and treatment of associated diseases, and in population genetics. Although much is known of the molecular basis of C4, including its genomic structure, knowledge of the molecular basis of individual C4A and C4B null alleles is incomplete, especially for non-Caucasoid populations. In this context, the high frequency of C4 deficiency alleles reported in Aboriginal Australian populations (Ranford et al., 1987) is of particular interest, due to a possible correlation with high prevalence and marked severity of the autoimmune disease, systemic lupus erythematosus (SLE) in these populations (Anstey et al., 1993).

C4 allotyping is a notoriously unreliable technique, available to only a limited number of specialised laboratories. An accurate definition of the various alleles associated with the C4 gene region is essential for reproducible results, and so a more reliable and accurate C4 allotyping method was a major part of studies for this thesis. Studies here are concerned principally with defining the molecular basis of C4 null alleles and other alleles in Aboriginal Australians, and the development of a molecular-based C4 allotyping protocol.

This chapter reviews the role of complement in human immune response, the association of C4 protein deficiencies with autoimmune diseases, the molecular structure and organisation of the C4 gene region, and the molecular basis of C4 null alleles in Caucasians. The main aims of each of three major areas of research undertaken are summarised at the conclusion of this review.
1.2 The role of complement in the human immune response

Complement consists of a complex series of 30 plasma and membrane proteins which are now grouped into families according to their degree of homology and functional activity (Kinoshita, 1991). These families consist of proteins of the activation pathways, terminal pathway, regulatory proteins, and complement receptors. A large number of biological activities important in the inflammatory response and in host resistance to infection are programmed at various points in the complement cascade. During an immune response, the ability of antibody to inactivate foreign antigen depends upon the co-operation of complement. The complement cascade is responsible for solubilisation of immune complexes, viral neutralisation, stimulation of the inflammatory response (a localised protective defense), opsonisation (promotes recognition by phagocyte complement receptors), and lysis of foreign cells from tissues and blood. These processes are achieved through the sequential activation of complement proteins and by complement protein interaction with antibodies and/or cells that express complement receptors. Clearance of immune complexes prevents their interaction with the endothelium and reduces local inflammation and tissue damage from the activation of complement.

The complement system targets microorganisms and other antigens for binding to complement receptor-bearing cells, attracts phagocytic cells to the area of complement activation, and destroys target membrane. These processes are initiated by the formation of the $C3$ convertase enzyme which activates $C3$ to $C3b$. $C3$ complement protein products are then responsible for activation of other complement proteins which lead to formation of the membrane attack complex (MAC), and finally cell lysis. There are two pathways to complement activation- either the 'classical pathway' which is triggered by antibody/antigen complexes, or the second, known as the 'alternative pathway', which has its own intrinsic ability to recognise foreign organisms. A summary of these pathways is shown in Figure 1.1.
Figure 1.1 Classical and alternative complement activation pathways. Ag: antigen, Ab: antibody, MAC: membrane attack complex, B: factor B, D: factor D, the remaining proteins are other complement pathway proteins
1.2.1 Classical and alternative complement pathways

The classical pathway involves $C1, C4, C2$ and $C3$ proteins, and is activated by antigen/antibody complexes containing IgM and IgG (aggregated immunoglobulins). $C1$ consists of three sub-components, $Clq,Clr$ and $Clr$. Activation of this pathway is specific, and requires the interaction of $Clq$ with the Fc portion of antibodies (aggregated IgG or IgM), which are bound to antigen. This interaction leads to conversion by limited sequential proteolysis of $Clr$ and $Clr$ to their activated enzymatic forms, $Clr^-$ and $Clr^+$. $Clq$ can also interact directly with surface components of certain bacteria and viruses (Tomlinson, 1993). $Clr^-$ activates complement component $C4$, cleaving 10 KDa ($C4a$ released) from the amino terminal of the alpha chain of $C4$ and leaving $C4b$ which binds covalently to the Ag:Ab:$C1$ complex. In the presence of magnesium ions, $C2$ binds to $C4b$ and is cleaved by $Clr$ to $C2b$ (released) and $C2a^-$ which has a proteolytic catalytic site (Figure 1.1). The resulting product is 'C4b2a', which is a $C3$ convertase enzyme.

The alternative complement pathway operates by continuously generating small amounts of activation products derived from $C3$. This pathway can function in the absence of antibodies, and is spontaneous. There are two possible mechanisms which could activate the alternative pathway. Firstly, initiation may be due to hydrolysis of the intramolecular thioester bond of $C3$, causing a conformational change in $C3$ resulting in a $C3b$-like structure which is often called $C3(H_2O)$ (Pangburn and Müller-Eberhard, 1984). The other mechanism involves the thioester bond of $C3$ which is directly attached by a surface-containing hydroxyl group resulting in covalent binding. $C3b$ bound to host cells is deactivated by host regulatory mechanisms, whereas $C3b$ bound to foreign substances such as bacteria, (where there is usually no regulatory mechanism), results in the formation of a surface-bound $C3$ convertase. Once $C3b$ is established, whether it be surface bound or in fluid phase, the $C3$ convertase enzyme is formed by an interaction with Factor B and Factor D. Factor B is activated by antigens particularly those carrying complex polysaccharides on their surface, such as bacteria, as well as large insoluble immune
complexes and IgA-carrying immune complexes. Factor D cleaves Factor B forming 'Ba' which is released, and 'Bb' which binds to $C3b$, creating the alternative-pathway $C3$ convertase 'C3bBb' (Figure 1.1). A catalytic site on Bb in the fluid phase of $C3(H_2O)Bb$ convertase, splits multiple molecules of $C3$ to $C3b$, some of which bind covalently to amino or hydroxyl groups on surrounding surfaces. An amplification loop is generated, as those $C3$ convertase complexes which are surface bound in turn produce more $C3b$ (Pangburn and Müller-Eberhard, 1984). As a result, once activation occurs, the generation of $C3b$ is accelerated. The invading organism becomes coated with $C3b$ which acts as a tag for complement receptors on phagocytic cells, and terminal components of the complement system (Ross and Medof, 1985).

$C3$ convertases of the classical and alternative pathways are homologous, as are $C4$ and $C3$, $C4b$ and $C3b$, $C2$ and B, and $C2a$ and Bb (Campbell et al., 1988). Phylogenetic evidence, (reviewed by Farries and Atkinson, 1991), suggests that complement probably arose as an independent immune mechanism like the 'alternative pathway' which can act independently, and provides a natural immunity against invading microorganisms. The 'classical pathway' most probably evolved along with the antibody response, enabling the host to cope with rapidly evolving pathogens, and respond to poor stimulators, to memorise, and to prevent recurrent infections.

1.2.2 Terminal pathways

Once the $C3$ convertase enzyme has been formed it cleaves $C3$, creating $C3a$ and $C3b$ through the proteolytic site on 'C2a' or 'Bb'. $C3a$, an anaphylotoxin, is released into the circulation, and $C3b$ binds to the $C4b2a$ or $C3bBb$ complex forming the $C5$ convertase $C4b2a3b$ or $C3bBb3b$. These latter enzymes in turn, cleave complement component $C5$ to $C5a$ and $C5b$. $C5b$ induces the association of the later-acting components $C6$, $C7$, $C8$, and $C9$ with itself, producing a large protein complex known as the membrane-attack complex 'C5b-9b'. This complex contains one molecule of $C5b$, $C6$, $C7$ and $C8$, and multiple molecules of $C9$ (up to 18). $C9$
has been shown to greatly enhance the activity of the membrane attack complex (Müller-Eberhard, 1986). The membrane-attack complex is extremely hydrophobic, and inserts into the lipid bilayer of bacterial and animal cells causing lysis. The precise model for the membrane-attack complex ('MAC' pore), is unknown, although various models have been proposed, (for review refer to Esser, 1991; Bhakdi and Tranum-Jensen, 1991). The terminal pathway in host defence is important, as individuals deficient in one of the components C5 to C9 suffer from infection by Gram-negative bacteria of the genus *Neisseria* (Ross and Densen, 1984), reflecting the breakdown in the immune response at this stage.

1.2.3 Inflammatory response

Activation of the complement cascade at the site of injury results in the production of pro-inflammatory mediators \( C5a, C4a, \) and \( C3a. \) These proteins cause a spasmogenic response in smooth muscle, vasoactive amine release, and an increase in vascular permeability. \( C5a \) is also a potent chemoattractant, and effects cellular responses from phagocytes, such as, stimulation of cytokine synthesis, oxidative metabolism, and release of eicosanoids and degradative enzymes (Morgan, 1990; Frank and Fries, 1991).

In addition to these complement effector systems described above, the host has developed a means of controlling the complement cascade and protecting its cells against complement attack. These include complement-regulatory proteins and complement-cellular receptors.

1.2.4 Regulatory proteins

Host cell membranes contain proteins with down-regulatory properties such as the membrane cofactor protein (MCP), decay-accelerating factor (DAF), and complement receptor CR1. These act on \( C3b \) bound to host cells by promoting the cleavage of their components by serine protease I and its cofactors. The \( C3b \) inactivator is a potent inhibitor of the alternative pathway. It acts enzymatically to destroy \( C3b \) which is surface-bound and in fluid
phase. Properdin stabilises the \( C3 \) convertase formed by the alternative pathway, and Factors I and H regulate the amplification of \( C3 \) convertase by this pathway. In the classical pathway, \( C1 \) is controlled by \( C1 \) esterase inhibitor, and \( C4 \) is inactivated by \( C4 \) binding protein and Factor 1 which cleaves \( C4b \) to \( C4c \) and \( C4d \). For future protection, the host-cell membranes contain inhibitors of the terminal pathways which regulate formation of the membrane-attack complex (MAC); they are, homologous restriction factor (HRF)/C8-binding protein (C8bp), and CD59. Some regulatory proteins have a number of functions. Factor H and CR1 for example, apart from their functions mentioned above, are involved in host cell protection (Kinoshita, 1991). Membrane-bound complement inhibitors are most important for protecting host cells. This is evidenced by individuals containing blood cells which have an increased susceptibility to complement (Kinoshita, 1991), and suffering from the disorder paroxysmal nocturnal haemoglobinuria (PNH) (Rosse, 1989; Rotoli and Luzzatto, 1989).

1.2.5 Complement receptors

Phagocytosis of most encapsulated microorganisms requires antibody and/or complement, which jointly promote binding and ingestion. Complement-dependent phagocytosis, mediated by the alternative pathway, depends on the interaction between deposited complement components and phagocyte receptors, such as \( CR1, CR3 \) and \( C5a \) (Ahearn and Fearon, 1989; reviewed by Krych et al., 1992). \( C5a \) induces margination and chemotaxis of neutrophils (Hugli, 1984), and is responsible for attracting other phagocytic cells to the area of complement activity. It renders these cells fully active in phagocytosis by up-regulating surface expression of \( CR1 \) and \( CR3 \) (Fearon and Collins., 1983; Arnaout et al., 1984). \( C3b \) and \( C4b \) immune complexes can bind to a receptor, \( CR1 \), which is found in large numbers on erythrocytes. The erythrocyte/immune complex is transported to the liver and spleen where the immune complex is transferred to a mononuclear phagocyte, and the erythrocyte returns to circulation (Schifferli et al., 1980). Clearance of immune complexes prevents their interaction with the endothelium and reduces local inflammation and tissue damage from the activation of
complement. $Clq$ has also been shown to enhance phagocytosis of both IgG and complement-$C3b$ coated organisms (Bobak et al., 1987 and 1988).

1.2.6 Additional roles of complement

It is well established that complement's principal role is host defence by eliminating foreign material. Current research however, has suggested complement also plays a role in the induction and modulation of the adaptive immune response which leads to antibody production. Two factors have suggested this additional role of complement. Firstly, animals depleted of $C3$ by treatment with cobra venom factor (CVF), as well as humans and animals genetically deficient in early complement components, have an impaired antibody response (Bottger and Bitter-Suermann, 1987; Heyman, 1990b). Secondly, CR2 transduces a growth signal in B cells (Cooper et al., 1988; Ahearn and Fearon., 1989; Matsumoto et al., 1991), and a monoclonal antibody that recognises the ligand-binding site of mouse CR2, inhibits the primary antibody response \textit{in vivo} (Heyman et al., 1990a; Kinoshita, 1991). Additional evidence which also support the theory of complement's involvement with adaptive immunity include the following: (1) Antigen-presenting cells and B cells which have complement receptors 1, 2 and 3; (2) Binding of immune complexes to follicular dendritic cells in the germinal centers of lymph nodes (important for generation of memory B cells), and also; (3) Depletion of $C3$ in experimental animals which impairs antigen localisation to germinal centres (reviewed by Morgan and Walport, 1991).

1.3 $C4$ deficiency and systemic lupus erythematosus (SLE)

SLE is an autoimmune connective tissue disease, the aetiology of which is unknown. This disease is complex having many clinical manifestations ranging from a mild cutaneous disorder to a more severe, life-threatening disorder which can affect multiple organs. The disease can also have both active stages and periods of remission. The extreme diversity of symptoms seen in SLE, many of which are non-specific, make the diagnosis of this disease very
difficult, especially when other connective tissue diseases have the same or similar features (Garfield, 1989). Due to the complexity of this disease an attempt has been made to develop criteria for the diagnosis of SLE to ensure some uniformity with regard to its cause, and to exclude other autoimmune connective tissue disorders (Tan et al., 1982). Since 1951 when activation of complement in SLE was first recognised, (Vaughan et al., 1951), there has been a major interest in the field of complement research and its association with this disease.

Deficiency of classical pathway proteins, especially Clq, C4 and C2, are associated with an increased susceptibility to SLE (Morgan and Walport, 1991a). The mechanism underlying the association of SLE with complement deficiency has not been established, but is thought to be related to the physiological inactivity of these proteins in removing immune complexes from circulation and tissues (Schifferli et al., 1986). Accumulation of immune complexes in tissues causes inflammation and the release of autoantigens, which in turn stimulate the production of autoantibodies, and therefore more immune complexes (Lachmann and Walport, 1987).

The strong association between complement deficiency and SLE suggests partial complement deficiency may also cause increased disease susceptibility. Research by Atkinson (1988), supported this hypothesis, finding homozygous C4A deficiency in 10-15%, and heterozygous C4A deficiency in 50-80% of Caucasian patients with SLE. Other studies found strong linkage disequilibrium between C4A null and HLA-DR 3 alleles in association with SLE (Fielder et al., 1983), and so confusing the relative contribution of each disease susceptibility gene. Some studies have found that in Caucasians with SLE, the B8 and DR 3 alleles associated with C4A null alleles represent the most important HLA-linked genetic risk factor for SLE (Howard et al., 1986; Dunckley et al., 1987). These act in a dose response manner, and synergistically increase disease risk when found in combination with another independent risk factor, DR 2 which is associated with C2 deficiency.
An attempt was made to resolve this problem by analysing the HLA association with SLE amongst different racial groups (reviewed by Ng and Walport, 1988). This analysis was used to determine which SLE candidate genes were maintained amongst different HLA haplotypes in each racial group. Unlike HLA-DR and DQ associations in SLE, the C4A null allele association with SLE was maintained across racial barriers. One group of researchers found a significant association between C4A null alleles and SLE in black and white Americans (Reveille et al., 1985). This association was stronger and often independent of HLA-B8 and/or DR 3's association with SLE. Others found that complete or partial C4A deficiency is a genetic determinant of SLE common to three ethnically distinct populations, namely Caucasoid, Chinese and Japanese (Dunckley et al., 1987). Reveille et al. (1985), detected a C4 null allele dosage effect in relation to disease predisposition. They found an increased frequency of C4 null alleles in individuals with SLE (60%), compared with healthy relatives (50%), and spouses (24%). As a result of these studies it can be said that C4 null alleles show a strong association with SLE which is common in a number of ethnic groups.

Most studies have indicated that C4A rather than C4B null alleles are associated with SLE. Fielder et al. (1983), found 15% of SLE patients, compared to none of the control patients were homozygous for C4A null alleles. Other studies of SLE groups from Australia, China, Japan and the United States found 10-15% of white SLE patients were completely deficient in C4A compared with less than 2% of control subjects (Howard et al., 1986). It is generally believed that the stronger association of C4A rather than C4B null alleles with SLE may be a reflection of C4A proteins ability to preferentially form amide linkages with amino groups such as those formed between C4 and immune complexes. Therefore, C4A protein deficiency is thought to lead to accumulation immune complexes within the body which is a feature of SLE. C4B protein on the other hand, more efficiently forms ester linkages with hydroxyl groups such as those found on the surface of encapsulated microorganisms. It has been shown that C4A protein is about twice as efficient as C4B protein in functionally restoring C4 deficient human serum ability to inhibit precipitation of immune complexes (Carroll et al.,
1989). As a result of the functional preferences of C4A and C4B proteins, C4A rather than C4B null alleles are more likely to be the strong susceptibility factor in SLE.

As a result of the association between complement deficiency and SLE, measuring complement activation is important for detecting and monitoring this disease state. The degree of complement activation is thought to reflect the intensity of the inflammatory response (Porcel et al., 1993), which in turn indicates the stage of disease activity. Assessing complement activation has mainly been carried out on patients with 'SLE' (reviewed by Porcel et al., 1993). This type of testing relies on measuring intact C3 and C4 protein levels in serum, or by determination of complement haemolytic function assays. These techniques measure complement activation indirectly, and are relatively insensitive (reviewed by Porcel et al., 1993).

New approaches for measuring complement activation are currently being developed. These depend on the detection of complement activation fragments, neo-antigens or complexes which directly result from complement activation. These methods are now being implemented in the clinical immunology laboratory for routine testing (Porcel et al., 1993).

1.4 C4 deficiency and other disease states

The majority of C4 deficiency biological research has concentrated on the association with the autoimmune disease SLE as described above (Dawkins et al., 1983; Fielder et al., 1983; Howard et al., 1984; Kemp et al., 1987). However, investigations of candidate genes for other diseases have found C4 deficiency to be associated with Myasthenia Gravis (Dawkins et al., 1983), Insulin Dependent Diabetes Mellitus (Dawkins et al., 1983; Hauptmann, 1980; Deschamps et al., 1988; Jenhani et al., 1992; Segurado et al., 1991), juvenile dermatomyositis (Robb et al., 1988), as well as chronic hepatitis and sub-acute sclerosis panencephalitis (Atkinson, 1988). The transient selective C4 deficiency in infancy results in erythema multiforme with outbreaks of bacterial and viral infection
C4B null alleles have been found in association with Brazilian Paracoccidioidomycosis (a thermomorph Dak et al., 1992). C4B deficiency is considered a genetic marker for IgA Nephropathy (Welch et al., 1987 and 1989). In addition, C4A and C4B protein deficiency have been found to be associated with rapid onset of AIDS after infection by HIV-1 in Caucasians (Hentges et al., 1992).

The HLA class III gene region has complement genes C2, C4, and factor B which are collectively referred to as complotypes. Specific complotypes are associated with various disease states (Dawkins et al., 1983). These complotypes include both functional and dysfunctional C4 alleles. Examples of complotypes as part of HLA-extended haplotypes include HLA-A 1 , Cw 7, B 8, Bf S, C2 C, C4A Q0, C4B 1, DR 3, associated with Myasthenia Gravis, SLE, and IDDM; HLA- B 18, Bf F1, C2 C, C4A 3, C4B Q0, DR 3, associated with IDDM; and HLA- B 14, C4A 2, C4B 2, DR 1, associated with late onset of 21-hydroxylase deficiency (Dawkins et al., 1983).

Disease association studies have confirmed the importance of the function of the C4 protein in the immune response. From an aetiological perspective, the molecular basis of C4 deficiency plays an important role in understanding association of the C4 null allele with different diseases. The majority of C4 disease association studies have been based on serological typing of C4 allotypes and thus do not indicate the molecular basis of this protein deficiency.

1.5 C4 protein and origin of its synthesis

In humans, the liver is the major producer of C4 messenger. It has been shown that after liver transplantation, patients take on expression of the donor's C4 and factor B allotypes (Koskimies et al., 1991). C4 genes have a transcript for a protein of about 200,000 Mr, consisting of three peptide chains, namely alpha (α) (95,000 Mr), beta (β) (75,000 Mr), and gamma (γ) (30,000 Mr), (Carroll, 1983). C4 protein is synthesised in the liver as a single peptide chain (pro-C4), which undergoes proteolytic cleavage on secretion into plasma to produce β, α, and γ chains in this order (Goldberger et al., 1980).
C4 transcripts have also been found in the kidney, thyroid, brain, and breast carcinoma (Feucht et al., 1989; Witte et al., 1990). Andoh et al. (1993), found C4 synthesis in the intestinal epithelial cell line 'Caco-2'. In murine studies, C4 is expressed in a variety of tissues including mammary gland, lung, spleen, and kidney (Saunders and Edidin, 1974; Cox and Robins, 1988; Passwell et al., 1988). In vitro studies have shown that macrophages are capable of synthesising C4 (Whaley, 1980; Alpert et al., 1983).

In the last few years a number of studies have concentrated on determining which cells are responsible for C4 protein synthesis in human kidneys, and if there is an association between C4 protein synthesis and immune-complex-mediated lesions in this organ (Witte et al., 1991; Zhou et al., 1993). Witte et al. (1991), found C4 is expressed at high levels in both normal and diseased kidneys, and in situ hybridisation analysis found the renal tubular epithelial cells are the source of these transcripts. Also, this study indicated that C4 expression is unlikely to contribute to glomerulonephritis, and in fact may serve as a local protective function for the epithelial lining. In addition, this group identified C4 expression in hepatocytes, thyroid follicular epithelial cells, and ductal epithelial cells of the submandibular salivary gland, indicating that some components of complement may have a physiological role in epithelial cell function. Zhou et al. (1993), found C4 transcripts in human glomerular epithelial cells. These cells form part of a charge and size-selective filtration barrier between the blood and urinary spaces (Adler et al., 1983; Couser, 1990). The function of this cell barrier was found to be disrupted in various types of complement-mediated injury. For example, it was shown to be the primary target of complement-mediated attack in glomerulonephritis (Couser and Abrass, 1988). The above studies have narrowed down the region in which cells transcribe C4 protein. However, there is some controversy over the functional and pathophysiological significance of C4 protein synthesis in the kidney and therefore requires further study.
1.6 Detection of \( C4 \) protein

Protein gel electrophoresis followed by immunofixation is the most common technique used for the identification of \( C4 \) protein allotypes. \( C4 \) protein allotyping identifies the phenotype of a \( C4 \) allele, and is denoted as for example \( C4A \, 3 \), whereas \( C4 \) alleles identified by analysis of DNA represents the allotype determined by DNA sequencing of a \( C4 \) allele and is therefore denoted as \( C4A \, \ast 3 \) in this thesis. Other methods for characterising \( C4 \) allotypes have been summarized by Mauff et al. (1990a-d), Rittner and Stradmann-Bellinghausen, (1990), and Kaufmann et al. (1990). These papers reported on a collaborative study of 136 individual's serum samples to define common and rare \( C4 \) alleles to be used as standard reference variants. Also, duplicated, deleted and non-deleted \( C4 \) null alleles, and hybrid \( C4 \) alleles were investigated. A series of different typing criteria including gel electrophoresis and immunofixation, haemolytic activity, reactivity with monoclonal antibodies and Rodgers/Chido reagents, \( \alpha \)-, and \( \beta \)-chain types, and \( C4/CYP21 \, Taq \, I \) RFLP's, were collectively used for the above studies. Each of these techniques is limited by the amount of information it can provide with respect to an individual's \( C4 \) allotype. Collectively, however, these techniques are quite informative but are unsuitable for routine diagnostic typing as each technique requires specific skills and are time consuming.

\( C4 \) \( \alpha \)-chain typing and immunoblotting with \( C4A \)- and \( C4B \)-specific monoclonal antibodies is used for classification of a number of expressed gene products and haplotypes aberrant with respect to their Rodgers and Chido determinants (Rittner and Stradmann-Bellinghausen, 1990). However, this technique is not able to detect every type of \( C4 \) allele including heterozygous null alleles at each \( C4 \) gene locus. More recently an enzyme-linked immunoassay (EIA), has been developed for \( C4 \) allotyping (Moulds et al., 1990). This assay uses heat-aggregated human IgG (Fc portion of antibody) to activate \( Cl \) and capture \( C4 \). \( C4 \) bound to Ab:Cl complex molecules are detected with horseradish peroxidase-conjugated monoclonal antibodies, and are specific for \( C4d \) (\( C4A/C4B \) cross-reactive), \( C4A \,(Rg1) \), and \( C4B \,(Ch1) \). This technique is useful in that it provides
an indication of the level of $C4$ protein in an individual. However, the normal level of $C4$ in an individual is not uniform and difficult to determine, and as stated above, monoclonal antibodies for Chido and Rodgers antigenic determinants do not detect all $C4$ alleles including $C4$ null alleles. A $C4$-specific haemolytic activity assay was designed to estimate the function of $C4$ in an \textit{in vitro} system, thereby reflecting as close as possible the \textit{in vivo} situation (Kaufmann \textit{et al.}, 1990). This assay was based on $C4$ dependent lysis of sensitised sheep erythrocytes, which involves activation of the whole complement cascade and thereby reflecting most of the \textit{in vivo} interactions of $C4$. However, estimation of the number of $C4$ null alleles depends on different activity levels by comparisons between the $C4A$ and $C4B$ isotypes. This does not provide a clear indication of the presence of $C4$ null alleles as some $C4$ null allotypes have been shown to result from duplicated $C4$ genes on the one haplotype. They therefore have functional $C4$ alleles at both $C4$ gene loci. Also, this assay does not indicate which $C4$ alleles are present. A Rodgers and Chido $C4$ antigenic determinant typing system has been developed which is based on haemagglutination inhibition (Giles, 1990). Although this technique is probably the best technique for accurate detection of rare $C4$ allotypes, the success of this typing technique depends on the correct combinations of antigenic determinants in haplotypes and in genotypes, and therefore relies heavily on family studies. Also, in the presence of other $C4$ allotypes it is not always possible to detect reversed antigenicity. An alternative to haemolytic overlays is immunoblotting with monoclonal antibodies (Doxiadis and Grosse-Wilde, 1990). This technique is very useful as it identifies $C4B$ and $C4A$ isotypes which tend to migrate by gel electrophoresis into the opposite anode and cathode ends, respectively. However, this technique does not enable double $C4$ null allele heterozygotes to be identified.

$C4$ protein gel electrophoresis followed by immunofixation has remained the most informative $C4$ typing protocol and is the method of choice by standard reference $C4$ typing laboratories around the world. Serological analysis by electrophoresis and immunofixation involves the resolution of $C4$ alleles according to their unique size and charge. Treatment of the resulting protein gel with antisera
specific for $C4A$ and $C4B$ gene products results in visualisation of three bands for each $C4$ gene (shown in Chapter 3), which represent pro-$C4(\beta-\alpha+\gamma)$, $C4S$ (secreted form of $C4$), and $C4P$ (predominant plasma form of $C4$) (Chan et al., 1984). Exactly which band is represented by which $C4$ form is unknown. Another team of researchers found that $C4$ digestion with pancreatic carboxypeptidase $B$ converts the two cathode located bands for each $C4$ protein into the most anodally located band which is intensified after digestion, and thus enables the allotypes to be more easily interpreted (Sim and Cross, 1986). As the $C4$ alleles are highly polymorphic, another modification of $C4$ protein gel electrophoresis has been developed to standardise the relative migration distances for each allele by measuring each $C4$ protein migration using laser densitometry (Mauff et al., 1990a).

Gel electrophoresis and immunofixation of $C4$ proteins does not reveal the number of $C4$ genes present on a particular haplotype, nor can it easily identify overlapping or closely migrating alleles. A molecular-based technique which could utilise genomic sequence features which are unique for each $C4$ allele, involving the use of only one protocol, and which could be easily implemented in a standard laboratory, would permit $C4$ molecular allotyping for routine diagnostic studies.

1.7 $C4$ protein as a polymorphic marker for population studies

As with the HLA class I and II alleles, a number of population and evolutionary studies have been performed by analysing the HLA class III complement component $C4A$ and $C4B$ polymorphic alleles. Serological typing of the $C4$ alleles has been performed on a number of populations including, Tunisians (Ayed and Gorgi, 1990), Caucasians, Arabs, African Blacks, German Caucasians (Schendel et al., 1984), Finnish (Partanen et al., 1986b; 1987), Aboriginal Australians from Darwin and Alice Springs (Ranford et al., 1987), Melanesians, Micronesians, Japanese, Chinese, Koreans, Thai, Indians and white Australians (Ranford et al., 1986; Ranford, 1989).
Only five of the $C4A$ alleles, 'A 2', 'A 3', 'A 4', 'A 6', and 'A Q0', and three $C4B$ alleles, 'B 1', 'B 2', and 'B Q0', are commonly found with frequencies above 10% in any population, and distribution of these alleles appears to define various racial groups (Ranford, 1989). Mutation is more common at the $C4B$ locus creating more allelic variants than at the $C4A$ locus, and in some cases these $C4B$ alleles may be used as markers for specific populations. For example, the 'B 5' allele was detected in all Asian groups and may be regarded as a marker for this population (Ranford, 1989).

Analysis of the molecular basis of alleles in isolated populations can indicate genetic mechanisms responsible for allelic diversity. Recent studies by Gao et al. (1992b) of Aboriginal Australian populations using molecular genetic typing techniques of the polymorphic HLA class II gene region, have identified class II allotypes which are commonly detected in Aboriginal Australian populations, and have provided some insight into how these alleles have evolved.

1.8 Complement $C4$ genetics

To understand the basis of complement $C4$ protein association with autoimmune disease, a sound knowledge of this protein's genetic structure and regulation is necessary.

1.8.1 Gene localisation

Complement component $C4$ is situated in the HLA region on the short arm of chromosome 6. The HLA region has three classes of genes. Class I and class II gene products are highly polymorphic cell-surface glycoproteins involved in the discrimination of self from non-self by the immune system (Steinmetz and Hood, 1983; Kaufman et al., 1984). The class III gene region is about 1,000 kb and contains many duplicated genes, all of which have different functions. Genes in this region which have been localised and characterised include the tumor necrosis factors 'α' and 'β' (Spies et al., 1986; Carroll et al., 1987a; Dunham et al., 1987), heat shock proteins HSP70 (Sargent et al., 1989), complement proteins including factor B, $C2$, $C4A$ and $C4B$ (Carroll et al., 1984a), and the adrenal steroid 21-
hydroxylase genes CYP21A and CYP21B (Higashi et al., 1986; White et al., 1986). These genes lie along chromosome 6 in the order C2, Bf, C4A, CYP21A, C4B, and CYP21B. C2 and Bf genes are separated by less than 2 Kb, and are approximately 30 Kb away from the C4 genes (Carroll et al., 1984a and 1984b). CYP21B is important for the synthesis of steroids in the adrenal gland, whereas CYP21A (highly homologous to CYP21B), is a dysfunctional pseudogene.

1.8.2 Structure of the C4 gene locus and surrounding gene region

There are two C4 genes, namely C4A and C4B, which are approximately 12 kb apart (Carroll et al., 1985a; Yu et al., 1990), and are transcribed in the same 5' to 3' direction (Figure 1.2). C4A is usually at locus I, and C4B at locus II in the C4 gene region. However, there have been exceptions reported (Awdeh and Alper, 1980; Carroll et al., 1984b; Yu and Campbell, 1987). Each C4 gene is accompanied by a 21-hydroxylase gene situated approximately 3 kb downstream from the 3' end of each C4 gene (Carroll et al., 1985a; White et al., 1985). The orientation and position of these genes was determined by pulsed field gel electrophoresis (PFGE) (Dunham et al., 1987). The C4B gene lies approximately 350 kb from the DRA locus, and is centromeric to the C4A gene, which lies approximately 600 kb from the HLA-B gene locus.

Recent chromosome and gene mapping studies have explored the gene region immediately surrounding the C4 and CYP21 gene loci for other genes which may be associated with autoimmune disease. A pair of genes with tenascin-like sequences,XA and XB, were reported overlapping the CYP21A and CYP21B genes, respectively (Morel et al., 1989; Gitelman et al., 1992). When the promoter regions of the C4A and C4B genes were identified, a polyadenylation (poly(A)) signal, 'AATAAA', was found 631 bp upstream of the transcriptional initiation sites of each C4 gene (Alper, 1991). Each of these poly(A) sites is followed by a stretch of GT-rich sequence which is characteristic of the 3' end of mammalian genes (Proudfoot, 1991).
Figure 1.2 C4 gene region structure and polymorphic amino acid residues situated in each region of the pro-C4 protein molecule
Shen et al. (1994), has recently identified two novel genes RP1 and RP2, located immediately upstream (611 bp), of the C4A and C4B genes, respectively. Molecular analysis suggests that the RP gene may encode a nuclear protein (Shen et al., 1994). This gene was found to undergo duplication and deletion together with neighbouring genes X, C4, and CYP21, which has triggered a special interest in the investigation of RP structure, genetics and function. Cumulative data suggest that the four tandemly arranged genes RP, C4, CYP21, and gene X, together form a modular structure, namely 'RCCX' (Shen et al., 1994). The number of RCCX modules varies in each individual from one to three. The bimodular structure is the most common. However, it has been reported that single RCCX modular structures have a frequency of 10-30% (Shen et al., 1994). Single modular RCCX structures consist of RP1, gene XB, and various types of C4 and CYP21 gene loci (Shen et al., 1994).

1.8.3 Exon and intron structure of C4

The exon and intron structure, and DNA sequence of a long (23 kb) C4A gene was reported by Yu (1991), who in addition provides a thorough overview of previous studies concerning the molecular organisation of the C4 gene locus. Yu (1991), found that the C4A gene spans 41 exons which encode a transcript for a precursor protein of 1744 amino acid residues (Figure 1.2). The intron size was found to vary from 82 bp to 7 kb (intron #9). The exon size varied between 51 bp to 232 bp. Analysis of intron sequences within the C4d region (polymorphic region of the α-chain), found these sequences to be highly conserved (Yu et al., 1986). The coding sequences of the proteolytic cleavage sites for the β-α and α-γ chain junctions are located in the middle of exons 16 and 33, respectively.

The β-chain consists of 656 amino acid residues, and is encoded by nucleotide bases in exons 1 to 16. Exon 1 encodes the leader sequence and the first three residues of the N terminus of the pro-C4 molecule. This fragment is unusual as some C4 genes encode a large intron (intron 9) of about 7 kb. This intron will be discussed in section 1.8.4 of this chapter.
The α-chain contains amino acid residues #661 to #1428, and is encoded by nucleotides in exons 16 to 33. The anaphylatoxin C4a is encoded by part of exon 16 and the whole exon 17, together, forming 77 amino acids. The two factor I cleavage sites in C4 are encoded by exons 23 and 30. These two cleavages generate the polymorphic C4d fragment, where thioester residues, isotypic residues, Rg/Ch antigenic determinants, and many of the allelic polymorphic residues are located. Exon 33 encodes the proteolytic cleavage site which forms the α- and γ-chains.

The γ-chain contains exons 33 to 41, and the nucleotides encode 291 amino acid residues. No specific function has yet been described for the γ-chain. This section of the gene is thought to be rather rigid as it possesses five intra-chain and two inter-chain disulfide linkages (Janatova, 1986; Seya et al., 1986).

1.8.4 C4 gene size variation

An interesting feature of the C4 gene locus, is the size variation of C4A and C4B genes. In general, the C4 gene located at locus I (usually C4A), is approximately 23 kb in length, whereas the gene at the second C4 gene locus (usually C4B), is either 23 kb or 16 kb in size (Prentice et al., 1986; Schneider et al., 1986; Palsdottir et al., 1987a) (Figure 1.2). C4 gene size variation results from the presence or absence of a 6-7 kb intron (separating exon 9 and 10), in the β-chain, which is approximately 2.5 kb from the 5' end of the C4 gene. It has been estimated that approximately half of the C4 genes at locus II are of the long 23 kb form (Schneider et al., 1986). The presence of the large 6-7 kb intron does not affect expression of the C4 gene harbouring it (Palsdottir et al., 1987b).

1.8.5 Molecular basis of C4A and C4B

Comparison of protein sequences derived from genomic DNA, cDNA, and partial protein sequencing data found 20 polymorphic residues, (reviewed by Yu, 1991). These polymorphic amino acid residues account for C4A/C4B isotypic differences, and Rg/Ch antigenic
determinants and some allelic differences. The distribution of these polymorphic residues between α, β, and γ C4 chains are shown in Figure 1.2. Those residues located in the C4d region of the α-chain are found in four discrete regions (Belt et al., 1984). The C4d region is a degradation fragment (Mr of about 40,000), from the middle of the α-chain (Reid, 1981). These polymorphic amino acid residues include residues #1054, #1090, #1101, #1102, #1105, #1106, #1157, #1182, #1188, and #1191, #1267, #1281, #1286, and #1287. Polymorphic nucleotides encoding amino acid residues #1101, #1102, #1105, and #1106 are known as the isotypic or class-specific nucleotides, and are either C4A or C4B nucleotides. Nucleotides encoding amino acid residues #1054, #1157, #1188, and #1191 are usually common to either C4A or C4B alleles, and are the Chido and Rodgers antigenic determinants. Six other polymorphic residues including three from the β-chain, two from the α-chain, and one from the γ-chain have been reported. These residues are thought to be C4 allelic differences.

Apart from nucleotide base differences, the two isotypes of C4 may also be distinguished by their slight differences in molecular weight. The α-chain of C4A has a slightly higher molecular weight than that of C4B (96 kD as opposed to 94 kD), as determined by polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulphate (Roos, 1982). The difference in molecular weight between these alleles is due to a specific amino acid residue difference ('Lys' or 'Pro'), for residue #1101 of the C4 precursor molecule (Rittner and Stradmann-Bellinghausen, 1990).

1.8.6 Functional differences between C4A and C4B isotypes

The isotypic amino acid residues situated in the C4d region are thought to be located on the surface of the C4 protein (Yu et al., 1988), where they are readily accessible for antigen/antibody interaction (Novotny et al., 1987). Each of the C4 isotypes differs markedly in haemolytic activity, covalent binding reactivity, and the expression of antigenic determinants, which can be detected using anti-Chido (detects C4B) and anti-Rodgers (detects C4A), alloantisera. In general, the C4A isotype has lower activity than the
C4B isotype in the classical haemolytic assay (Awdeh and Alper, 1980). This in turn is related to the ability of the C4B isotypic amino acid residues to preferentially bind covalently with acceptor hydroxyl groups that predominate amongst the carbohydrates of red cell surfaces (Isenman and Young., 1984). In contrast, C4A isotypic amino acid residues bind preferentially to amino groups such as those found in immune complexes (Law et al., 1984). It has been demonstrated that three to four times as much C4A binds to immunoglobulin as C4B (Kishore et al., 1988). This difference in haemolytic strength has been shown to result from polymorphism of residue #1106 (Carroll et al., 1989). Even though each C4 isotype has slightly different binding affinities, both C4A and C4B isotypes are capable of binding amino and hydroxyl groups, and therefore both isotypes are not required for survival. The importance of their preferential binding affinities has yet to be elucidated.

Peake et al. (1989), studied differences in the metabolism of C4A and C4B isotypes in healthy control subjects as well as in patients with active immunological diseases, such as SLE and rheumatoid arthritis. These studies investigated the significance of functional differences between these isotypes in vitro. Peake et al. (1989) has found that the C4B 1 protein is catabolised more rapidly than C4A 3 in patients with pathological complement activation compared to healthy controls. However, the molecular basis for this finding is unresolved.

1.8.7 Chido and Rodgers antigenic determinants

To test the antigenicity of C4, plasma containing this protein is examined for its ability to inhibit agglutination in an antiglobulin test. This type of serological test was originally developed by Grubb (1956), and was first applied to C4 antigenic determinant testing by Middleton and Crookston (1972). The Chido (Ch) antigenic determinants were originally identified by Harris et al. (1967), and Middleton and Crookston, (1972). Rodgers (Rg) antigenic determinants were defined by Longster and Giles (1976), who also found that some plasma samples exhibited partial inhibition in antigen testing, suggesting anti-Rg and anti-Ch are polyspecific and
that there may be more than one antigenic determinant for Rodgers and Chido. Giles (1985), proposed a model of multiple Ch and Rg antigenic determinants, which led to the development of standardised polyspecific anti-Rg and anti-Ch alloantisera. This in turn has allowed serological characterisation of many rare C4 allotypes and the subdivision of common allotypes. C4A proteins generally have Rodgers type 1 and 2 antigenic determinants (blood group antigens), and C4B proteins usually have Chido types 1, 2, 3, 4, 5 and 6 antigenic determinants.

These antigens were mapped to the HLA gene region by serological studies (Longster and Giles, 1976), and later were found to be part of the C4d region of C4A and C4B genes (Tilley et al., 1978). A characteristic antigenic site for a protein is thought to span about six amino acid residues (Kabat, 1970, Schechter, 1971), and these may be continuous (sequential), or discontinuous (conformational) in nature (Benjamin et al., 1984; Van Regenmortel, 1987). The proposed locations for Rodgers and Chido antigenic determinants in the C4 amino acid sequence proposed by Yu et al. (1988), correspond to these characteristics of antigen/antibody binding areas within a protein. C4A isotypic residues #1101-#1106 and residue 'Gln' encoded by nucleotide bases for residue #1157, are involved in the formation of the discontinuous Rg2 epitope. C4B isotypic residues on the other hand are related to the continuous Ch4, and discontinuous Ch2 epitopes, with residue 'Gly' encoded by nucleotides for residue #1154. Residues 'Val'-Asp'-Leu'-Leu' encoded by nucleotides for residues #1188 to #1191 of a C4A allele are thought to be related to the formation of the Rg1 epitope. The C4B residues 'A1a'-Asp'-Leu'-Arg' encoded by nucleotides for residues #1188 to #1191 are thought to be related to the continuous Ch1 and discontinuous Ch3 epitopes, the latter having residue 'Ser' encoded by nucleotides for residue #1157 (Yu et al., 1988; Giles et al., 1988).

1.8.8 C4 polymorphism and method of detection

Researchers investigating C4 antigenic determinants were the first to establish the molecular basis for much of the polymorphism of C4 proteins. The C4A 1 and C4B 5 alleles have C4A and C4B isotypic
properties respectively. However, these alleles were found to express the reversed antigenic determinants, where \( C4A \) \( 1 \) had \( C4B \) and \( C4B \) \( 5 \) had \( C4A \) antigenic determinants (Roos \textit{et al.}, 1984a and b). Detailed serological analysis of 'B \( 5 \) ' led to the discovery of three additional specificities of anti-Ch and their corresponding antigen determinants on \( C4 \), which were designated Ch4, 5 and 6 (Giles, 1987).

Researchers investigating Chido and Rodgers antigenic determinants, found it difficult to assign specific Rg/Ch determinants to each \( C4 \) allele in individuals with heterogeneous \( C4A \) and/or \( C4B \) alleles. Family studies were therefore necessary (Yu \textit{et al.}, 1988). Serological observations by Roos \textit{et al.} (1984a) and Giles (1987), enabled a number of conclusions to be drawn concerning the relationship between Chido/Rodgers antigenic determinants and \( C4A/C4B \) isotypes (summarized by Yu \textit{et al.}, 1988). However, a degree of uncertainty was still associated with correlating Ch/Rg antigenic determinant types with specific \( C4A \) and \( C4B \) alleles.

The more commonly used method for detecting \( C4 \) polymorphism is \( C4 \) protein gel electrophoresis followed by immunofixation. This technique does not rely on family analysis. Originally this method could only detect 'fast' and 'slow' types of \( C4 \) alleles. However, with the advent of neuraminidase treatment of serum samples prior to electrophoresis (Awdeh and Alper, 1980), extensive polymorphism at both the \( C4A \) and \( C4B \) loci has been detected. A total of 35, \( C4 \) alleles (13 \( C4A \) and 22 \( C4B \)), including the less common variants, have been assigned (Mauff \textit{et al.}, 1983).

In parallel with advances in \( C4 \) protein gel electrophoresis, progress was made in the structural studies of \( C4 \) including determination of the \( C4d \) region amino acid sequence (Campbell \textit{et al.}, 1981; Harrison \textit{et al.}, 1981; Chakravarti \textit{et al.}, 1983; Hellman \textit{et al.}, 1984). This was followed by cDNA and genomic sequencing (Carroll \textit{et al.}, 1983; Carroll \textit{et al.}, 1984a,b; Belt \textit{et al.}, 1984, 1985). The estimate of 35, \( C4 \) alleles probably represents a minimum. This is because the most common alleles, 'A \( 3 \)' and 'B \( 1 \)', \{which have gene frequencies of approximately 60 \% to 70 \% (Schendel \textit{et al.}, 1984)\}, have been
subdivided on the basis of serological determinants (Giles, 1984a), DNA restriction fragment length polymorphism (RFLP) (Palsdottir et al., 1987a), and nucleotide sequencing (Belt et al., 1985).

1.9 C4 analysis in other animals - purpose and findings

Apart from humans, organisation and structure of C4 genes in the C4 gene region has been studied in a variety of other mammals such as those shown in Tables 1.1, 1.2 and 1.3. The molecular basis for evolution of the C4 gene region has also been studied. The C4 gene region has been used as a model for eukaryotic gene regulation and multigene family studies.

Many studies concerning organisation of the C4 gene region, structure of C4 genes, and most importantly C4 gene regulation studies, have centered on murine C4 genes. A summary of C4 research findings in mice is shown in Table 1.1. Many other mammals have also been used for analysis of the C4 gene region for both gene organisation and evolutionary studies (Table 1.2).

Guinea-pigs have been used to investigate the molecular basis for C4 protein deficiency (review by Colten, 1982; Whitehead et al., 1983). Whitehead et al. (1983), used a human cDNA clone encompassing the human C4 γ-chain for hybridisation analysis of normal and C4-deficient guinea-pig-liver mRNA. A 5 kb transcript was found in normal C4 productive liver. However, in C4-deficient liver a 7 kb transcript was detected, indicating the presence of C4 precursor RNA. This analysis suggested that the basis of C4 deficiency in guinea-pig is a post-transcriptional defect in the processing of C4 precursor RNA to mature C4 mRNA. Whitehead's study has shown the usefulness of human cDNA probes in the investigation of the molecular basis of C4 deficiency. Evolutionary studies of the C4 gene region have focussed on the analysis of primates, and results of these findings are shown in Table 1.3. In addition to evolutionary studies, the analysis of C4 in primates has been aimed at identifying ancestral haplotypes which are common to both humans and other primates. These haplotypes may represent genes which have a
<table>
<thead>
<tr>
<th>location</th>
<th>gene structure</th>
<th>animal/human homology</th>
<th>investigation</th>
<th>findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>chromosome 17</td>
<td>two C4 gene loci</td>
<td></td>
<td></td>
<td>- two major C4 alleles, C4-high and C4-low</td>
</tr>
<tr>
<td>MHC</td>
<td>functional C4 gene</td>
<td></td>
<td></td>
<td>for high and low C4 protein-producing strains</td>
</tr>
<tr>
<td>S region of the H-2 complex</td>
<td>sex-limited (Slp) gene dysfunctional</td>
<td></td>
<td></td>
<td>(5). These strains show 10 to 20 fold difference</td>
</tr>
<tr>
<td></td>
<td>due to three amino acid deletion in the α-chain (1)</td>
<td></td>
<td></td>
<td>in serum C4 levels</td>
</tr>
<tr>
<td></td>
<td>- C4 shows 96% homology to Slp</td>
<td></td>
<td></td>
<td>- a cis-acting element closely linked to C4</td>
</tr>
<tr>
<td></td>
<td>- complete sequence known (2)</td>
<td></td>
<td></td>
<td>controls the serum C4 level at the pre-</td>
</tr>
<tr>
<td></td>
<td>- 15.5 kb in length</td>
<td></td>
<td></td>
<td>translational level (6) and post-transcriptional</td>
</tr>
<tr>
<td></td>
<td>- 41 exons</td>
<td></td>
<td></td>
<td>level (7). Low C4 production is liver tissue-</td>
</tr>
<tr>
<td></td>
<td>- encodes 1738 amino acids</td>
<td></td>
<td></td>
<td>specific, whereas steady state levels of C4 mRNA</td>
</tr>
<tr>
<td></td>
<td>- mRNA 5372 bp</td>
<td></td>
<td></td>
<td>in monocytes and macrophages (second major site</td>
</tr>
<tr>
<td></td>
<td>- Slp 1.1 kb shorter than C4 due to</td>
<td></td>
<td></td>
<td>of C4 synthesis), are similar in low and high C4</td>
</tr>
<tr>
<td></td>
<td>deletion of repetitive sequence and an MT element</td>
<td></td>
<td></td>
<td>producing strains of mice (8)</td>
</tr>
<tr>
<td></td>
<td>- duplicated 21-hydroxylase gene one functional and one dysfunctional</td>
<td></td>
<td></td>
<td>- the genetic basis for C4 low producing mice</td>
</tr>
<tr>
<td></td>
<td>- dysfunctional 21-hydroxylase gene is lacking the second exon and some adjacent sequence (3)</td>
<td></td>
<td></td>
<td>results from abnormally processed C4 mRNA by</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>model to show allelic variation of multigene families in MHC</td>
<td>alternative splicing. A retroposon-like insertion</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>of B2 sequence (short repetitive genetic element</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>in rodents), into intron 13 of the C4 gene (8 and 9)</td>
</tr>
</tbody>
</table>

References for the above studies include: (1) Ferreira et al. (1978); (2) Ogata et al. (1989); (3) Miller et al. (1992); (4) Farries and Atkinson et al. (1991); (5) Shreffler (1976); (6) Ogata and Sepich (1984); (7) Nakayama et al. (1990a); (8) Pattanakilsakul et al. (1992a); (9) Zheng et al. (1993) (10) Huang et al. (1991).
<table>
<thead>
<tr>
<th>Mammal Description</th>
<th>C4 Gene Locus</th>
<th>Protein Binding Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dog (1 and 2), Cat (3)</td>
<td>Single</td>
<td>C4B</td>
</tr>
<tr>
<td>Guinea-pig (4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syrian-Hamster (5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Several species of whale (6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mice (7), Pig (8), Horse (9)</td>
<td>Duplicated</td>
<td>C4A and C4B</td>
</tr>
<tr>
<td>Rat (10), Goat (11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cattle (12), Sheep (13)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primate (14), Human (15)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

References for C4 gene locus studies for each mammal include; (1) Kay and Dawkins (1984a); (2) Doxiadis et al. (1985); (3) Dekroone et al. (1986); (4) Bitter-Suermann et al. (1977); (5) Levi-Strauss et al. (1985); (6) Spilliaert et al. (1990); (7) Roos et al. (1978); (8) Kirstenbaum et al. (1985); (9) Kay et al. (1987a); (10) Tosi et al. (1985); (11) Cameron et al. (1990); (12) Groth et al. (1987); (13) Groth et al. (1988); (14) Kawaguchi et al. (1992a); (15) Yu (1991). Dodds and Law (1990), investigated the C4 isotypic binding activity for the majority of mammals described above.
Table 1.3 C4 gene features of humans and other primates

<table>
<thead>
<tr>
<th>C4 gene feature</th>
<th>human</th>
<th>chimpanzee</th>
<th>gorilla</th>
<th>orangutan</th>
</tr>
</thead>
<tbody>
<tr>
<td>C4A and C4B gene duplication</td>
<td>yes</td>
<td>yes (1)</td>
<td>yes (1)</td>
<td>yes (1)</td>
</tr>
<tr>
<td>gene duplication</td>
<td>yes</td>
<td>yes (1,3 and 4)</td>
<td>yes (2)</td>
<td>yes (2)</td>
</tr>
<tr>
<td>gene size</td>
<td>long and short</td>
<td>short (1)</td>
<td>short (1)</td>
<td>long and short (1)</td>
</tr>
<tr>
<td>null alleles</td>
<td>deleted and non-deleted</td>
<td>non-deleted (1)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>21-hydroxylase defect</td>
<td>many types including an 8 bp deletion</td>
<td>8 bp deletion (1)</td>
<td>non-deleted (1)</td>
<td>non-deleted (1)</td>
</tr>
<tr>
<td>human/primate homology</td>
<td>extensive homogenisation</td>
<td>extensive homogenisation</td>
<td>extensive homogenisation</td>
<td>extensive homogenisation</td>
</tr>
<tr>
<td>C4A/C4B hybrid</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>polymorphism</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>

References for these investigations include; (1) Christiansen et al. (1991); (2) Kawaguchi et al. (1992a); (3) Kawaguchi et al. (1990); (4) Bontrop et al. (1991). Human/primate homology, C4A/C4B hybrid and C4 polymorphism studies were by Kawaguchi et al. (1992b). All human analyses were described in Chapter 1 of this thesis. ‘-’ not investigated.
selective advantage for the elimination of infectious agents, and may be able to withstand harsh environmental conditions.

Recent studies of the \( C4 \) gene region in sheep and cattle (ungulates) (Table 1.2), have concentrated on the residues encoded by nucleotides at the \( C4A/C4B \) isotype gene region with an aim to gaining some insight into the possible evolutionary route of the \( C4A \) and \( C4B \) genes in comparison with that of the primates (Ren et al., 1993). Comparing the nucleotide sequences of the \( C4A \) and \( C4B \) isotype region between these species enabled Ren et al. (1993) to propose an evolutionary model for this region of the genome. This is described in the following section.

1.10 Evolution of the \( C4 \) gene region

1.10.1 Multiple copies of a \( C4 \) module

The \( C4 \) and \( CYP21 \) genes are part of a modular structure which includes other closely linked genes (see Chapter 1. section 1.8.2). Some species carry multiple copies of this module which appears to have undergone continuous expansion and contraction (Kawaguchi et al., 1991). The original \( C4 \) gene may have been a \( C4B \)-like gene, as this gene has been detected in all mammals. The \( C4A \) gene may have arisen independently after speciation in some mammals such as mice, primates, and ungulates, as sequences in both \( C4A \) and \( C4B \) genes from these mammals are quite different from each other (Ren et al., 1993). Alternatively, duplication of the \( C4 \) gene region may have occurred only once in the history of placental mammals, and as a result of continuous homogenisation (concerted evolution) in each species, these genes became quite different between the species (Kawaguchi et al., 1991).

Two copies of the \( C4 \) module appear to be standard in humans. However, haplotypes with one, three, or four modules are quite common. In Caucasians, for example, haplotypes with one or three modules occur at frequencies of 16 % and 18 %, respectively (Teisberg et al., 1988). This modular polymorphism is thought to arise by unequal cross-over between misaligned modules (Raum et
al., 1984; Carroll et al., 1985a and b; Donohoue et al., 1986a; Rodrigues et al., 1987). This cross-over theory has been supported by pedigree analysis of one family (Sinnott et al., 1990). Originally it was thought unequal cross-over only occurred between the boundaries of \textit{C4} and \textit{CYP21} genes (White et al., 1984). However, it has more recently been shown that cross-overs can occur between individual exons of these genes (Donohoue et al., 1989). Analysis of many haplotypes has shown that no matter how the \textit{C4} modules rearrange, they retain their modular organisation.

If originally there was a single copy of the \textit{C4} module, it is difficult to explain how the second copy arose. It has been suggested that possibly the original cross-over between chromatids, possessing the \textit{C4} alleles, occurred between homologous sequences flanking the \textit{C4} module (Kawaguchi et al., 1991). This exchange would result in one chromosome where the gene has been deleted, and another where gene duplication has occurred (Figure 1.3). Alternatively, it is possible that the duplication arose as a result of insertion of movable repetitive elements epitomized by \textit{Alu} repeats. Frequent occurrence of recombination leading to changes in gene copy numbers has been reported to be associated with the insertion of \textit{Alu} elements in several genetic systems (Ottolenghi and Giglioni, 1982; Lehrman et al., 1985; Rouyer et al., 1987; Miura et al., 1989). As neither \textit{Alu}, nor any other repetitive element has been found in the vicinity of the \textit{C4} and \textit{CYP21} genes in any species, it has been assumed that the original duplication arose via the first described mechanism of duplication and deletion (Kawaguchi et al., 1991).

A study by Gitelman et al. (1992), located the duplication boundaries between \textit{CYP21A} and \textit{C4B} by sequencing this region known as the \textit{XA} locus. This analysis indicated that the duplication occurred by non-homologous recombination. These boundaries are substantially different to those in the mouse genome, suggesting similar gene duplications have occurred independently in ancestors of rodents and primates after mammalian speciation. Those proponents of the \textit{C4} module duplication arising prior to speciation of mammals, believe that even though mice \textit{C4} genes form one
Figure 1.3 Model for deletion and duplication within the C4 gene region
phylogenetic group, and humans another, it does not prove the two groups originated by independent duplications (Kawaguchi et al., 1991). These researchers believe unequal cross-over would have homogenised the genes to look more alike within each of the two orders than between them.

1.10.2 $C4A$ and $C4B$ isotypic residues and $C4$ gene size variation

The $C4A$ and $C4B$ isotypic residues appear to have diverged before that of the orangutan and the African-ape lineages more than 12 million years ago, whilst the remaining $C4d$ region of the $C4$ gene appeared to diverge separately in each of the four species after their formation (Kawaguchi et al., 1991). Some hypothesise that most of the $C4d$ region (apart from the isotypic residues) have homogenised. Analysis of the presence or absence of a 7 kb intron in humans and other primates indicates that the original $C4$ gene was of the short type, and that the insertion must have occurred before the divergence of the orangutan and the African-ape lineages, but after the original duplication in one of the two $C4$ genes (Kawaguchi et al., 1991). Also, the transfer of the insert from one copy ($C4A$ or $C4B$), to its duplicate must have occurred by unequal cross-over (Kawaguchi et al., 1991).

1.10.3 Evolution of $C4$ polymorphism

The origin of, and genetic mechanisms responsible for the extensive polymorphism detected in HLA class I, II and III loci, are of considerable interest in evolutionary studies. Discovery of the remarkable patchwork pattern of allelic sequence diversity at the human class II loci (Bodmer et al., 1991; Gyllensten et al., 1991), and the example of gene conversion in the bm12 mouse mutation (McIntyre and Seidman, 1984; Mengle-Gaw et al., 1984), suggested recombination events may have been important for the generation of polymorphism, (reviewed by Erlich and Gyllensten, 1991).
An important factor which has played a major role in the evolution of polymorphism in the \textit{C4} gene region is the mechanism of genetic exchange between chromatids. Many of these studies have investigated the \textit{CYP21} genes, as the dysfunctional types appear to have arisen from genetic exchange between the \textit{CYP21P} and \textit{CYP21} genes. Two explanations have been put forward for the apparent transfer of sequence between these two genes, namely, gene conversion (Donohoue \textit{et al}., 1986b; Matteson \textit{et al}., 1987; Higashi \textit{et al}., 1988; Miller \textit{et al}., 1988), and unequal cross-over (Raum \textit{et al}., 1984; Carroll \textit{et al}., 1985a, and b; Donohoue \textit{et al}., 1986a; Rodrigues \textit{et al}., 1987).

1.10.3.1 Gene conversion mechanism

Gene conversion can be defined as a recombination event over a short region between alleles or loci where the flanking markers remain in the parental configuration, a pattern which is indistinguishable from a double cross-over event. The donor gene remains unchanged. The length of DNA sequence exchanged is uncertain and may range from a hundred bases to kilo bases (Kourilsky, 1986).

Gene conversion has gained some attention over the years as molecular studies have shown many instances of non-reciprocal recombination in bacteria, fungi, yeast and mammals (reviewed by Kobayashi, 1992). Evidence of this type of genetic exchange is commonly found in repeated genes, and is thought to have played an important role in the evolution of these gene regions. This mechanism appears to have specific functions in different organisms. For example, yeast cells alter their mating type by non-reciprocal recombination. This mechanism creates diversity of immunoglobulin genes in chickens, and is thought to lead to antigenic variation, and to be responsible for the movement of introns between homologous genes.
The mechanism of gene conversion was most probably observed as early as 1926 in studies of fungi, when Brunswick noticed that recombination may not always be reciprocal. In 1930, Winkler hypothesised that those cases of non-Mendelian segregation were a result of recombination where a physiological process of one allele influences its counterpart on the homologous chromosome, and physiologically converts it to its own kind. This hypothesis was not recognised by the scientific community at that time, although was resurrected in 1949 by Lindgren as a synonym for non-reciprocal genetic recombination. The hypothesis has been retained in this form by fungal geneticists.

The current genetic theory, from studies of yeast and fungi, states that non-reciprocal recombination (gene conversion), is the result of DNA repair following intragenic cross-over (Holliday, 1964). The theory postulates that cross-over between either the gene itself or its RNA transcript (Kourilsky, 1986), leads to the formation of DNA heteroduplexes (molecules containing one strand from one chromatid and the other from the homologous chromatid). An alternative theory put forward is that a block of donor strand DNA is accidently copied into the cognate position of the recipient strand during DNA synthesis (Kourilsky, 1986). As the point of synthesis migrates along the gene and reaches nucleotides which are different between the two alleles, the non-complementary nucleotides remain unpaired. DNA synthesis proof-reading enzymes then recognise the unpaired nucleotides, excise one of the nucleotides, and replace it with a complementary nucleotide, and thereby convert one of the alleles to the other (Kawaguchi et al., 1991).

One of the major problems which confounds the understanding of gene conversion is the process of heteroduplex formation, and whether heteroduplex formation actually occurs. A study of monkey cells detected heteroduplex formation in vitro and their fate was observed in vivo. This study found that blocks rather than single nucleotides were exchanged, resulting in patchworks of the parental sequences. A technique was designed to detect heteroduplex formation in yeast during meiotic recombination, by using a denaturing gel procedure (Lichten et al., 1990). Recent studies of
this mechanism were based on detection of heteroduplex formation by identifying stem-loop or hairpin configuration of nucleotide base sequence using yeast as a model system (Nag and Petes, 1993). Until recently, the majority of studies on the molecular basis of this mechanism have involved lower eukaryotes, as all of the products of a single meiosis can be recovered and analysed, whereas this has not been possible in mammals.

1.10.3.2 Gene conversion in mammals

Slightom *et al.* (1980), were the first to suggest that a gene conversion-like mechanism occurs in mammalian genes. The study of this mechanism in mammals is difficult, as geneticists must study products of many cell generations, often, many generations removed from the initial meiosis which created the recombinant. In addition, low probability recombination or mutation events are most likely to have occurred after a number of generations. Therefore, instead of being the product of gene conversion, the recombinant sequence could be the result of two independent, sequential reciprocal cross-over events. Also, under most circumstances, it is virtually impossible to differentiate between the products of gene conversion and those of double cross-over. If the occurrence of the recombinant sequence is associated with deletion or duplication of chromatin, then it is more likely to be the outcome of unequal cross-over than gene conversion (Collier *et al.*, 1989).

1.10.3.3 Gene conversion and the C4/CYP21 module

Early studies of the *CYP21* gene region in congenital adrenal hyperplasia (CAH) patients by RFLP analysis, found 30 % suffered from this disease due to a deleted *CYP21* gene (Schneider *et al.*, 1986; Werkmeister *et al.*, 1986; Jospe *et al.*, 1987; Rumsby *et al.*, 1986; White *et al.*, 1988). These deletions were thought to result from unequal cross-over. Short-range restriction mapping (Donohoue *et al.*, 1986a; Jospe *et al.*, 1987; Matteson *et al.*, 1987; Rumsby *et al.*, 1986; White *et al.*, 1988), based on the use of specific oligonucleotide probes (Higashi *et al.*, 1988), and sequencing (Harada *et al.*, 1987; Higashi *et al.*, 1988), of the *CYP21-C4* gene region,
provided evidence of inter-locus transfer of sequences between alleles in this region. Since this time there have been over 200 reported cases of apparent gene conversion events between $CYP21P$ and $CYP21$ (Collier et al., 1993). Some investigators (Matteson et al., 1987; Miller, 1988), argue that the apparent deletions were in fact genes which had evolved by gene conversion. However, pulsed-field gel electrophoresis (PFGE) studies have authenticated deletions of this gene region by demonstrating that portions of chromatin are missing in some of the CAH patients (Collier et al., 1989), and therefore more likely to result from unequal cross-over.

1.10.4 Gene duplication and evolution of the $C4$ gene region

Bimodular $C4/CYP21$ haplotypes are common in humans, chimpanzee, gorilla, mouse, rat, pig and the macaque (Mevag et al., 1983). Monomodular haplotypes are common in Syrian hamster, dog, cat, guinea-pig and several species of whale, whilst trimodular haplotypes have been reported in humans and the orangutan. The presence of multi-modular haplotypes provides the potential for unequal cross-over, and therefore instability of the $C4/CYP21$ gene region (Kawaguchi et al., 1991). The fact that these multimodular species have not reverted back to a single module suggests there is a selective advantage in having more than one module. There does not appear to be an advantage of having multiple copies of $CYP21$, as all but one $CYP21$ allele are dysfunctional when multiple copies of this allele are present on a haplotype (Kawaguchi et al., 1991). This observation also suggests that multiple copies of $CYP21$ may produce an undesirable condition (Kawaguchi et al., 1991). We are still learning about other genes in the $C4$ gene region, and therefore uncertain of the contribution that these genes make on the stability of this region. Much attention has been given therefore to the duplicated $C4$ genes. It appears there may be a selective advantage to having more than one $C4$ module, as complete deficiency is very rare, and most $C4$ deficient patients develop SLE and other autoimmune diseases. This is presumably because of decreased level of circulating $C4$ protein which is required to clear immune complexes from the circulation.
Partial C4 deficiency occurs frequently within a population. Approximately 10% of Caucasians have only two C4 genes expressed, 13% - 24% have three genes expressed, and 69% - 76% have four genes expressed in a diploid genome of four C4 genes (Kawaguchi et al., 1991). In the Caucasian population, the frequency of homozygous C4A deficient individuals is 2.5% - 3.5%, and homozygous C4B deficient individuals is 1% - 2% as determined by C4 protein gel electrophoresis (Hauptmann et al., 1986). The effects of partial C4 deficiency are far less drastic than those of complete C4 deficiency. However, studies have found an association between partial C4 deficiency and susceptibility to development of a range of diseases, such as, SLE, IDDM, and Graves disease (Dawkins et al., 1983). These studies showed that individuals with more than one copy of C4 in their genome have less chance of contracting these types of diseases, thereby suggesting a selective advantage of possessing multiple copies of this gene. It is also possible that other recently identified genes in this gene region may convey a selective advantage by possessing multiple copies. In these cases where a single C4 module has been detected in a species by C4 protein analysis, it is possible that more than one C4 module is present because there has not been confirmation by cloning. However, in the mouse, where a bimodular arrangement of C4 genes has been detected, only one of these genes is functional (Ferreira et al., 1978). It is possible that species with a monomodular arrangement of C4 genes may have a C4 gene which has been re-modelled, thus enabling the product of this gene to carry out functions executed by both C4A and C4B type genes (Kawaguchi et al., 1991).

Researchers who have studied the evolution of the C4/CYP21 gene region have suggested it may be under two opposing selective pressures, one which selects for the presence of two C4 genes, and the other which selects for a single CYP21 gene (Kawaguchi et al., 1991). This differential selection may be responsible for the presence of monomodular haplotypes in some species and bimodular haplotypes in others. Kawaguchi et al. (1991), suggest that if this arrangement between C4 and CYP21 genes is undesirable, this may explain why all but one copy of CYP21 is dysfunctional in a multi-modular haplotype.
1.11 Molecular analysis of \textit{C4} null alleles

A major factor preventing the development of a molecular-based technique for \textit{C4} allotyping is the accurate characterisation of \textit{C4} \textit{A} and \textit{C4B} null alleles. Four types of \textit{C4} null alleles have been reported and include: (1) Deletion of a \textit{C4} gene locus which usually occurs together with a \textit{CYP21} gene (Carroll \textit{et al.}, 1985b and c; Schneider \textit{et al.}, 1986); (2) \textit{C4A} mutation as a result of a small insertion (2 bp) leading to a stop codon and truncated mRNA product (Barba \textit{et al.}, 1993); (3) Two \textit{C4A} or \textit{C4B} alleles on the same haplotype which are associated with a \textit{C4B} or \textit{C4A} null allele {for example \textit{C4A} 2,3 \textit{B} \textit{QO} (isoexpression), or \textit{C4A} 3,3 \textit{BQO} (homoexpression)}, (Raum \textit{et al.}, 1984; Rittner \textit{et al.}, 1984; Uring-Lambert \textit{et al.}, 1984; Yu and Campbell, 1987); and (4) Uring-Lambert \textit{et al.} (1989), and Barba \textit{et al.} (1993) have reported a non-deleted type of \textit{C4} null allele, the molecular basis of which is yet to be established.

1.11.1 \textit{C4} deletions

Only large deletions of 28 kb to 30 kb including a \textit{C4} gene locus and usually a neighbouring \textit{CYP21} gene have been detected. Deletion of the entire \textit{C4} gene region has not been reported. It has been suggested that the large intron (7kb), (absent in short forms of \textit{C4B}), may affect the site of recombination of \textit{C4B} and \textit{CYP21B} in an unequal cross-over event leading to \textit{C4} gene locus deletions (Yu \textit{et al.}, 1986). Studies of haplotypes with a deleted \textit{C4A} gene have found that the remaining \textit{C4B} genes were almost always the short form of \textit{C4} gene (Carroll and Alper, 1987b).

The types of \textit{C4} deletions detected thus far include \textit{CYP21A/C4B}, \textit{C4B/CYP21B}, and \textit{C4A/CYP21A} leaving a long 23kb \textit{C4B} gene, and \textit{C4A/CYP21A} leaving a short 16kb \textit{C4B} gene on the chromatid (Schneider \textit{et al.}, 1986). These large deleted types of \textit{C4} null alleles can be detected by RFLP analysis using restriction enzyme \textit{Taq I}, or restriction enzymes \textit{Bgl II}, \textit{Kpn I}, \textit{Bam HI}, and \textit{Hind III} (Carroll \textit{et al.}, 1985b).
1.11.2 C4A gene mutation resulting from a small insertion

Recent studies by Barba et al. (1993), detected an intact 23 kb C4A type of gene which is dysfunctional due to a small 2 bp insertion in exon 29. This mutation was found to lead to a termination codon during transcription in 10 of 12 individuals with non-deleted types of C4A null alleles. For one of the two remaining C4A non-deleted null allele individuals, evidence was obtained for gene conversion to the C4B isotype. The basis for the other type of C4A null allele was unresolved.

1.11.3 Isoexpression, homoexpression and hybrid C4 null alleles

There are two types of non-deleted C4 haplotypes, associated with C4 null alleles, which have been identified (Partanen and Campbell, 1989). One type has C4 genes with the same C4A or C4B polymorphisms and antigenic determinants otherwise known as isoexpression and homoexpression, for example, C4A null alleles with Rodgers antigenic determinants, and C4A isotypic polymorphisms. The second type has C4 genes with mixed C4A and C4B polymorphisms and antigenic determinants, also described as C4A/C4B hybrid types of alleles, for example, C4A null alleles with Chido antigenic determinants and C4A isotypic polymorphisms. The reason for non-expression of the hybrid type of C4 null allele, which shares both C4A and C4B characteristics, has yet to be established.

Identification of homoexpression and isoexpression types of C4 haplotypes was based on extensive family studies (Schneider et al., 1986; Palsdottir et al., 1987b; Fredrikson et al., 1991). Schneider et al. (1986), analysed RFLP typing patterns of 126 haplotypes from 92 individuals in 32 Caucasian families, including eight patients with CAH, five patients with celiac disease, and four patients with multiple sclerosis, to determine the molecular basis of C4 null and duplicated variant allotypes. Healthy subjects were selected for the presence of haplotypes carrying C4 null alleles at one or both of the C4 gene loci. More than half of the C4 genes typed as C4 null by protein gel
electrophoresis resulted from large deletions. Also, several unrecognised homo-duplicated C4 alleles were detected.

It is thought isoexpression and homoexpression on a haplotype are the second product of unequal cross-over between homologous chromosomes (Schneider et al., 1986). This results in one haplotype with duplicated C4 genes, whilst the second haplotype has a deleted C4 gene locus (Figure 1.3). Schneider et al. (1986), suggested that the frequencies of null alleles determined by Fielder et al. (1983), and duplications determined by Raum et al. (1984), (both of which were based on C4 protein typing data), may be underestimated, and that C4 gene deletions and duplications may be more common than previously thought. Schneider et al. (1986), found all non-deleted C4B null alleles were of the large 23 kb gene size. This was likewise reported by other researchers such as Carroll and Alper (1987b), and Jenhani et al. (1992). This is an interesting observation as all duplicated C4 genes resulting in isoexpression and homoexpression are associated with long C4 genes at both loci I and II. It may be possible in these cases, that the haplotype having the C4B null allotype may consist of two functional C4A alleles and no C4B allele.

The C4A/C4B hybrid type of C4 null allele, described above, has similar properties to the C4A *1 and C4B *5 allotypes (Yu et al., 1986). It has been suggested that a gene conversion-like mechanism is responsible for mixing the generally discrete serological allotypes between C4A and C4B in these alleles (Palsdottir et al., 1987b). The hybrid type of C4 allele reported by Yu and Campbell (1987), was detected by RFLP analysis. Restriction analysis of a C4A gene with Nla IV results in fragment sizes 276 bp and 191 bp. A C4B gene, however, produces a fragment size of 467 bp. DNA encoding Rg1 antigenic determinants produce 565 bp Eco0109 fragments by restriction analysis, whereas DNA encoding Ch1 produces 458 bp fragments (Partanen and Campbell, 1989).

In summary, both large deletions of C4A and C4B alleles, and the small 2 bp insertion in C4A null alleles, can be identified without family analysis, as the molecular basis of these C4 null allotypes has been determined. Gene duplications are associated with the C4 null
allotype, and occur with quite high frequencies. However, gene duplications reported as homoexpression and isoexpression cannot be assigned with a high degree of certainty without family studies. There is now evidence for hybrid types of C4 genes being associated with the C4 null allotype. However, the molecular basis for this association has not been established. In addition, as mentioned above, there are a number of other C4 alleles which appear to be intact but are dysfunctional. The molecular basis of these is unknown.

1.11.4 Association of C4 null alleles with HLA haplotypes

C4A and C4B null alleles appear to be associated with specific haplotypes in different ethnic populations. Certain C4 gene deletion combinations were found to segregate with specific extended haplotypes in Caucasians, such as C4B/CYP21 with HLA-Bw 47, DR 7, FC 91,0, and C4A/CYP21A with HLA-A 1, Cw 7, B 8, DR 3, C4A Q0, C4B 1 (Schneider et al., 1986). The allelic combination HLA-B 18, Bf F1, C2 C, C4A 3, C4B Q0, DR 3 was shown to be associated with IDDM in Caucasians (Dawkins et al., 1983), and HLA-A 2, C w5, B 44, Bf S, C4A 3, C4B Q0, DR 4 associated with D-Penicillamine-induced Thrombocytopenia in Caucasian patients (Dawkins et al., 1983). However, the molecular basis of these null alleles was not determined in these studies. The 2 bp insertion mutation (described above), was found to be in linkage disequilibrium with the haplotype HLA-B60, DR6 in seven of 10 Caucasian donors (Barba et al., 1993).

In Thai/Chinese over 90% of C4A Q0 alleles are associated with two particular allelic combinations, namely HLA-A w33, B 44, C4A Q0, C4B 1, Bf S, DR 7, and HLA-A w33, B 17, C4A Q0, C4B 1, Bf F, DR x (where x is an unassigned DR antigen), (Kay et al., 1987b). However, the molecular basis for the C4 null allotype was not determined. These two allelic combinations have minimum population frequencies of 6% and 3%, respectively. In the Thai/Chinese population, 70% of C4B Q0 alleles were associated with the allelic combination HLA-A w33(11), B 17, C4A 3, C4B Q0, Bf S, DR 3, and the null allele had a minimum frequency of 8% (Kay et al., 1987b).
Very few studies have investigated the association of non-deleted C4 null alleles with particular HLA alleles. Schneider et al. (1986), did not detect linkage disequilibrium of non-deleted C4A *Q0 alleles with any specific HLA type. However, a study by Kemp et al. (1987), suggested that non-deleted C4A *Q0 alleles are associated with DR 2 in white Caucasian SLE patients. Schneider et al. (1986) found that 12 of 17 non-deleted C4B QO alleles were associated with HLA-B 44, Bf S, C2 C, C4A 3, C4B Q0, and seven with the extended haplotype B 44, DR 4, Bf S, C2 C, C4A 3, C4B Q0. All non-deleted C4B null alleles were found in the common Bf S, C2 C, C4A 3, C4B Q0 and Bf F, C2 C, C4A 3, C4B Q0 complotypes.

1.12 Summary of review material most relevant to research studies undertaken for this thesis

A wealth of information is available concerning the C4 protein, the gene region which encodes this protein, and the role of this protein within the classical complement activation pathway. Studies of the association between C4 null alleles and autoimmune disease have indicated a number of areas needing further research. These include the role of C4 protein in the immune response, the factors which regulate C4 protein activity, the molecular basis of C4 alleles including C4 null alleles, and genetic mechanisms operating within the C4 gene region which have led to its evolution.

Many of the earlier C4 protein functional and disease association studies were based on C4 protein allotyping. However, availability of contemporary molecular techniques such as restriction fragment length polymorphism (RFLP), allele-specific oligonucleotides (ASO), pulsed-field gel electrophoresis (PFGE), polymerase chain reaction (PCR), and DNA sequencing, has facilitated the molecular analysis of the C4 gene. This has provided much of the basic information and methodology required for developing a molecular-based C4 allotyping protocol, a major aim of this thesis.

The C4 gene locus has proved to be more complicated than originally thought. Not only does the C4 gene region have two genes, C4A and C4B, but occasionally C4 may have two C4A- or two C4B- like
genes on the same haplotype. C4 alleles including some null alleles have been found to have mixed C4A and C4B antigenic determinants. Also, some C4A alleles have antigenic determinants commonly found in C4B alleles, and some C4B alleles have antigenic determinants commonly found in C4A alleles.

Previous research has led to changing hypotheses as to which factors generate a C4 Q0 allotype. Initial studies of the C4 null allele found that large deletions were the molecular basis of this allotype. However, family studies have shown that some C4 Q0 allotypes are not the result of gene deletion but rather two alleles of the same isotype. The frequencies of null alleles and frequencies of duplications in association with C4 Q0 alleles (0.10-0.15, and 0.0075-0.01, respectively), do not correspond in the Caucasian population. This indicates that other genetic mechanisms besides C4 gene duplication are operating to create this allotype. More recent molecular studies of C4 null alleles have shown that a number of genetic mutations are responsible for the C4 null allotype (as occurs with the accompanying CYP21 gene loci). Besides the earlier described large deletion type of C4 null alleles, a 2 bp insertion, and apparently intact alleles which have unusual combinations of C4A and C4B isotypic amino acid residues, have been found in association with this allotype. The molecular basis of the C4A/C4B hybrid types of C4 Q0 allotype remains unresolved. Only RFLP analysis has been applied to these aberrant null alleles to study the known polymorphic sites.

The presence or absence of the large 7 kb intron at the second C4 gene locus has probably increased the instability of this gene region. This would therefore be a promising site to investigate the molecular basis of C4 null alleles. The interaction of this intron with the C4 gene region from the accompanying chromatid, may be important in misalignment of the C4 alleles, and could be involved in gene-conversion-like events between the C4A and C4B alleles. The reported frequencies of C4 Q0 alleles in the general population as determined by protein gel electrophoresis may be inaccurate. These frequencies could possibly be lower than previously thought, if those haplotypes which have two of the same active C4 isotypes on
one chromatid (isoexpression and homoexpression) are at all common. This type of $C4$ gene organisation could not be identified by $C4$ protein allotyping, and therefore emphasises the need for a $C4$ molecular typing method. $C4$ protein allotyping has, however, shown the frequency of $C4\ QO$ alleles to be higher in disease-risked individuals, and therefore reinforces the importance of the association of $C4$ null alleles with disease.

Polymorphism of $C4$ alleles makes them valuable genetic markers for population and evolutionary studies. A molecular-based $C4$ allotyping protocol would be useful in studying population origin and migration by recording the distribution of polymorphic $C4$ alleles. This is made possible by the fact that $C4$ alleles occur with different frequencies and segregate with specific complotypes in specific ethnic groups (Schendel et al., 1984; Partanen et al., 1987; Ayed and Gorgi, 1990).

At the present time, although analysis of $C4$ allelic distribution as determined by protein gel electrophoresis withstands the test of fit to the Hardy-Weinberg equilibrium in the majority of cases, molecular analysis of $C4$ alleles indicates that the serological method may not always provide a true description of $C4$ alleles nor necessarily identify all of the $C4$ alleles carried by a donor. A technique using optical density analysis (Christiansen et al., 1983) may be used to rectify the problem of overlapping $C4$ alleles by protein gel electrophoresis. However, $C4$ alleles with identical migration patterns will not be easily identified and therefore can lead to a false indication of the distribution of these alleles in a population.

$C4$ allele frequency studies of Aboriginal Australians and Papuans from Papua New Guinea may be useful in resolving unanswered questions with respect to their relationships and origin. However, study of isolated populations has previously been hindered by the difficulty in obtaining non-degraded samples of $C4$ protein due to the unavailability of suitable storage facilities. A molecular-based protocol which accurately defines $C4$ allotypes would not only be
valuable in disease association studies, but would enable typing of DNA, a more stable source of material for study of remote populations.

There are a number of sound reasons for analysing the \textit{C4} gene region, both from an evolutionary perspective in relation to population and gene origins, and from a disease association perspective in relation to the aetiology of different disease states. The wealth of information available on studies of this gene region in humans and animals, especially with recently developed molecular technologies, provides a sound basis for research conducted here.
1.13 Hypothesis of studies for this thesis

1. To test whether indigenous people of PNG and Aboriginal Australians have shared a common gene pool by investigation of the distribution of polymorphic \( C4 \) alleles.

2. To test whether the high frequency of \( C4 \) null alleles reported for the northern located Darwin Aboriginal population was an isolated incident or whether other northern located Aboriginal Australian population's have a high frequency of \( C4 \) null alleles.

3. To test whether the polymorphic \( C4d \) region of \( C4 \) genes may be used by molecular means to identify and type \( C4 \) alleles.

4. To test if the genetic basis of Aboriginal Australian \( C4 \) null alleles is the same as that reported for Caucasian null alleles.
Chapter 2
Materials and Methods

2.1 Materials

Reagents used in experimental work, and their place of origin are listed below in alphabetical order of manufacturer:

**Ajax chemicals, Auburn, N.S.W, Australia:** acetic acid; barbituric acid; boric acid; bromophenol blue; chloroform; EDTA; ethanol; HCl; iso-amyl alcohol; iso-propyl alcohol; KCl; methanol; MgCl₂; Na barbiturate; NaCl; NaHCO₃; NaH₂PO₄·2H₂O; Na₂HPO₄; NaHPO₄; Na₂EDTA; Na₄EDTA; CH₃COONa; xylene cyanol

**Aldrich Chemicals, Castle Hill, N.S.W, Australia:** tetramethyl-NH₄Cl

**Amersham Int., Amersham, Buckinghamshire, UK:** mp8 ssDNA standard; Nick translation kit containing reaction reagents dCTP, dGTP, dTTP and DNA polymerase I

**Amicon Inc, MA, USA:** microcon 100 columns

**Applied Biosystems, Foster City, CA, USA:** deoxy/dideoxynucleotide triphosphate (d/ddNTP) mixes; dye-labelled oligonucleotide primers

**BDH Chemicals, Poole, Dorset, UK:** 2-mercaptoethanol; NH₄OAc; polyethylene glycol₆₀₀₀ (PEG₆₀₀₀); sodium citrate; ultra pure sodium dodecyl sulphate (SDS)

**Boehringer Mannheim, Castle Hill, NSW, Australia:** alkaline phosphatase; EcoRI-, HaeIII-, MspI-, Smal- restriction endonucleases; X-gal

**Bresatec Ltd, Adelaide, SA, Australia:** (α-³²P)dATP, 3000 Ci/mmol; (γ-³²P)ATP, 4000 Ci/mmol; PvuII restriction endonuclease

**CUNO laboratory products, NM, USA:** Zetabind nylon membrane

**DIFCO Laboratories, Detroit, MI, USA:** bacto agar; bacto tryptone; yeast extract

**Fluka, Chemische Fabrik, Buchs, Switzerland:** diethyl pyrocarbonate (DEP); formamide; guanidinium thiocyanate

**FMC Corporation, Rockland, ME, USA:** Seakem ME agarose; Gel Bond

**Fuji Photo Film Co., Tokyo, Japan:** X-ray film

**Gambro AB, Lund, Sweden:** Lundia IC 3¹ plate dialyzer cellophane
Hoefer Scientific Instruments, TKO, USA: Hoechst fluorimeter dye
M&B (Mary and Baker) Laboratory Reagent, Australia Pty. Ltd, Victoria: \( \text{Na}_2\text{CO}_3 \)
Mallinkrodt Inc, Kentucky, USA: \( \text{CaCl}_2 \); dimethyl sulfoxide (DMSO); glycerol
Merck, Darmstadt, Frankfurter Strasse, FRG: \( \text{NaOH}; \text{Na(BH}_4 \) 
National Diagnostics, Atlanta, Georgia, USA: polyacrylamide 'Accugel' 40™
Pharmacia, Uppsala, Sweden: ATP; dextran sulphate; DNA polymerase I large fragment (Klenow); dNTP's; Ficoll 400; Moloney murine leukemia virus reverse transcriptase; \( M13 \) mp18 and mp19 vector DNA; plasmid pBR322; T4 polynucleotide kinase; T4 DNA ligase; terminal deoxynucleotide transferase
Polaroid Co., Cambridge, MA, USA: polaroid film
Progen, Queensland, Australia: SSP-I bacteriophage, DNA-molecular weight marker restricted with Eco RI; \( \lambda \) DNA restricted with Hind III and Eco RI
Promega, Madison, WI, USA: Not I Poly 'T' primer; sequencing grade Thermus aquaticus (Taq) DNA polymerase; Taq I-restriction endonuclease
Sigma Chemical Co., St Louis, MO, USA: agarose Type II; \( \text{AgNO}_3 \); bovine serum albumin (BSA); Carboxypeptidase B (Type I); coomassie brilliant blue dye; deoxyribonuclease; dithiothreitol (DTT); dialysis tubing; ethidium bromide; formaldehyde; gelatin; glycine; guanidinium hydrochloride; IPTG; mineral oil; MOPS; neuraminidase (Type VI); polyvinyl pyrroldione (PVP); proteinase K; salmon sperm DNA; sodium sarcosyl; tris
Silenus Laboratories, Hawthorn, Australia: Anti-human C4 antiserum
Stratagene, La Jolla, USA: \( \text{Pfu Taq} \) DNA polymerase
Whatman International Ltd, Molestone, England: Whatman #1 filter paper; Whatman 3MM chromatography paper
Wako Pure Chemical Ind., Novachem, Switzerland: phenol
2.2 Methods

The methods used in this thesis are listed below. Those methods, such as the polymerase chain reaction (PCR), which have specific conditions are outlined in this chapter and described in detail in each relevant chapter.

2.3 Study populations

Aboriginal Australians from two Cape York communities, Groote Island and Darwin were selected for analysis of the polymorphic \( C_4 \) gene locus in population studies and development of a molecular based \( C_4 \) typing technique, as well as for analysis of the molecular basis of \( C_4 \) null alleles. Papuans from three different geographic regions of PNG were chosen for analysis of the polymorphic \( C_4 \) gene locus in population studies. Aboriginal Australians and Papuans were chosen as little is known about the distribution of \( C_4 \) alleles in these indigenous people. Individuals selected for analysis were chosen at random and are believed to be representative of the respective populations. Peripheral blood specimens from 101 Aboriginal donors in two Cape York communities were obtained by Dr Anthony Veale (National Center for Epidemiology and Population Health, ACT). Immediately after collection samples were placed on dry ice for transport to Australia, and then stored at -70 °C. PNG blood specimens provided by Dr Kuldeep Bhatia (PNG Institute of Medical Research, Goroka), were collected from 49 mothers attending the Goroka Base Hospital for routine antenatal check-ups, and from 51 patients participating in a population genetic survey of the Wosera Subdistrict. Samples (120), were collected from Koki, Port Moresby. Plasma samples were separated from buffy coats within one to two hours of collection, treated with 0.2 M EDTA to prevent protein deterioration, and snap frozen in liquid nitrogen. Samples were transported on dry ice to the John Curtin School of Medical Research (JCSMR), Canberra. Specimens from a Groote Island Aboriginal family were provided by Drs Bart Currie and Tim Burt (Menzies School of Health Research). Three Darwin Aboriginal cell line samples H11, H14 and H19, known to be homozygous for \( C4A \) or \( C4B \) null alleles as ascertained by electrophoresis and immunofixation, and Groote
Island Australian Aboriginal peripheral blood samples, were provided by Bart Currie, collected by Tim Burt, and were used for DNA analysis.

2.4 Caucasian donors

Staff members of JCSMR, ANU, Canberra, provided blood samples for use as controls in C4 protein allotyping and in the development of the C4 molecular typing technique. Additional samples of interest were ascertained from the records of the Human Genetics Group (JCSMR). These included certain patients with insulin-dependent diabetes mellitus (IDDM), some of which carried C4A or C4B null alleles, as determined by routine C4 allotyping by electrophoresis and immunofixation. IDDM patient cell lines were provided by the Royal Childrens Hospital (RCH) in Melbourne, and Woden Valley Hospital (denoted with ASD), Canberra. Both RCH and ASD groups were used in the development of the molecular C4 typing protocol. These groups were used as they had rare C4 alleles which were not found in the Aboriginal Australians and provided Caucasian samples of different C4 alleles for comparison with the Aboriginal Australian C4 alleles. 10th International Histocompatibility Workshop cell lines, including EW-VAVY, EW-DUCAF and EW-L0081785, of known homozygosity for either C4A or C4B null alleles, were used for C4 null allele analysis and for development of the molecular C4 typing technique. A series of 50 SLE Caucasian patients numbered SLE502 to SLE686 were from Canberra and were provided by Paul Gatenby. SLE Caucasian patients were chosen as many of these patients had C4B Q0 alleles, as determined by protein gel electrophoresis, compared to other Caucasian donors which had been C4 allotyped in our laboratory. These patients were used in screening for a novel C4A *CAN1 allele which was identified in these studies and shown to be in association with a C4B null allele by protein gel electrophoresis in Aboriginal Australians.
2.5 *C4* protein allotyping

*C4* allotyping was performed by a protein electrophoretic migration technique described by Awdeh and Alper (1980). Unusual *C4* allotypes including *C4* null alleles were confirmed by treatment of serum samples with *carboxypeptidase B* prior to electrophoresis, as described by Sim and Cross (1986).

2.5.1 Gel preparation

The agarose gel solution consisted of 0.75 % SEAKEM ME agarose, 16 ml of electrophoresis buffer (0.2 M Tris, 0.4 M Glycine, 30 mM Na Barbiturate, 6 mM Barbituric acid, pH approximately 8.8), 1.6 ml of 0.2 M EDTA solution (0.1 M Na2EDTA, 0.1 M Na4EDTA), and distilled water to a total volume of 40 ml. The gel solution was boiled and placed in a gel mould prepared in the following way. The hydrophobic side of a piece of 'Gel Bond' (24.3 cm by 12.5 cm), was placed on a glass plate of identical size. A mould was formed by using a second siliconized glass plate and two 1 mm gaskets to separate the glass plates, and was held together using bulldog clips. Gels were cooled to 4 °C prior to use. EDTA solution was stored at -20 °C, and once thawed was discarded after one week.

2.5.2 Dialysis

Dialysis buffer stock solution (pH approximately 6.8), consisted of 0.3 M NaH2PO4·2H2O, 0.2 M Na2HPO4, and 0.03 M Na4EDTA. This solution was discarded after one week. 1 L of diluted dialysis buffer (1 part stock solution to 5 parts distilled water) was used for each dialysis. Cellophane membrane was placed on towelling (previously soaked in dialysis buffer), extending across a wick with one end suspended in dialysis buffer and the other in an empty container. This arrangement allowed the movement of buffer from one container to the next. Serum samples were dialysed at 4 °C overnight by mixing 10 μl serum sample with 2μl neuraminidase (50 units dissolved in 1.4 ml of 0.2 M EDTA solution and stored at -20 °C) on the cellophane.
2.5.3 Electrophoresis

An applicator strip, the length of the glass plate with slots 8-10 mm, 0.75 mm wide and 2 mm apart, was made from gel bond. This applicator strip was placed (hydrophobic side up) along the cathode side of the gel where excess water had been removed by blotting with Whatman No.1 filter paper. Dialysed samples (10 μl), were loaded into each slot. A haemoglobin sample containing HbA was co-electrophoresed as a marker. Gels were left to stand for 30 min prior to electrophoresis to allow the movement of serum into the gels. 1 L of undiluted electrophoresis buffer was used in each gel tank. Saturated wicks, made from cotton wool, were extended from the cathode and anode ends of the gel into the electrophoresis buffer. Gels were covered with a glass plate to prevent dehydration, and electrophoresed at 85 mA for 6 hours at 6 °C.

2.5.4 Immunoprecipitation and staining

Gels were overlaid with diluted anti-human C4 antiserum (0.4 ml antiserum and 2 ml saline), and incubated for 1 hour at room temperature in a moist chamber to prevent desiccation. Gels were washed gently with running water, then water was removed by blotting the gel with absorbent paper for 10 min. Gels were soaked overnight in 0.9 % saline. Gels were washed again, blotted, air dried and stained in 0.2 % Coomassie Brilliant Blue in 9 parts methanol, 2 parts acetic acid, and 9 parts distilled water, then destained in the solvent.

2.5.5 Carboxypeptidase B treatment

Unusual C4 allotypes and C4 null alleles were confirmed by treatment with Carboxypeptidase B, where 1.6 units of Carboxypeptidase B in 100 mM NaCl was added to 10 μl of plasma, incubated at room temperature for 30 min, then dialysed as described in section 2.5.2.
2.6 Statistical analysis

Maximum likelihood estimates of \( C4 \) allele frequencies were calculated from phenotype distributions using an iterative computer programme to determine the best fit to Hardy-Weinberg equilibrium. That is \( p^2 + q^2 + 2pq = 1 \), where \( p \) = frequency gene 1, and \( q \) = frequency gene 2. This computer analysis was necessary to estimate the population frequency of double heterozygosity for null alleles which could not be detected by \( C4 \) protein allotyping.

2.7 Genomic DNA extraction

Heparinised whole blood samples were centrifuged at 2000 rpm for 10 min at room temperature, prior to removal of the plasma and red blood cell interphase (buffy coat), with siliconised glass pipettes. Buffy coats were washed multiple times with \( T_{20}E_{5} \) (20 mM Tris/HCl, 5 mM EDTA, pH 8.0), and centrifuged at 4000 rpm for 10 min at room temperature, until the remaining cell pellet appeared white. Cell line material was washed once with this solution. Genomic DNA was extracted by resuspending pellets in 3 ml \( T_{10}E_{1} \) (10 mM Tris/HCl, 1 mM EDTA, pH 8.0), 400 \( \mu \)l of 10 % SDS, and 1.2 mg of Proteinase K, followed by incubation at 37 °C overnight. Cell debris was precipitated by the addition of 1 ml saturated NaOAc, mixed vigorously for 15 sec, and centrifuged at 1500 rpm for 15 min. DNA in the supernatant, was precipitated in a fresh tube by the addition of an equal volume of iso-propyl alcohol. DNA was removed with siliconised looped pipettes, air-dried and resuspended in \( T_{10}E_{1} \) to a final concentration of 0.3 \( \mu \)g/\( \mu \)l.

2.8 DNA digestion

Digestion of DNA samples was carried out according to procedures described by the manufacturer of each enzyme. Core buffers were provided by the manufacturer at ten times the concentration required for digestion. Samples were digested at 65 °C for two hours, or overnight at 37 °C, depending on specific enzyme used. The
digestion reaction was stopped by the addition of 0.5 M EDTA (pH 8.0), and heating to 65 °C for 10 min.

2.9 Gel electrophoresis

A variety of agarose (Type II, Sigma), and polyacrylamide gel concentrations were used depending on the size of DNA fragments to be separated. In each case a 6 x loading buffer (30 % (v/v) glycerol, 0.25 % (w/v) xylene cyanol, and/or 0.25 % (w/v) bromophenol blue) was added to 1 x final concentration, samples were heated to 65 °C for 10 min and cooled on ice for 10 min prior to loading. Agarose gels were electrophoresed at room temperature, with constant voltage, in 1 x Tris-acetate EDTA (TAE) buffer (0.04 M Tris-acetate, 1 mM EDTA, pH 8.0), whilst polyacrylamide gels were electrophoresed at 6 °C with constant current and 1 x Tris-borate EDTA (TBE) buffer (0.045 M Tris-borate, 1 mM EDTA, pH 8.0). Standard DNA molecular weight markers (~0.5 μg), listed in Figure 2.1, were used to estimate DNA fragment sizes.

2.10 Gel staining with EtBr and silver

Agarose gels were prepared containing EtBr at a final concentration of 0.5 μg/ml, whereas polyacrylamide gels were stained with EtBr (0.5 μg/ml) in distilled water, after electrophoresis for 15 min. Gels were photographed with UV transmitted light (254 nm) using an orange filter and a Polaroid camera. Gels which contained DNA fragments not easily interpreted by EtBr staining were stained with silver. This method of staining was based on the protocol supplied with Qiagen TGGE apparatus (Diagen GmbH, Dusseldorf, FRG), and adapted from Merril et al. (1981). Each gel was fixed by two 3 min washes with 9.95 % ethanol and 0.5 % acetic acid. Silver adhered to the DNA by the addition of 0.1 % AgNO₃ for 10 min, followed by two rinses with water. Gels were developed for approximately 20 min with 1.5 % NaOH, 0.01 % NaBH₄, and 0.1 % formaldehyde, and then fixed for 10 min in 0.75 % Na₂CO₃. Silver stained gels were photographed with a Polaroid camera using a light box and a violet coloured filter.
Figure 2.1 DNA molecular weight standards
<table>
<thead>
<tr>
<th>( \lambda, \text{HindIII} ) (Kb)</th>
<th>( \lambda, \text{Hind III} ) and ( \text{Hind III/EcoRI} ) (Kb)</th>
<th>( \lambda, \text{Hind III} ) and ( \text{EcoRI} ), (Kb)</th>
<th>( \text{SSP-1} ) bacteriophage, ( \text{EcoRI} ), (Kb)</th>
<th>( \text{pBR322, Msp I} ) (bp)</th>
<th>( \text{pBR322, Hae III} ) (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>23.130</td>
<td>23.130</td>
<td>21.226</td>
<td>7.840</td>
<td>622</td>
<td>587</td>
</tr>
<tr>
<td>6.557</td>
<td>9.416</td>
<td>4.973</td>
<td>5.860</td>
<td>404</td>
<td>504</td>
</tr>
<tr>
<td>4.361</td>
<td>2.322</td>
<td>4.268</td>
<td>3.370</td>
<td>309</td>
<td>458</td>
</tr>
<tr>
<td>4.268</td>
<td>1.904</td>
<td>1.175</td>
<td>2.680</td>
<td>242</td>
<td>434</td>
</tr>
<tr>
<td>2.322</td>
<td>0.947</td>
<td>0.947</td>
<td>1.890</td>
<td>238</td>
<td>267</td>
</tr>
<tr>
<td>2.027</td>
<td>0.831</td>
<td>1.450</td>
<td>1.800</td>
<td>217</td>
<td>234</td>
</tr>
<tr>
<td>1.904</td>
<td>0.564</td>
<td>1.330</td>
<td>1.450</td>
<td>201</td>
<td>213</td>
</tr>
<tr>
<td>1.584</td>
<td>1.090</td>
<td>1.090</td>
<td>1.330</td>
<td>190</td>
<td>192</td>
</tr>
<tr>
<td>1.375</td>
<td>0.880</td>
<td>0.880</td>
<td>1.090</td>
<td>180</td>
<td>184</td>
</tr>
<tr>
<td>0.947</td>
<td>0.660</td>
<td>0.660</td>
<td>0.880</td>
<td>160</td>
<td>124</td>
</tr>
<tr>
<td>0.831</td>
<td>0.480</td>
<td>0.480</td>
<td>0.660</td>
<td>147</td>
<td>123</td>
</tr>
<tr>
<td>0.564</td>
<td>0.380</td>
<td>0.380</td>
<td>0.480</td>
<td>110</td>
<td>104</td>
</tr>
<tr>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.380</td>
<td>90</td>
<td>76</td>
</tr>
<tr>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.380</td>
<td>76</td>
<td>89</td>
</tr>
<tr>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.380</td>
<td>67</td>
<td>80</td>
</tr>
<tr>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.380</td>
<td>34</td>
<td>64</td>
</tr>
<tr>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.380</td>
<td>26</td>
<td>57</td>
</tr>
<tr>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.380</td>
<td>15</td>
<td>51</td>
</tr>
<tr>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.380</td>
<td>5</td>
<td>21</td>
</tr>
<tr>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.380</td>
<td>18</td>
<td>11</td>
</tr>
<tr>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.380</td>
<td>18</td>
<td>8</td>
</tr>
</tbody>
</table>
2.11 Southern Blotting

DNA fragments were transferred from a gel to nylon membrane (Zetabind), using the general method described by Maniatis et al. (1982). Gels were soaked in 0.25 M HCl for 10 min to depurinate the DNA prior to capillary blotting using 0.4 M NaOH as a transfer buffer (Reed and Mann, 1985). After transfer was complete membranes were neutralised by rinsing in 1.5 M NaCl, 0.5 M Tris (pH 7.5) for 5 min, and then DNA was cross-linked to the membrane with 254 nm UV light for 5 min and dried at 65 °C for one hour.

2.12 Radioactive labelling of DNA probes and hybridisation

2.12.1 Nick translation of C4 pAT-A

The C4 pAT-A cDNA probe containing an entire plasmid vector 'pAT-A', and a C4 specific insert, was provided by Dr M. Carroll (The Childrens Hospital, Boston, USA). This probe (0.5 µg) was radio-labelled using an Amersham nick translation kit following the procedure described by Maniatis et al. (1982). The following reagents were combined: 1 µl (0.5 µg) of C4 pAT-A; 2µl of each dNTP 'C', 'G', and 'T'; 3 µl DNA polymerase I enzyme; 3 µl of α-32P dATP; and 17 µl of distilled water. The reaction was carried out at <15 °C for 1 hour, and then stopped by the addition of 1 µl 0.5 M EDTA (pH 8.0). Samples were precipitated with 15 µl of 7.5 M ammonium acetate and 90 µl of absolute ethanol, collected by centrifugation, and resuspended in 50 µl T10E1.

2.12.2 Hybridisation of nick translated probes

Prior to pre-hybridisation, filters were incubated in a 0.1 x SSC, 0.5 % SDS solution for 1 hour at 65 °C to minimise background hybridisation created by the dye which had transferred from the gel to the filter during blotting. Pre-hybridisation and hybridisation were performed in a rotisserie Hybaid oven (Hybaid, Middlesex, United Kingdom) at 65 °C. Pre-hybridisation and hybridisation solution consisted of 150 µl heat denatured salmon-sperm DNA
(10 mg/ml), 7.5 ml Nasmyth solution (18.5 % dextran sulphate, 1.85 % sodium sarcosyl, 0.6 M NaCl, 0.18 M Na₂HPO₄, 6 mM Na₂EDTA) and 7.5 ml distilled water. Following pre-hybridisation of filters overnight, nick translated probes were boiled in 5 ml of hybridisation solution for 10 min and then added to filters for incubation overnight. To remove excess and non-specifically bound probe, filters were washed in 2 x SSC and 0.1 % SDS at room temperature for 5 min, 2 x SSC, 0.1 % SDS at 65 °C for 15 min, 1 x SSC, 0.1 % SDS at 65 °C for 15 min, and 0.5 x SSC, 0.1 % SDS at 65 °C for 15 min. Once the filters had a reading of 10-20 cps, they were exposed to autoradiographic X-ray film, using single sided intensifying screens (0.5 dpm/mm² sensitivity), at -70 °C for 3 to 5 days, after which time films were developed using an automatic film developer.

2.12.3 Labelling 5' end of DNA oligonucleotides

The following protocol was used for the radio-labelling of small DNA fragments of approximately 20 nucleotide bases in length. T4 polynucleotide kinase was used to catalyse the transfer of the terminal gamma phosphate of ATP to the 5'-hydroxyl terminus of the DNA fragment to be labelled. 10 pmoles of oligonucleotide were radio-labelled by the addition of 3 μl of 10 x kinase buffer (50 mM Tris/HCl pH 7.5, 20 mM MgCl₂, 20 mM DTT), 3 μl of γ³²P-ATP, 1 μl of T4 polynucleotide kinase (10 units), and distilled water to a total volume of 30 μl. Samples were incubated at 37 °C for 30 min and then stopped by the addition of 1 μl 0.5 M EDTA pH 8.0.

2.12.4 Hybridisation of 5' end-labelled oligonucleotides

Filters were prepared for hybridisation by washing in 0.1 x SSC, 0.5 % SDS for 1 hour at 65 °C, followed by pre-hybridisation for 2 hours at 42 °C in 15 ml of a solution containing 15 % formamide, 0.1 % Denhardt's solution (0.1 % BSA, 0.1 % PVP, 0.1 % Ficoll 400), 5 x SSPE (5 mM EDTA, 50 mM NaH₂PO₄.H₂O, 900 mM NaCl, pH 7.4), 0.2 mg/ml heat denatured salmon sperm, and 1 % SDS. Radio-labelled oligonucleotides were added to the pre-hybridisation solution and hybridised to filters at 42 °C for 1 hour. Excess and non-specifically
bound probe was removed by washing the filters twice in 3 x SSC at
room temperature, and 30 min at 59 °C in a filtered solution of 3 M
tetramethyl NH₄Cl, 50 mM Tris/HCl pH 8.0, 2 mM EDTA, and 1 % SDS.
Filters were autoradiographed by exposure to X-ray film with an
intensifying screen (0.5 dpm/mm² sensitivity), for 2-24 hours at
-70 °C.

2.13 Total RNA extraction

2.13.1 RNA preparation

Isolation of total RNA was by a single extraction with an acid
guanidinium thiocyanate-phenol-chloroform mixture, as described by
Chomczynski and Sacchi (1987). RNA was extracted from cell line
material (1 x 10⁷ cells) by firstly, denaturing cellular protein in
500 µl of solution D (4 M guanidinium thiocyanate, 25 mM sodium
citrate {pH 7.0}, 0.5 % sodium sarcosine, 0.1 M 2-mercaptoethanol).
Secondly, 50 µl of 2 M diethyl pyrocarbonate (DEP) treated NaOAc
pH 4.0, 500 µl phenol and 100 µl chloroform:iso-amyl alcohol (24:1)
were added, mixed thoroughly and cooled on ice for 15 min. Cell
debris was collected by centrifugation at 10,000 g for 20 min at
4 °C, and the supernatant containing RNA was removed to a fresh
tube. RNA was precipitated by the addition of 500 µl iso-propyl
alcohol and incubation at -20 °C for 1 hour. The RNA precipitate was
collected by centrifugation at 10,000 g for 20 min at 4 °C, and then
resuspended in 150 µl of solution D before a second precipitation
with iso-propyl alcohol. The final precipitate was washed in 20 µl of
75 % ethanol, vacuum dried for 15 min, and then dissolved in 20 µl
DEP-treated distilled water.

2.13.2 Estimation of RNA concentration and protein contamination

A spectrophotometer was used to measure the concentration of RNA
by recording the optical density for 1 µl of RNA in 1 ml water, at 260
nm and 280 nm. The presence of protein contamination in an RNA
sample was detected if the ratio of 260 nm : 280 nm readings was
less than 2.
2.13.3 Electrophoresis of RNA

Total RNA extractions were separated on a 1.4 % agarose gel (Type II, Sigma) containing 36 ml of DEP-treated water, 10 ml of 10 x MOPS buffer (0.4 M MOPS pH 7.0, 0.1 M NaOAc, 0.01 M EDTA, 0.1 % DEP), 2 ml of formaldehyde (37 %), and 5 μl of EtBr (10 mg/ml). This solution was especially formulated to prevent the degradation of RNA. Approximately 10 μg of RNA suspended in T10E1, 2 mM of dithiothreitol (DTT), and 1 unit/μl of RNasin, were prepared for electrophoresis by the addition of 2 μl 10 x MOPS buffer, 3.3 μl formaldehyde solution (37 %) and distilled water to 20 μl. Prior to electrophoresis samples were heated at 55 °C for 15 min, cooled on ice for 5 min, followed by the addition of 2 μl 10 x RNA sample buffer (50 % glycerol, 0.5 mM EDTA {pH 8.0}, 0.25 % {w/v} bromophenol blue). RNA samples were electrophoresed for 5 hours at 65 volts, and then visualised by UV transmitted light.

2.14 Northern blots and hybridisation

RNA was transferred to a nylon membrane (Zetabind), by blotting overnight using 1 x MOPS buffer, following the method described in section 2.11. Pre-hybridisation and hybridisation of filters were carried out in a Hybaid oven, using a high stringency buffer containing 5 x SSPE (5 mM EDTA, 50 mM NaH2PO4.H2O, 900 mM NaCl, pH 7.4), 7 % SDS and 40 % formamide. Filters were pre-hybridised for 2 hours at 42 °C, and then hybridised overnight with C4 pAT-A nick translated probe prepared as described in section 2.12.1. Excess, and non-specifically bound probe was removed by washing filters in 2 x SSC, 0.1 % SDS for 5 min at room temperature, 2 x SSC, 0.1 % SDS for 15 min at room temperature, 0.5 x SSC, 0.1 % SDS for 15 min at room temperature, 0.1 x SSC, 0.1 % SDS for 15 min at room temperature, and 0.1 x SSC and 1 % SDS for 30 min at 42 °C. Filters (~10-20 cps) were exposed to X-ray film with an intensifying screen at -70 °C for 3 to 5 days.
2.15 First strand cDNA synthesis

C4-specific and non-specific first strand cDNA were synthesised from total RNA by using a 3' C4-specific antisense primer 5'-TGGGA CTTGAACCCATTTCCG-3' (corresponding to nucleotide bases encoding amino acid residues #1317 to #1323 in exon 30), and a 3' non-specific antisense primer Not 1, Poly 'T' primer 5'-CAATTCGCGGCCG C(T)15-3', respectively. Both C4-specific and non-specific cDNAs were prepared by incubating 10 μg of total RNA (previously incubated at 65 °C for 3 min, and on ice for 5 min), with 1 x cDNA synthesis buffer \{500 mM Tris/HCl (pH 8.3), 750 mM KCl, 30 mM MgCl₂, 100 mM DTT\}, 50 pmoles of antisense primer, 2 mM of dNTP's, 15 units of RNasin, and 40 units of reverse transcriptase in a total volume of 50 μl, at 37 °C for 4 hours. cDNA was purified by the addition of 150 μl DEP-treated water, 100 μl phenol and 100 μl of chloroform:iso-amyl alcohol (24:1), mixed vigorously, and centrifuged for 3 min at 14,000 g. The aqueous phase was recovered and the cDNA concentrated by using a microcon 100. A sample (1 μl) of the cDNA preparation was electrophoresed on a 0.8 % agarose (Type II, Sigma) gel containing EtBr (0.5 μg/ml), at 120 volts for 30 min, and visualised by UV light.

2.16 PCR amplification

Polymerase chain reaction (PCR) was used to amplify a segment of DNA between two regions of known sequence. Each PCR reaction contained 0.3 μg/μl of template DNA, 10 % DMSO, 0.2 mM dNTP mix (dATP, dCTP, dGTP, dTTP), 1 x Promega Taq buffer containing MgCl₂ (25 mM), 0.3 μg/μl of each oligonucleotide, and 0.25 units of either Promega Taq or Pfu Taq DNA polymerase in a total volume of 100 μl. Each reaction was covered by 70 μl of mineral oil to prevent evaporation. For each PCR cycle template DNA was denatured by heating to 96 °C and then cooled to a temperature between 58 °C and 60 °C, which allowed the oligonucleotides to anneal to their target sequences, followed by primer extension at 72 °C. This cycle of DNA denaturation, annealing and primer extension was repeated for 30-35 cycles. Optimum conditions varied for each PCR reaction depending on the base composition of each oligonucleotide, and the
length of template DNA to be amplified. Annealing temperatures for each PCR were determined by the theoretical melting temperature (Tm) of each oligonucleotide, given by Tm °C = 4 (C + G) + 2 (A + T). PCR products were visualised by electrophoresis at 120 volts for 30 min on a 1 % agarose (Type II, Sigma) gel containing EtBr (0.5 μg/ml).

2.17 Electro-elution of PCR fragments from agarose gels

PCR fragments of interest were visualised by EtBr staining and long wavelength (300-360 nm) UV transmitted light, then excised from the agarose gel with sterile scalpel blades. Agarose containing DNA was placed in 2 cm wide dialysis tubing (held together with dialysis clips), with 500 μl of 0.1 x Tris-acetate buffer. Elution was carried out in a gel electrophoresis tank for 30 min at 290 volts, followed by reversal of the current for 30 sec to prevent binding of DNA to the dialysis tubing. Samples were purified with a phenol/chloroform:iso-amyl alcohol (24:1) extraction, and then precipitated with 1/10 th the total volume 7.5 M NH₄OAc. Samples were air-dried and resuspended in 100 μl distilled water.

2.18 Cloning of C4 specific DNA fragments

2.18.1 Sticky-end cloning

2.18.1.1 Preparation of PCR DNA

Total PCR reactions (100 μl) were concentrated using a centricon 100 to a volume of approximately 35 μl. PCR samples (35 μl) were restricted at 37 °C for 4 hours with Eco RI (50 units) and Sal I (20 units), using 1 x reaction buffer 'H' (high salt buffer provided by the manufacturer), in a total volume of 100 μl. Restriction reactions were stopped by the addition of 2 μl of 0.5 M EDTA, followed by ethanol precipitation with 1/10 th the total volume of 7.5 M NH₄OAc. DNA was pelleted and rinsed in 70 % ethanol. Digested PCR samples were air-dried and resuspended in 35 μl of distilled water.
2.18.1.2 Preparation of M13 cloning vectors

M13 vectors (2 μg of each) mp18 (for sequencing the sense strand) and mp19 (for sequencing the anti-sense strand), were restricted with Eco RI and Sal I using the same reaction conditions, as described for the preparation of PCR DNA, and resuspended in 60 μl of sterile distilled water. Restricted vector fragments were de-phosphorylated to prevent re-ligation by the addition of 1 μl of 1 M Tris (pH 8.0), and 1 μl of alkaline phosphatase, in a total volume of 100 μl, followed by incubation for 30 min at 37 °C. Vector samples were then phenol extracted and ethanol precipitated (as described above), air-dried, and resuspended in 100 μl of T10E1 to a final concentration of 20 ng/μl.

2.18.2 Blunt-end cloning

2.18.2.1 Preparation of PCR DNA

Total PCR reactions (100 μl) were purified by phenol/chloroform:iso-amyl alcohol (24:1) extraction, followed by ethanol precipitation using 1/10 th volume of 7.5 M NH₄OAc, and rinsing with 70 % ethanol. PCR samples were air-dried and resuspended in 13 μl of distilled water. PCR fragments were phosphorylated by incubating for 45 min at 37 °C in a 20 μl reaction containing 50 mM Tris/HCl (pH 7.5), 20 mM MgCl₂, 20 mM DTT, 2 mM ATP and 10 units of T4 polynucleotide kinase. Phosphorylated PCR-DNA was then purified by electro-elution as described in section 2.17.

2.18.2.2 Preparation of M13 cloning vectors

Vector M13 was linearised by digestion at 30 °C for 2 hours in a reaction containing 2 μg of M13 (mp18 or mp19), 1 x One-Phor-All reaction buffer (Pharmacia), and 20 units of restriction enzyme Sma I. To minimise recircularisation, vector DNA was dephosphorylated with 1 μl Tris (pH 8.4) and 1 μl alkaline phosphatase, in a total volume of 100 μl at 37 °C for 30 min. Vector preparation was purified by phenol/chloroform:iso-amyl alcohol (24:1) extraction, and precipitated with 1/10th volume 7.5 M NH₄OAc and 2 volumes
absolute ethanol, then rinsed once with 80 % ethanol and once with 95 % ethanol. Vector samples were air-dried and resuspended in 100 μl of T\textsubscript{10}E\textsubscript{1}.

2.18.3 Ligation of PCR insert and vector DNA

Both sticky-end and blunt-end ligation reactions were carried out using a molar ratio of 3:1 of PCR insert to vector. This ratio was determined by visualisation of electrophoresed vector and insert PCR DNA. Each ligation reaction contained PCR insert, vector DNA, 1 x BRL ligation buffer (50 mM Tris/HCl (pH 7.6), 10 mM MgCl\textsubscript{2}, 5 % polyethylene glycol (PEG\textsubscript{6000}), 1 mM ATP, 1 mM DTT), and 1 unit of T\textsubscript{4} DNA ligase in a total volume of 10 μl. Sticky-end cloning reactions were incubated at 15 °C overnight, whereas blunt-end cloning reactions were left at room temperature overnight.

2.18.4 Preparation of competent \textit{E.Coli} cells

Strain TG-1 \textit{E.Coli} competent cells were prepared on the same day as transformation by diluting (1 in 100) an overnight culture of cells in 20 mls of 2YT media (1.6 % bacto tryptone, 1 % yeast extract, 0.5 % NaCl), followed by incubation with shaking at 37 °C for 2 hours. Once the cells had reached exponential growth phase, they were harvested by centrifugation at 2000 rpm, for 7 min at 4 °C. The supernatant was removed, and the cells were gently resuspended in 10 ml of cold 50 mM CaCl\textsubscript{2}, and left on ice for 30 min. Cells were harvested again by centrifugation at 2000 rpm for 7 min, at 4 °C, and then gently resuspended in 2 ml of 50 mM CaCl\textsubscript{2}, and incubated on ice for 30 min prior to transformation.

2.18.5 Transformation of competent cells

To transform the competent cells 10 μl of the ligation mixture was incubated on ice for 45 min with 200 μl of competent cells, and then heat-shocked for 3 min at 42 °C, and left on ice for 10 min. Transformed cells were combined with 200 μl of exponentially grown TG-1 cells, 10 μl of 100 mM IPTG, 35 μl of 20 μg/ml X-gal, and 3 ml
of 48 °C H-Top agar (1 % bacto tryptone, 0.5 % yeast extract, 0.8 % NaCl, 0.8 % bacto agar), and then spread onto H-plates (1 % bacto tryptone, 0.8 % NaCl, 1.2 % bacto agar). After the H-top mixture had set, plates were inverted and incubated overnight at 37 °C.

2.18.6 Selection of recombinant plaques

Cells containing M13 recombinants appeared as colourless plaques whereas cells infected with M13 vector appeared as blue plaques. For PCR DNA inserts representing a gene in heterozygous form, a minimum of 8 plaques were recovered from transformation, whereas for those representing a gene in the homozygous form, a minimum of 4 plaques were chosen. Plaques were extracted using sterile glass pipettes, resuspended in 2 ml of 2 x YT media, with 10 μl of exponentially grown TG-1 cells, and incubated overnight at 37 °C with shaking. Supernatant containing M13 recombinants were harvested by taking 1 ml of culture and centrifuging for 10 min, and then stored at 4 °C.

2.18.7 Direct Gel Electrophoresis (DIGE), and hybridisation selection of recombinant clones

To determine whether the selected M13 recombinants contained a C4 insert, 20 μl of phage M13 supernatant was combined with 5 μl of DIGE dye (3 % SDS and 0.25 % bromophenol blue in 30 % glycerol), heated to 65 °C for 10 min, placed on ice for 10 min, and then separated by gel electrophoresis on a 0.7 % agarose gel (Type II, Sigma) as described in section 2.9, for 4 hours at 60 volts. Samples containing an insert could be identified by differential size to a control M13 sample with no insert (blue plaque). Gels were blotted by capillary transfer as described in section 2.11, probed with nick-translated C4 pAT-A (section 2.12.1), and hybridised as in section 2.12.2 to confirm that samples contained a C4 specific insert.
2.19 Automated DNA sequencing of recombinant clones

2.19.1 Preparation of single stranded DNA template

Single stranded DNA (ssDNA) was prepared from positive recombinant clones by combining 1 ml of M13 supernatant with 250 µl of PEG (20 % PEG₆₀₀₀, 2.5 M NaCl), and incubated at room temperature for 30 min. To collect the PEG precipitate containing ssDNA, samples were centrifuged for 10 min, the supernatant was discarded, and the precipitate was resuspended in 100 µl T_10E₁. ssDNA was purified with a phenol/chloroform:iso-amyl alcohol (24:1) extraction, followed by ethanol precipitation with 7.5 M NH₄OAc. Precipitates were washed with 80 % cold ethanol, air-dried, and resuspended in 20 µl of distilled water.

2.19.2 Estimation of concentration of ssDNA template

The concentration of ssDNA was determined by a DNA fluorometer (Hoefer Scientific Instruments, TKO 100, U.S.A), and a fluorescent dye called 'Hoechst dye' (0.1 µg/ml) which was provided by Hoefer Scientific Instruments. A 100 ng/µl ssDNA standard 'mp8', was used to compare with prepared ssDNA template. Readings ranged from 70 ng/µl to 200 ng/µl.

2.19.3 Single cycle sequencing

Single cycle sequencing was performed using an automated sequencing kit from ABI (Applied Biosystems) which contained dATP, dCTP, dGTP, dTTP, M13 dye primers for each nucleotide sequencing reaction, and Taq DNA polymerase (0.25 units). A total of 1600 ng ssDNA template was required for sequencing of cloned material. Template and reagents were divided into four reaction mixes containing 200 ng of ssDNA template, 1.8 µl of 5 x sequencing buffer (50 mM Tris/HC1 {pH 8.4}, 50 mM MgCl₂, 250 mM NaCl), and 0.4 pmoles of dye primer ('Joe' for 'A' mix and 'Fam' for 'C' mix), in the 'A' and 'C' reaction mixes in a total volume of 6 µl, and double these amounts for all reagents in the respective 'G' and 'T' reaction
mixes (using 'Tamra' for 'G' and 'Rox' for 'T' reaction mixes), for a total volume of 12 μl. Dye primers annealed to template by heating each reaction mix to 65 °C for 10 min, and cooling to room temperature for 20 min. Prior to primer extension 1 μl of the respective dNTP for 'A' and 'C' reactions, and Taq polymerase diluted to 0.5 units in 50 mM KCl, 10 mM Tris (pH 8.3), 1.5 mM MgCl₂, and 0.01 % gelatin (w/v), were added to each reaction. Quantities were doubled for 'G' and 'T' reactions. Extension of primers with Taq polymerase was performed by incubating samples at 70 °C for 10 min, followed by 37 °C for 10 min. Interpretation of the first 50 to 100 base pairs is often difficult due to 5' to 3' exonuclease activity of DNA Taq polymerase. To rectify this problem a second reaction was performed known as 'Klenow Chase'. The Klenow enzyme (large fragment of *E.coli* DNA polymerase I), was diluted to 0.7 units/μl in 1 x sequencing buffer prior to use. For 'A' and 'C' sequencing reactions 0.7 units of enzyme were used, whereas 1.4 units were used in the 'G' and 'T' reactions. Samples were incubated at 37 °C for 5 min, and then immediately placed on ice. Contents of 'A', 'C', 'G', and 'T' reactions for each sample were precipitated together by adding sequentially to 130 μl of 95 % ethanol and 5 μl of 3 M NaOAc (pH5.2), followed by incubation on ice for at least 10 min. Precipitates were collected by centrifugation on ice for 30 min, and rinsed with 100 μl of 70 % ethanol, followed by a second centrifugation for 5 min. DNA pellets were air-dried for 7 min, and stored at -20 °C in the dark until electrophoresis. Samples were electrophoresed by the Biomolecular Resources Facility at JCSMR, Canberra, using an ABI (Applied Biosystems) version 1.30 software package to perform automatic fluorescence analysis and base calling.

2.20 Computer analysis of derived C₄ protein sequences to estimate their isoelectric point, denoted as 'PI' (Predicted Isoelectric Point)

A Sequence Analysis Software package, version 7.2, called 'isoelectric' was provided by the Genetics Computer Group, Madison, Wisconsin, USA. This programme was used to estimate the PI of a given protein sequence by calculating the sum of the number of positively charged residues (protonated lysine, arginine, histidine), minus the number of negatively charged residues (deprotonated tyrosine, cysteine,
glutamate, aspartate), plus the number of protonated amino termini, minus the number of deprotonated carboxyl termini.

For each amino acid residue of interest, the protein analysis programme calculated the number of protonated residues by the following equation: 

\[ N_p = N_t \frac{[H^+]}{([H^+] + K_N)} \]

where \( N_p \) = number of protonated residues, \( N_t \) = total number of residues of a specific amino acid, \([H^+]\) = hydrogen ion concentration, and \( K_N \) = dissociation constant for the amino acid of interest that is equal to the following: 

\[ 10^{-pK_N} \]
Chapter 3
Investigation of complement C4 alleles in Aboriginal Australian and Papua New Guinea (PNG) populations

3.1 Introduction

Aboriginal Australians have been isolated from other indigenous people since the post-Pleistocene elevation in sea level which separated Australia from New Guinea. Since this time these populations have genetically differentiated as a result of random genetic drift, differential selection, new mutations, and the introduction of different genes as a result of occasional contact with other ethnic populations (Kirk, 1972).

Investigation into the origin of Aboriginal Australians has included populations from the Cape York Peninsula (north-eastern Australia), because of their geographical proximity to other indigenous people, such as those from PNG. These populations are thought to be more likely to show evidence of gene flow into Australia compared with centrally and southern located populations. Linguistic (Simmons et al., 1958), skeletal (Macintosh and Larnach, 1972), and serum protein studies (Kirk, 1972), provide evidence which suggests that Aboriginal Australians from the Cape York region are similar to Melanesian and PNG populations. Comparative studies of the genetic make-up of each ethnic group would provide further evidence for any relationship.

Comparison of the C2 and Bf loci in Aboriginal Australian and PNG highlander populations found each population to be monomorphic, encoding C2 I and Bf S (Ranford, 1989). These studies and those by Kirk (1972), suggest that Aboriginal Australians and PNG populations (especially highland populations) may have a common genetic origin. This observation has prompted population genetic analysis of polymorphic gene regions between indigenous people of these countries. Hence the polymorphic C4 gene region was considered to be an ideal gene region for this analysis.
There is a lack of information on \( C4 \) allotyping in the Southwest Pacific. Only two other studies have investigated the distribution of \( C4 \) alleles in this region, one, involving Aboriginal Australian populations from central and northern Australia (Ranford et al., 1987), and the other, involving isolated populations of PNG (Bhatia et al., 1991). \( C4 \) proteins are extremely labile, and therefore deteriorate rapidly unless stored below -70 °C soon after collection. Suitable storage facilities are difficult to obtain in remote areas of PNG and Australia, accounting for the lack of \( C4 \) allotyping data in these regions.

\( C4 \) allotype analysis of two populations of Aboriginal Australians by Ranford (1989), found them to be different from both Asians and Pacific Islanders, as well as from each other. Ranford et al. (1987), found the highest frequency ever reported of \( C4 \) null alleles in a northern Aboriginal Australian population at both the \( C4A \) and \( C4B \) loci (29 % and 26 %, respectively). Also, each Aboriginal Australian population had two more frequently occurring \( C4A \) alleles, 'A 3' and 'A 4', whereas other populations of the Southwest Pacific usually had one frequently occurring \( C4A \) allele (summarised by Ranford, 1989). At the \( C4B \) gene locus of Aborigines, 'B 1' and 'B 2' were found to be the most frequently occurring alleles, whereas 'B 1' was usually the more frequently occurring \( C4B \) allele in other Southwest Pacific populations (Ranford, 1989).

Concurrent with studies in this thesis, Bhatia et al. (1991) conducted complement MHC class III studies on PNG highland and lowland populations, and detected genetic variation at the \( Bf, C4A, \) and \( C4B \) gene loci. There was significant allelic diversity associated with all three loci in the lowland population, but significant allelic diversity occurred only at the \( C4B \) locus in the highland population. The estimated percentage of \( C4 \) null alleles at both \( C4A \) and \( C4B \) gene loci (9.0 % and 7.4 %, respectively), was high, but not as high as that found in the northern Aboriginal Australian population by Ranford et al. (1987). In addition, a low frequency of \( Bf F \) was detected in the highland region of PNG, \( Bf S \) being the most frequently occurring
allele (Ranford, 1989; Bhatia et al., 1991), whereas Bf F was the most frequently occurring allele in the lowland region (Bhatia et al., 1991).

3.2 Aim

The main aims of this study were: (1) To investigate genetic similarities and differences between northern Aboriginal Australians and populations from coastal, highland and inland regions of PNG; (2) To identify individuals with null or rare C4A and C4B alleles for subsequent analysis at the molecular level, and to identify C4 allotypes for subsequent correlation with C4 alleles identified by DNA analysis for the purpose of developing a method for molecular typing of C4. In addition, this study enabled a comparison of C4 null allotypes of other northern located Aboriginal Australians with the high frequencies found in this region by Ranford et al. (1987).

3.3 Materials and Methods

3.3.1 Aboriginal Australian and PNG population donors

Peripheral blood specimens were obtained from 101 Aboriginal Australian donors in two Cape York communities, namely East Cape York (ECY) and West Cape York (WCY). Immediately after collection, samples were placed on dry ice, then stored at -70°C. PNG blood specimens were collected from 49 mothers attending the Goroka Base Hospital for routine antenatal check-ups in the highland (H) region of PNG. Further PNG specimens were taken from 51 donors participating in a population genetics survey of the Wosera inland (I) sub-district, and from 120 donors of the coastal (C) region of Koki, Port Morseby. All PNG plasma samples were separated from buffy coats within one to two hours of collection, treated with 0.2 M EDTA to prevent protein deterioration, and snap frozen in liquid nitrogen. Samples were transported on dry ice to the John Curtin School of Medical Research (JCSMR), Canberra for C4 protein analysis.
3.3.2 C4 protein allotyping

C4 allotyping was performed by protein gel electrophoresis and immunofixation, following the method described by Awdeh and Alper (1980), and as described in Chapter 2, section 2.5. C4 allotyping results in the separation of C4 proteins according to their size and overall net charge. Unusual C4 allotypes and C4 null alleles were confirmed by treatment with Carboxypeptidase B prior to electrophoresis, as outlined in section 2.5.5. Homozygous null alleles were identified by the absence of C4 protein at either the anodal or cathodal region of the gel. Resulting C4 allotypes were analysed independently by at least one other person. To ensure typing results were reproducible all donors were C4 allotyped at least two times.

3.3.3 Estimation of C4 allelic frequencies

Maximum likelihood estimates of C4 gene frequencies were calculated from allotype distributions using an iterative computer programme to determine the best fit to Hardy-Weinberg equilibrium. This computer analysis was necessary to estimate the population frequency of double heterozygous null alleles which cannot be detected by protein gel electrophoresis. C4 allotypes which have one C4A or C4B null allele, such as C4A 3,Q0 B 2, can be recognised by a less intense banding pattern for the 'A 3' protein in comparison to the 'B 2' protein. The banding pattern for the 'B 2' alleles, in this allotype would be twice as dense as that of the A 3,Q0 alleles. A double heterozygous C4 null allotype, such as A 3,Q0 B 2,Q0, would have C4A and C4B protein bands of equal intensity. Due to the variability in the normal level of circulating C4 protein in a population, it is virtually impossible to identify double heterozygous C4 null alleles in any particular allotype, and therefore these must be estimated by statistical analysis.

3.3.4 Statistical analysis

Chi-square ($X^2$) was used to test the significant difference between observed and expected numbers of C4 alleles for each population and for comparing the estimated C4 gene frequencies. Differences in
the distribution of C4A and C4B allotypes between each population were tested by chi-square with continuity correction. If a allotype frequency was below five within the cell of a 2 x 2 contingency table, the exact probability was calculated (Armitage, 1971). To establish if linkage disequilibrium existed between the most frequently occuring C4 haplotypes in each population, frequencies of these haplotypes were calculated according to Mittal et al. (1973). The non-random association of C4A and C4B alleles was measured by the disequilibrium statistic delta (Mittal et al., 1973). The significance of linkage disequilibrium was tested by chi-square. A kinship coefficient equation (Cavalli-Sforza and Bodmer, 1971a), was used to estimate the genetic distance between each population based on C4A and C4B gene frequencies.

3.4 Results

3.4.1 C4 allotypes identified by protein gel electrophoresis and immunofixation

Figure 3.1, shows various C4A and C4B allotypes in East and West Cape York Aboriginal populations, and Figures 3.2 and 3.3, show C4 allotypes from each PNG population. C4 allotypes represented by three protein bands (as in Figure 3.1 'C'), had been desialysed with neuraminidase prior to electrophoresis, and those represented by one protein band (as in Figure 3.1 'D'), had been treated with Carboxypeptidase B and neuraminidase. Carboxypeptidase B was used on serum samples suspected to be degraded, and was used to distinguish C4 allotypes of individuals who had overlapping C4 protein bands.

3.4.2 C4 allelic polymorphism

Four to six different C4A alleles were detected in both Aboriginal Australian (Table 3.1) and PNG populations (Table 3.2), except for the relatively monomorphic inland population of PNG where only two types of C4A alleles were detected. Of the PNG populations, the inland population was least polymorphic followed by the highland
Figure 3.1 C4 allotypes in East and West Cape York Aboriginal Australian populations. Schematic diagram showing the relative migration distances of C4 allotypes (VI complement genetics workshop reference typing), is reproduced with permission from WHO Bulletin, 70: 531-535, (1992). Proteins represented by three bands, for example in 'C', were treated with neuraminidase prior to electrophoresis, and those represented by one band such as in 'D', were treated with neuraminidase and Carboxypeptidase B.
**Figure 3.2** C4A allotypes in coastal, inland and highland regions of PNG. Schematic diagram showing the relative migration distances of C4 allotypes (VI complement genetics workshop reference typing), is reproduced with permission from *WHO Bulletin*, 70: 531-535, (1992). Proteins in 'A', 'B', 'C' and 'D' were treated with neuraminidase and *Carboxypeptidase B*, whereas those in 'E' were treated with neuraminidase prior to electrophoresis.
Figure 3.3 *C4B* allotypes in coastal, inland and highland regions of PNG. Schematic diagram showing the relative migration distances of *C4* allotypes (VI complement genetics workshop reference typing), is reproduced with permission from *WHO Bulletin*, 70: 531-535, (1992). Proteins in 'A', 'B' (third lane), 'C' and 'E' (first lane) were treated with neuraminidase, whereas those in 'B' (first two lanes), 'D', and 'E' (second lane) were treated with neuraminidase and *Carboxypeptidase B* prior to eletrophoresis.
Table 3.1 Observed and expected numbers of C4A allotypes in East and West Cape York Aboriginal Australians

<table>
<thead>
<tr>
<th>Allotype</th>
<th>ECY Observed</th>
<th>ECY Expected</th>
<th>WCY Observed</th>
<th>WCY Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>F</td>
<td>n</td>
<td>F</td>
</tr>
<tr>
<td>A 2,2/2,Q0</td>
<td>1</td>
<td>1.3</td>
<td>1.91</td>
<td>0</td>
</tr>
<tr>
<td>A 2,3</td>
<td>1</td>
<td>1.3</td>
<td>1.94</td>
<td>1</td>
</tr>
<tr>
<td>A 2,4</td>
<td>7</td>
<td>9.3</td>
<td>4.33</td>
<td>1</td>
</tr>
<tr>
<td>A 2,6</td>
<td>0</td>
<td>0.0</td>
<td>0.42</td>
<td>0</td>
</tr>
<tr>
<td>A 2,X</td>
<td>0</td>
<td>0.0</td>
<td>0.06</td>
<td>0</td>
</tr>
<tr>
<td>A 3,3/3,Q0</td>
<td>15</td>
<td>20.0</td>
<td>9.58</td>
<td>8</td>
</tr>
<tr>
<td>A 3,4</td>
<td>*6</td>
<td>8.0</td>
<td>15.83</td>
<td>*1</td>
</tr>
<tr>
<td>A 3,6</td>
<td>#5</td>
<td>6.7</td>
<td>1.53</td>
<td>1</td>
</tr>
<tr>
<td>A 3,X</td>
<td>0</td>
<td>0.0</td>
<td>0.22</td>
<td>0</td>
</tr>
<tr>
<td>A 4,4/4,Q0</td>
<td>36</td>
<td>48.0</td>
<td>31.04</td>
<td>14</td>
</tr>
<tr>
<td>A 4,6</td>
<td>2</td>
<td>2.7</td>
<td>3.41</td>
<td>0</td>
</tr>
<tr>
<td>A 4,X</td>
<td>1</td>
<td>1.3</td>
<td>0.48</td>
<td>0</td>
</tr>
<tr>
<td>A 6,6/6,Q0</td>
<td>0</td>
<td>0.0</td>
<td>1.46</td>
<td>0</td>
</tr>
<tr>
<td>A 6,X</td>
<td>0</td>
<td>0.0</td>
<td>0.05</td>
<td>0</td>
</tr>
<tr>
<td>A X,X,X,Q0</td>
<td>0</td>
<td>0.0</td>
<td>0.19</td>
<td>0</td>
</tr>
<tr>
<td>A Q0,Q0</td>
<td>1</td>
<td>1.3</td>
<td>2.55</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>#75</td>
<td>100.0</td>
<td>75.00</td>
<td>26</td>
</tr>
</tbody>
</table>

* = 0.01 < p < 0.05; # = p < 0.01; 'n' = number; 'F'= frequency of observed numbers shown as a percentage; 'X' refers to a novel C4A allele which migrates between an A 2 and an A 1 allele by protein gel electrophoresis
Table 3.2 Observed and expected numbers of C4A alleles in three PNG populations

<table>
<thead>
<tr>
<th>Allotype</th>
<th>Highland</th>
<th>Inland</th>
<th>Coastal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed</td>
<td>Expected</td>
<td>Observed</td>
</tr>
<tr>
<td></td>
<td>n  F  n</td>
<td>n  F  n</td>
<td>n  F  n</td>
</tr>
<tr>
<td>A 2,2/2,Q0</td>
<td>0</td>
<td>0.00</td>
<td>0</td>
</tr>
<tr>
<td>A 2,3</td>
<td>1 2.0 0.98</td>
<td>1 2.0 0.99</td>
<td>0</td>
</tr>
<tr>
<td>A 2,5</td>
<td>0 0.01</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>A 2,6,6</td>
<td>0 0.01</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>A 3,3/3,Q0</td>
<td>46 93.9 46.02</td>
<td>50 98.0 50.01</td>
<td>99 82.5 99.62</td>
</tr>
<tr>
<td>A 3,4</td>
<td>0</td>
<td>0</td>
<td>8 6.7 7.05</td>
</tr>
<tr>
<td>A 3,5</td>
<td>1 2.0 0.99</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A 3,51</td>
<td>0</td>
<td>0</td>
<td>1 0.8 0.91</td>
</tr>
<tr>
<td>A 3,6</td>
<td>0</td>
<td>0</td>
<td>2 1.7 2.80</td>
</tr>
<tr>
<td>A 3,6,6</td>
<td>1 2.0 0.98</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A 4,4/4,Q0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A 4,5</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A 4,51</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A 4,6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A 5,5/5,Q0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A 5,51</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A 5,6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A 5,6,6</td>
<td>0 0.01</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A 51,51/51,Q0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A 51,6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A 6,6/6,Q0</td>
<td>0</td>
<td>0</td>
<td>1 0.8 0.12</td>
</tr>
<tr>
<td>A 6,6,6,6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A Q0,Q0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>49 100.0 49.00</td>
<td>51 100.0 51.00</td>
<td>120 100.0 120.00</td>
</tr>
</tbody>
</table>

# = p < 0.01; 'n' = number; 'F' = frequency of observed numbers shown as a percentage
then coastal populations. Both East and West Cape York Aboriginal populations were found to be equally polymorphic at the C4A gene locus. The C4B gene locus was equally polymorphic, where four to five different alleles were detected in the East and West Cape York Aboriginal populations (Table 3.3) and each of the three PNG populations (Table 3.4).

3.4.3 Testing the goodness of fit of the C4 allele frequencies to the Hardy-Weinberg Equilibrium

Numbers of observed and expected C4 allotypes and frequencies of observed C4 allotypes of each population studied are shown in Tables 3.1-3.4. Chi-square statistical analysis of the C4A and C4B allele frequencies did not detect a significant difference between the number of observed and expected C4A and C4B alleles for each population, except for the C4A alleles of East Cape York Aborigines. In this population, the total chi-square value for the observed allele frequencies was significantly different to that of the expected allele frequencies (p < 0.01). This discrepancy resulted from a lack of C4A 3,4 heterozygotes (0.01 < p < 0.05), and an abundance of C4A 3,6 heterozygotes (p < 0.01). Similarly, a lack of C4A 3,4 heterozygotes were detected in the West Cape York population (0.01 < p < 0.05) when this particular allotype was tested. However, a significant difference was not detected when the total chi-square value of all observed C4A allotype frequencies in the West Cape York population were tested. The only other population where C4A 3,4 heterozygotes were detected was the coastal PNG population. In this population a significant difference between observed and expected allotype frequencies was not evident when testing for Hardy-Weinberg equilibrium.

3.4.4 C4 gene frequency analysis

From the observed C4A and C4B allele frequencies of each population, the iterative computer programme estimated gene frequencies expressed as percentage for each C4 allele. C4A and C4B gene frequencies for the two Aboriginal Australian and three PNG populations are shown in Tables 3.5 and 3.6, respectively.
Table 3.3 Observed and expected numbers of $C4B$ alleles in East and West Cape York Aboriginal Australians

<table>
<thead>
<tr>
<th>Allotype</th>
<th>ECY</th>
<th>WCY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed n F</td>
<td>Expected n</td>
</tr>
<tr>
<td>--------------------</td>
<td>--------------</td>
<td>--------------</td>
</tr>
<tr>
<td>$B\ 1,1/1,Q0$</td>
<td>19 25.3</td>
<td>22.25</td>
</tr>
<tr>
<td>$B\ 1,12$</td>
<td>1 1.3</td>
<td>0.44</td>
</tr>
<tr>
<td>$B\ 1,2$</td>
<td>32 42.7</td>
<td>26.50</td>
</tr>
<tr>
<td>$B\ 1,4$</td>
<td>0 0.0</td>
<td>0.82</td>
</tr>
<tr>
<td>$B\ 12,12/12,Q0$</td>
<td>0 0.0</td>
<td>0.15</td>
</tr>
<tr>
<td>$B\ 12,2$</td>
<td>0 0.0</td>
<td>0.43</td>
</tr>
<tr>
<td>$B\ 12,4$</td>
<td>0 0.0</td>
<td>0.01</td>
</tr>
<tr>
<td>$B\ 2,2/2,Q0$</td>
<td>18 24.0</td>
<td>21.85</td>
</tr>
<tr>
<td>$B\ 2,4$</td>
<td>2 2.7</td>
<td>0.81</td>
</tr>
<tr>
<td>$B\ 4,4/4,Q0$</td>
<td>0 0.0</td>
<td>0.28</td>
</tr>
<tr>
<td>$B\ Q0,Q0$</td>
<td>3 4.0</td>
<td>1.46</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>75 100.0</strong></td>
<td><strong>75.00</strong></td>
</tr>
</tbody>
</table>

'\text{n}' = number; 'F' = frequency of observed numbers shown as a percentage.
Table 3.4 Observed and expected numbers of $C4B$ alleles in three PNG populations

<table>
<thead>
<tr>
<th>Allotype</th>
<th>Highland Observed</th>
<th>Highland Expected</th>
<th>Inland Observed</th>
<th>Inland Expected</th>
<th>Coastal Observed</th>
<th>Coastal Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n F n</td>
<td></td>
<td>n F n</td>
<td></td>
<td>n F n</td>
<td></td>
</tr>
<tr>
<td>$B\ 1.1/1,Q0$</td>
<td>30 61.2 28.92</td>
<td></td>
<td>42 82.4 41.31</td>
<td></td>
<td>101 84.1 100.12</td>
<td></td>
</tr>
<tr>
<td>$B\ 1.2$</td>
<td>12 24.5 14.14</td>
<td></td>
<td>7 13.7 6.67</td>
<td>1 0.8 0.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B\ 1.4$</td>
<td>0 0</td>
<td></td>
<td>0</td>
<td>1 0.8 2.19</td>
<td>7 5.8 7.08</td>
<td></td>
</tr>
<tr>
<td>$B\ 1.5$</td>
<td>0 0</td>
<td></td>
<td>0</td>
<td>0.94 0</td>
<td>7 5.8 7.08</td>
<td></td>
</tr>
<tr>
<td>$B\ 1.6$</td>
<td>0 0</td>
<td></td>
<td>0</td>
<td>0.94 0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>$B\ 1.95$</td>
<td>1 2.0 0.67</td>
<td></td>
<td>0</td>
<td>0.94 0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>$B\ 1.96$</td>
<td>0 1 0.94 0</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>$B\ 2.2/2,Q0$</td>
<td>6 12.2 4.42</td>
<td></td>
<td>0</td>
<td>0.07 0</td>
<td>0</td>
<td>0.24</td>
</tr>
<tr>
<td>$B\ 2.4$</td>
<td>0 0</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>$B\ 2.5$</td>
<td>0 0</td>
<td></td>
<td>0</td>
<td>0.07 0</td>
<td>0</td>
<td>0.04</td>
</tr>
<tr>
<td>$B\ 2.6$</td>
<td>0 0</td>
<td></td>
<td>0</td>
<td>0.07 0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>$B\ 2.95$</td>
<td>0 0.21</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>$B\ 2.96$</td>
<td>0 0</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>$B\ 4.4/4,Q0$</td>
<td>0 0</td>
<td></td>
<td>1</td>
<td>0.8 0.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B\ 4.5$</td>
<td>0 0</td>
<td></td>
<td>1</td>
<td>0.8 0.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B\ 5.5/5,Q0$</td>
<td>0 0</td>
<td></td>
<td>2</td>
<td>1.7 2.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B\ 5.6$</td>
<td>0 #1 2.0 0.01</td>
<td></td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B\ 5.96$</td>
<td>0 0</td>
<td></td>
<td>0</td>
<td>0.01 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B\ 6.6/6,Q0$</td>
<td>0 0</td>
<td></td>
<td>0</td>
<td>0.01 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B\ 6.96$</td>
<td>0 0</td>
<td></td>
<td>0</td>
<td>0.01 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B\ 95,95/95,Q0$</td>
<td>0 0.11</td>
<td></td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B\ 96,96/96,Q0$</td>
<td>0 0</td>
<td></td>
<td>0</td>
<td>0.01 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B\ Q0,Q0$</td>
<td>0 0.52</td>
<td></td>
<td>0</td>
<td>0.06 6</td>
<td>5.0 6.19</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>49 100.0 49.00</td>
<td></td>
<td>#51 100.0 51.00</td>
<td>120 100.0 120.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

# = p < 0.01; 'n' = number; 'F' = frequency of observed numbers shown as a percentage
Table 3.5 *C4A* gene frequencies of two Aboriginal and three PNG populations

<table>
<thead>
<tr>
<th>Allele</th>
<th>Aboriginal East Cape York n=75</th>
<th>Aboriginal West Cape York n=26</th>
<th>PNG Highland n=49</th>
<th>PNG Inland n=51</th>
<th>PNG Coastal n=120</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GF</td>
<td>SE</td>
<td>GF</td>
<td>SE</td>
<td>GF</td>
</tr>
<tr>
<td>A 2</td>
<td>5.9</td>
<td>0.020</td>
<td>3.8</td>
<td>0.027</td>
<td>1.0</td>
</tr>
<tr>
<td>*A 3</td>
<td>21.8</td>
<td>0.034</td>
<td>26.5</td>
<td>0.066</td>
<td>97.4</td>
</tr>
<tr>
<td>*A 4</td>
<td>48.5</td>
<td>0.048</td>
<td>43.4</td>
<td>0.079</td>
<td>0.0</td>
</tr>
<tr>
<td>A 5</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>A 5I</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>A 6</td>
<td>4.7</td>
<td>0.017</td>
<td>2.0</td>
<td>0.019</td>
<td>0.0</td>
</tr>
<tr>
<td>A 6,6</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>AX</td>
<td>0.7</td>
<td>0.007</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>*A Q0</td>
<td>18.4</td>
<td>0.041</td>
<td>24.3</td>
<td>0.078</td>
<td>0.0</td>
</tr>
</tbody>
</table>

GF= gene frequency shown as a percentage, SE= standard error '*': major C4A allotypes in the respective populations
Table 3.6 *C4B* gene frequencies of two Aboriginal and three PNG populations

<table>
<thead>
<tr>
<th>Allele</th>
<th>Aboriginal</th>
<th></th>
<th></th>
<th>PNG</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>East Cape York n=75</td>
<td>West Cape York n=26</td>
<td>Highland n=49</td>
<td>Inland n=51</td>
<td>Coastal n=120</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GF</td>
<td>SE</td>
<td>GF</td>
<td>SE</td>
<td>GF</td>
<td>SE</td>
<td>GF</td>
</tr>
<tr>
<td><em>B1</em></td>
<td>42.3</td>
<td>0.048</td>
<td>24.8</td>
<td>0.065</td>
<td>67.2</td>
<td>0.059</td>
<td>93.6</td>
</tr>
<tr>
<td>B12</td>
<td>0.7</td>
<td>0.007</td>
<td>4.1</td>
<td>0.028</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td><em>B2</em></td>
<td>41.8</td>
<td>0.042</td>
<td>51.8</td>
<td>0.080</td>
<td>21.5</td>
<td>0.044</td>
<td>7.0</td>
</tr>
<tr>
<td>B4</td>
<td>1.3</td>
<td>0.009</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1.3</td>
</tr>
<tr>
<td>B5</td>
<td>0.0</td>
<td>0.000</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
<td>4.1</td>
</tr>
<tr>
<td>B6</td>
<td>0.0</td>
<td>0.000</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>B95</td>
<td>0.0</td>
<td>0.000</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
<td>0.010</td>
<td>0.0</td>
</tr>
<tr>
<td>B96</td>
<td>0.0</td>
<td>0.000</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td><em>BQ0</em></td>
<td>14.0</td>
<td>0.044</td>
<td>19.2</td>
<td>0.079</td>
<td>10.3</td>
<td>0.057</td>
<td>0.0</td>
</tr>
</tbody>
</table>

GF = gene frequency shown as a percentage, SE = standard error *"* = major *C4B* allotypes in the respective populations
Results from this analysis are reported below.

3.4.4.1 C4A gene frequency analysis

Analysis of the C4A gene frequencies in East and West Cape York Aboriginal Australian populations found three most frequently occurring alleles, 'A 3', 'A 4' and 'A Q0', at the C4A gene locus, where A 4 had the highest frequency in each population. Ranford et al. (1987), also found that in studies of central and northern Aboriginal Australian populations the C4A gene locus had the same most frequently occurring alleles. 'A 3' and 'A 4' allele frequencies (42.8 % and 39.9 %, respectively) were not significantly different in the central population, however, the 'A 3' allele frequency was significantly higher than 'A 4' (47.2 % and 22.0 %) in the northern population. The 'A 3' allele frequency between central and northern Aboriginal Australian populations was not significantly different, however, a significant difference between the 'A 4' allele frequency in these two populations was detected. The frequency of C4A Q0 alleles in West Cape York Aboriginals was high (24.3 %), which was similar and not significantly different to that of the northern Aboriginal Australian population (28.8 %) studied by Ranford et al. (1987). East Cape York has a lower frequency of C4A Q0 alleles (18.4 %), which is not significantly different to the central Aboriginal Australians (12.3 %), reported by Ranford et al. (1987), and to the Caucasian population (16.8 %), reported by Ranford et al. (1986).

As distinct from the Aboriginal Australians, the C4A gene locus of each PNG population was dominated by one allele, 'A 3'. All other C4A alleles had relatively low frequencies. The highest frequencies of the 'A 3' allele were detected in the inland and highland PNG populations (99.5 % and 97.4 %, respectively), whilst the coastal population was more heterogeneous resulting in a slightly lower and significantly different 'A 3' frequency (88.2 %). Bhatia et al. (1991), found the 'A 3' allele to be the most frequently occurring C4A allele in a study of PNG populations. Bhatia et al. (1991) reported the highland population to have a significantly higher 'A 3' frequency of 88.4 %, compared to the lowland population with a frequency of 72.2 %. The highest 'A 3' allele frequency was found in the inland PNG
population, in this study, which is significantly different to that of the highland PNG population reported by Bhatia. The lowest 'A 3' allele frequency, found in the coastal PNG population is significantly different to that reported for the lowland PNG population by Bhatia. A significant difference was found here between the frequency of C4A Q0 alleles in PNG populations compared to the Aboriginal Australian populations. C4A Q0 alleles were absent in both highland and inland populations of PNG, and occurred with a very low frequency (3.0 %), in the coastal population. By comparison, Bhatia et al. (1991), found a slightly higher frequency of C4A Q0 in PNG, 10.8 % in the lowland population (not significantly different to the Aboriginal Australians of Cape York) and 7.4 % in the highland population (significantly different to the Aboriginal Australians of Cape York).

Besides the most frequently occurring C4A alleles, rare C4A alleles were also detected in each population, and many were unique to either Aboriginal Australian or PNG populations. Rare C4A alleles which were detected in both Aboriginal Australian and PNG populations included 'A 2', and 'A 6'. Those which appeared to be unique to a particular population included; 'A 5', detected only in the highland and coastal PNG populations; 'A 51', found in the coastal PNG population; 'A 6', in Aboriginal Australian and PNG coastal populations; a duplicated 'A 6,6' allele (as determined by different protein band intensities), found only in the PNG highland population, and 'A X' (an allele which has not previously been reported), was found in the East Cape York Aboriginal population. The 'A X' allele migrated between 'A 2' and 'A 1' alleles by protein gel electrophoresis. A similar array of rare C4A alleles was detected in the Aboriginal populations studied by Ranford et al. (1987). However, Ranford et al. (1987) found another C4A allele, 'A 12', in the central Australian population which was not detected here. Bhatia et al. (1991), found 'A 12' and 'A 55' alleles in a lowland population of PNG which were absent in PNG populations studied here.
3.4.4.2 **C4B** gene frequency analysis

Estimated **C4B** gene frequencies of ECY and WCY Aboriginal Australian populations and of the PNG populations are listed in Table 3.6. Results show that, similar to the **C4A** gene locus of East and West Cape York Aborigines, a number of more frequently occurring types of **C4B** alleles were found at the **C4B** gene locus. Alleles 'B 1' and 'B 2' occurred with the highest frequencies, followed by the 'B Q0' allele which had a slightly lower frequency. Both 'B 1' and 'B 2' alleles occurred with similar and not significantly different frequencies in the East Cape York population (42.3 % and 41.8 %, respectively), whereas the 'B 2' allele occurred with a higher and significantly different frequency than the 'B 1' allele (51.8 % and 24.8 %, respectively) in the West Cape York population. Similarly, Ranford et al. (1987), found that the central and northern Aboriginal populations had high frequencies of the 'B 1', 'B 2' and 'B Q0' alleles, the most frequently occurring being 'B 1' and 'B 2'. Ranford et al. (1987), found that frequencies of these alleles varied between populations. The 'B 2' allele of the central Aboriginal Australian population, like that of the West Cape York population, had the highest frequency, however was not significantly different to the 'B 1' allele which was the most frequently occurring **C4** allele in the northern Aboriginal Australian population.

Frequencies of **C4B Q0** alleles (14.0 % to 19.2 %), found here, were similar and not significantly different to those found by Ranford et al. (1986), in Caucasians (20.0 %), and central Aboriginal Australian (18.0 %), and northern Aboriginal Australian populations (25.7 %) reported by Ranford et al. (1987).

In the PNG populations, **C4B** gene frequencies were similar to those of the **C4A** gene, having one frequently occurring **C4B** allele, namely 'B 1' (Table 3.6). The highest frequency of 'B 1' was detected in the inland PNG population (93.6 %), whereas both highland and coastal populations had similar, although slightly lower 'B 1' allele frequencies (67.2 % and 71.4 %, respectively). The 'B 2' allele was the second most frequently occurring **C4B** allele in the highland population (21.5 %), but low frequencies of the 'B 2' allele occurred
in both inland and coastal populations (7.0 % and 0.4 %, respectively). Bhatia et al. (1991), found the 'B 1' and 'B 2' alleles were the most frequently occurring C4B alleles in each PNG population, where the 'B 1' allele occurred with the highest frequencies of 65.4 % and 71.4 % in the highland and lowland populations, respectively. The 'B 1' allele frequencies, reported by Bhatia et al. (1991), are not significantly different to each other or to those found for the highland and coastal PNG populations in this study, but are significantly different to the inland PNG population in this study. The 'B 2' allele frequencies in both highland and lowland populations (24.5 % and 15.0 %, respectively) reported by Bhatia et al. (1991), are similar and not significantly different to the frequency of the 'B 2' allele found here for the highland population (21.5 %), but are significantly different to the inland and coastal PNG populations. C4B Q0 alleles in this study were found in both highland and coastal populations (10.3 % and 22.7 %, respectively), but not in inland PNG populations. Frequency of C4B Q0 alleles in the coastal PNG population found here is similar and not significantly different to that reported in the northern Aboriginal Australians (25.7 %) by Ranford et al. (1987). Bhatia et al. (1991), detected a higher and significantly different frequency of C4B Q0 alleles in the lowland region of PNG compared to the PNG highland region (11.4 % and 3.6 %, respectively). These frequencies are not significantly different to those detected in the highland PNG region, however, they are significantly different to those detected in the coastal PNG region in this study.

Rare allele variants were also detected at the C4B gene locus, many of which were only detected in one or two of the populations studied. The rare C4B allele 'B 12', occurred at a low frequency in both the ECY and WCY Aboriginal Australian populations of this study (0.7 % and 4.1 %, respectively), but was absent in the PNG populations. This allele was also detected with low frequencies in both central and northern Aboriginal Australian populations by Ranford et al. (1987). The 'B 12' allele frequency was not significantly different between each of these populations. Another C4B allele, 'B 4', was found without a significant difference of frequency in the East Cape York Aborigines and coastal population of PNG in this study. Ranford et al.
(1987) also detected this allele, without a significant difference in frequency, in central Aboriginal Australians, as well as the 'B 3' allele in the northern Aboriginal population. The 'B 3' allele was not found in either of the Cape York Aboriginal populations in this study. Of other rare C4B alleles, recorded in this study, 'B 5', was only detected in the inland and coastal regions of PNG, 'B 6', was unique to the inland PNG population, 'B 95', was detected only in the highland PNG population, and 'B 96', was recorded in the inland PNG population. This study did not find the 'B 3' allele in both highland and lowland PNG populations as reported by Bhatia et al. (1991).

### 3.4.4.3 Significant differences in C4 gene frequencies between different Aboriginal Australian and PNG populations

C4 gene frequencies were compared between different Aboriginal Australian and PNG populations using chi-square to indicate populations' degree of similarity at both the C4A and C4B loci. A summary of those C4 gene frequencies which were significantly different between populations is shown in Table 3.7.

The majority of significant differences recorded between populations were between the most frequently occurring C4 alleles. C4A and C4B gene frequencies were very similar for both Aboriginal populations with only the 'B 1' gene frequency being found to be significantly different. There were no significant differences of C4A gene frequencies between the PNG highland and inland populations. However, the 'A 3' gene frequencies of the PNG highland and coastal populations were significantly different, as were the 'A 3' and 'A 5' gene frequencies of the inland and coastal populations. There were more C4B gene frequency differences (namely 'B 1', 'B 2', 'B 5', and 'B Q0' alleles), between PNG populations than there were for the C4A gene.

Comparing gene frequencies of each of the Aboriginal Australian populations with each of the three PNG populations, found many significant differences for both the C4A and C4B gene loci. The Aboriginal East Cape York and PNG coastal populations were most dis-similar with respect to their C4 gene frequencies. By
Table 3.7 Statistically significant differences of C4 gene frequencies between Aboriginal Australian and PNG populations

<table>
<thead>
<tr>
<th>Allotype</th>
<th>WCY/ECY</th>
<th>H/I</th>
<th>H/C</th>
<th>I/C</th>
<th>ECY/C</th>
<th>ECY/I</th>
<th>ECY/H</th>
<th>WCY/C</th>
<th>WCY/I</th>
<th>WCY/H</th>
</tr>
</thead>
<tbody>
<tr>
<td>A 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>O</td>
<td>X</td>
<td>O</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* A 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>* A 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>A 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>* A Q0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

* B 1     | X       | O   | O   | O   | O     | O     | O     | O     | O     | O     |
| B 12     |         |     |     |     |       |       |       | O     | X     | X     |
| * B 2    |         |     |     |     | O     | O     | O     | O     | O     | O     |
| B 5      |         |     |     |     |       | X     |       | X     |       |       |
| * B Q0   |         |     |     |     | O     | O     | O     | X     | O     |       |
| Total    |         | X   | O   | O   | O     | O     | O     | O     | O     | O     |

'X' = 0.01 < p < 0.05; 'O' = p < 0.01; 'C' refers to coastal, 'I' refers to inland, and 'H' refers to highland PNG populations. 'WCY': West Cape York Aboriginal population; 'ECY': East Cape York Aboriginal population. '*' : C4 alleles occurring most frequently in the respective populations.
comparison, there was greater similarity between the Aboriginal East Cape York and PNG highland populations, and between the Aboriginal West Cape York and PNG highland populations.

3.4.5 Distribution of $C4A$ and $C4B$ allotypes within Aboriginal Australian populations

Comparisons of the $C4$ haplotype (including both $C4A$ and $C4B$ alleles), expressed as a percentage, in each population, (Table 3.8), found that for the West Cape York Aboriginal population, the most frequently occurring allotype was 'A 4 B 2' (42.3 %), followed by 'A 4 B 1,2' and 'A 3 B 1' with equal frequency, (11.5 %). The frequencies of these allotypes in the East Cape York population were slightly different, 'A 4 B 2' (21.3 %) and 'A 4 B 1,2' (20.0 %), followed by 'A 3 B 1' (13.3 %).

3.4.6 Distribution of $C4A$ and $C4B$ allotypes within three PNG populations

The distribution of $C4A$ and $C4B$ allotypes are shown as numbers in Table 3.9, and as percentage distribution in Table 3.10. Comparing all $C4$ allotypes found in all three populations 'A 3 B 1' was the most frequently occurring, with the highest frequency of 82.4 % occurring in the inland population, followed by 73.3 % in the coastal population, and 55.1 % in the highland population. Only one other allotype was found frequently occurring in the inland population, namely, 'A 3 B 1,2', with a frequency of 13.7 %. $C4$ allotypes, 'A 3 B 1,2' and 'A 3 B 2', were reasonably common in the highland population (24.5 % and 12.2 %, respectively), whereas, 'A 3.5 B 1', 'A 3.4 B 1.5', and 'A 3 B Q0', were other frequently occurring allotypes in the coastal population (7.5 %, 5.8 %, and 5.0 %, respectively).
<table>
<thead>
<tr>
<th>C4B allotype</th>
<th>1,1/1,Q0</th>
<th>1,12</th>
<th>1,2</th>
<th>12,12/12,Q0</th>
<th>2,2/2,Q0</th>
<th>2,4</th>
<th>Q0,Q0</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>C4A allotype</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ECY N %</td>
<td>WCY N %</td>
<td>ECY N %</td>
<td>WCY N %</td>
<td>ECY N %</td>
<td>WCY N %</td>
<td>ECY N %</td>
<td>WCY N %</td>
</tr>
<tr>
<td>2,2/2,Q0</td>
<td>1 1.3 0</td>
<td>0 0 0</td>
<td>0 0 0</td>
<td>0 0 0</td>
<td>0 0 0</td>
<td>0 0 0</td>
<td>0 0 0</td>
<td>0 0 0</td>
</tr>
<tr>
<td>2,3</td>
<td>0 1.3 0</td>
<td>0 0 0</td>
<td>1 1.3 0</td>
<td>0 0 0</td>
<td>0 0 0</td>
<td>0 0 0</td>
<td>0 0 0</td>
<td>0 0 0</td>
</tr>
<tr>
<td>2,4</td>
<td>0 0 0</td>
<td>0 0 0</td>
<td>6 8.0 0</td>
<td>0 0 0</td>
<td>1 1.3 0</td>
<td>0 0 0</td>
<td>7 9.3 1 3.8</td>
<td></td>
</tr>
<tr>
<td>3,3/3,Q0</td>
<td>10 13.3 3 11.5</td>
<td>1 1.3 1 3.8 2 2.7 1 3.8 0</td>
<td>1 1.3 1 3.8 2 7.7 0 0</td>
<td>1 1.3 0</td>
<td>15 20.0 8 1 30.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3,4</td>
<td>1 1.3 0</td>
<td>0 0 0</td>
<td>5 6.7 1 3.8 0</td>
<td>0 0 0</td>
<td>0 0 0</td>
<td>0 0 0</td>
<td>6 8.0 1 3.8</td>
<td></td>
</tr>
<tr>
<td>3,6</td>
<td>4 5.3 1 3.8</td>
<td>0 0 0</td>
<td>1 1.3 0</td>
<td>0 0 0</td>
<td>0 0 0</td>
<td>0 0 0</td>
<td>5 6.7 1 3.8</td>
<td></td>
</tr>
<tr>
<td>4,4/4,Q0</td>
<td>2 2.7 0</td>
<td>0 0 0</td>
<td>15 20.0 3 11.5 0</td>
<td>0 0 0</td>
<td>16 21.3 11 42.3 1 3.0</td>
<td>2 2.7 0</td>
<td>36 48 1 4 53.8</td>
<td></td>
</tr>
<tr>
<td>4,6</td>
<td>0 0 0</td>
<td>0 0 0</td>
<td>1 1.3 0</td>
<td>0 0 0</td>
<td>1 1.3 0</td>
<td>0 0 0</td>
<td>2 2.7 0</td>
<td>0 0 0</td>
</tr>
<tr>
<td>4,X</td>
<td>0 0 0</td>
<td>0 0 0</td>
<td>1 1.3 0</td>
<td>0 0 0</td>
<td>0 0 0</td>
<td>0 0 0</td>
<td>1 1.3 0</td>
<td>0 0 0</td>
</tr>
<tr>
<td>Q0,Q0</td>
<td>1 1.3 0</td>
<td>0 0 0</td>
<td>0 0 0</td>
<td>0 0 0</td>
<td>0 0 0</td>
<td>0 0 0</td>
<td>1 1.3 0</td>
<td>0 0 0</td>
</tr>
<tr>
<td>Total</td>
<td>19 25.3 5 19.1</td>
<td>1 1.3 1 3.8 32 42.7 5 19.1 0</td>
<td>1 3.8 18 24 14 53.8 2 2.7 0</td>
<td>3 4 0</td>
<td>75 100 26 100</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

'X' refers to a C4A allele which migrates between an A 2 and A 1 allele by protein electrophoresis. 'N' = Number of a particular C4 allotype. 'WCY': West Cape York Aboriginal population; 'ECY': East Cape York Aboriginal population.
Table 3.9 Numbers of C4A and C4B allotypes in three PNG populations

<table>
<thead>
<tr>
<th>C4A Allele</th>
<th>C4B Allele</th>
<th>1.111/1.00</th>
<th>1.14</th>
<th>1.15</th>
<th>1.195</th>
<th>1.196</th>
<th>1.195/1.00</th>
<th>1.5</th>
<th>1.5/1.00</th>
<th>4.5</th>
<th>4.5/1.00</th>
<th>4.5</th>
<th>4.5/1.00</th>
<th>5.6</th>
<th>5.6/1.00</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>C</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>H</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>I</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>I</td>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>I</td>
<td>H</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>I</td>
<td>E</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>I</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>H</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

'H' represents the highland population. 'I' represents the inland population, and 'C' represents the coastal population.
Table 3.10 Distribution of C4A and C4B alleles expressed as percentages in three PNG populations

<table>
<thead>
<tr>
<th>C4A allotype</th>
<th>1-1/1.Q0</th>
<th>1-2</th>
<th>1-4</th>
<th>1-5</th>
<th>1-95</th>
<th>1-96</th>
<th>2-2/2.Q0</th>
<th>4-4/4.Q0</th>
<th>4-5</th>
<th>5-5/5.Q0</th>
<th>5-6</th>
<th>Q0-Q0</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-3</td>
<td>2.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.0</td>
</tr>
<tr>
<td>3-3/3, Q05.1</td>
<td>82.4</td>
<td>73.3</td>
<td>24.5</td>
<td>13.7</td>
<td>0.8</td>
<td>0</td>
<td>0</td>
<td>2.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.7</td>
<td>93.9</td>
</tr>
<tr>
<td>3-4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5.8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.8</td>
<td>6.7</td>
</tr>
<tr>
<td>3-5</td>
<td>2.0</td>
<td>0</td>
<td>0</td>
<td>7.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.0</td>
</tr>
<tr>
<td>3-51</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.8</td>
</tr>
<tr>
<td>3-6</td>
<td>0</td>
<td>0</td>
<td>1.7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.7</td>
</tr>
<tr>
<td>3-6,6</td>
<td>2.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.0</td>
</tr>
<tr>
<td>6-6/6, Q0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.8</td>
</tr>
<tr>
<td>Total</td>
<td>61.2</td>
<td>82.4</td>
<td>84.2</td>
<td>14.3</td>
<td>23.5</td>
<td>0.8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

'H' represents the highland population, 'I' represents the inland population, and 'C' represents the coastal population.
3.4.7 Significant differences of C4 allotype frequencies between different Aboriginal Australian and PNG populations

The degree of similarity between different Aboriginal Australian and PNG populations with respect to their C4A and C4B allotype frequencies was determined using chi-square with continuity correction and Fishers exact test. A summary of C4 allotype frequencies which were significantly different between populations is shown in Table 3.11.

Comparing the two Aboriginal Australian populations, results show that all of the C4 allotype frequencies were not significantly different except for allotypes having 'B2'. Comparative analysis of the different C4 allotype frequencies in the PNG populations found that the inland and highland populations were most similar. There were, by contrast, several significant differences in C4 allotype frequencies between the highland and coastal populations. Most differences in C4 allotype frequencies were found to occur between Aboriginal and PNG populations. Of the two Aboriginal populations, the East Cape York population was most different to the PNG populations, especially the coastal population.

In summary, as was previously found with comparisons of gene frequencies (Table 3.7), frequencies of different C4 allotypes of different Aboriginal Australian populations were quite similar. PNG populations, on the other hand were distinctly different from one another, whilst the majority of significant differences in C4 allotype frequencies were detected between Aboriginal Australians and the populations from PNG.

3.4.8 C4A and C4B linkage disequilibrium relationships in Aboriginal Australian and PNG populations

Results of C4A and C4B linkage disequilibrium relationships are shown in Table 3.12. 'A 3' and 'B 1' alleles were found to be more strongly linked in PNG populations compared to the Aboriginal Australians. This was especially so for the coastal and inland PNG populations. Linkage disequilibrium of 'A 3' and 'B 1' alleles was
<table>
<thead>
<tr>
<th>Allotype</th>
<th>WCY/ECY</th>
<th>H/I</th>
<th>H/C</th>
<th>I/C</th>
<th>ECY/C</th>
<th>ECY/I</th>
<th>ECY/H</th>
<th>WCY/C</th>
<th>WCY/I</th>
<th>WCY/H</th>
</tr>
</thead>
<tbody>
<tr>
<td>A 2,4 B 1,2</td>
<td>O</td>
<td></td>
<td>O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A 3 B 1</td>
<td>O</td>
<td>X</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>A 3 B 1,2</td>
<td>O</td>
<td></td>
<td></td>
<td>X</td>
<td>O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A 3 B 2</td>
<td>X</td>
<td>O</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A 3,4 B 1,2</td>
<td>O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A 3,4 B 1,5</td>
<td>O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A 3,5 B 1</td>
<td>O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A 4 B 1,2</td>
<td>O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A 4 B 2</td>
<td>O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Total A 2,4| O       |     |     |     |       |       |       |       |       |       |
| Total A 3   | O       |     |     |     |       |       |       |       |       |       |
| Total A 3,4 | O       |     |     |     |       |       |       |       |       |       |
| Total A 3,5 | O       |     |     |     |       |       |       |       |       |       |
| Total A 4   | O       |     |     |     |       |       |       |       |       |       |
| Total B 1   | X       | O   |     |     | O     | O     | O     | O     | O     | O     |
| Total B 1,2 | O       |     |     |     |       |       |       |       |       |       |
| Total B 1,5 | O       |     |     |     |       |       |       |       |       |       |
| Total B 2   | X       | O   |     |     | O     | O     | O     | O     | O     | O     |

X - 0.01 < p < 0.05  O - p < 0.01

'Total', refers to the total number of allotypes with a specific combination of C4A or C4B alleles. 'H' is highland PNG, 'T' is inland PNG, and 'C' is coastal PNG. 'WCY' and 'ECY': West and East Cape York Aboriginal populations, respectively.
Table 3.12 \(C4A\) and \(C4B\) allele linkage disequilibrium relationships in Aboriginal Australian and PNG populations

<table>
<thead>
<tr>
<th>Population</th>
<th>HF</th>
<th>(\Delta)</th>
<th>(\chi^2)</th>
<th>HF</th>
<th>(\Delta)</th>
<th>(\chi^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECY</td>
<td>0.066</td>
<td>0.052</td>
<td>0</td>
<td>0.109</td>
<td>0.074</td>
<td>0</td>
</tr>
<tr>
<td>WCY</td>
<td>0.055</td>
<td>0.037</td>
<td>0</td>
<td>0.230</td>
<td>0.127</td>
<td>0</td>
</tr>
<tr>
<td>Inland PNG</td>
<td>0.580</td>
<td>0.081</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highland PNG</td>
<td>0.130</td>
<td>-0.154</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coastal PNG</td>
<td>0.442</td>
<td>0.092</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(x = 0.01 < p < 0.05; \ o = p < 0.01\)

HF: Haplotype Frequency; \(\Delta\): delta; \(\chi^2\): chi-square

Linkage analysis was only performed on \(A 3 B 1\) allotypes in PNG populations as homozygous \(A 4 B 2\) allotypes were not detected in PNG populations, respectively.
found to be significant in the East Cape York Aboriginal population. Linkage disequilibrium of 'A 3' and 'B 1' was not detected in the highland PNG population nor in the West Cape York Aboriginal population. There was a greater degree of linkage disequilibrium of 'A 4' with 'B 2' alleles in each Aboriginal Australian population than there was between 'A 3' and 'B 1' alleles. Homozygous 'A 4 B 2' allotypes were not detected in the PNG populations.

3.4.9 Determination of the genetic distance between Aboriginal Australian and PNG populations

Calculating a kinship coefficient between populations (Cavalli-Sforza and Bodmer, 1971a), was used to determine the genetic distance between each population. This analysis brought together the $C4A$ and $C4B$ gene frequency data, and compared this information for each population. Kinship coefficients between populations and a summary of this analysis, represented by a dendrogram, is shown in Figure 3.4.

This analysis showed the Aboriginal Australian populations to be more closely related to each other than were the PNG populations. The inland and highland PNG populations were more closely related than either was with the coastal population. In contrast to the gene frequency and allotype frequency analysis, these results showed the East Cape York Aboriginal population, rather than the West Cape York population, to be more closely related to the PNG populations. The analysis also showed that the coastal PNG population was most closely related to the Aboriginal Australians. However, both $C4$ gene frequency and $C4$ allotype frequency analyses found that most differences occurred between the East Cape York Aborigines and the coastal PNG population (Tables 3.7 and 3.11).

3.5 Discussion

The distribution of observed $C4A$ alleles of East Cape York Aborigines, did not correspond to that expected when tested for Hardy-Weinberg equilibrium. A statistically significant difference
Figure 3.4 Genetic distance dendrogram based on $C4A$ and $C4B$ gene frequencies
H, I and C represent highland, inland and coastal PNG populations, respectively.
was found between observed and expected frequencies of 'A 3,4' and 'A 3,6' heterozygous allotypes. This indicated that an abundance of homozygous 'A 3' alleles may be masking the presence of 'A 3,4' heterozygotes. Such a discrepancy may result from technical error, as it is difficult to distinguish 'A 3,4' heterozygotes from 'A 3,3' homozygotes due to close migration of 'A 3' and 'A 4' alleles by protein gel electrophoresis. The combination of 'A 3' and 'A 4' alleles often appears as a smeared banding pattern, even after treatment with Carboxypeptidase B. As a result of this discrepancy, many donors in this population were subjected to re-analysis by C4 allotyping and Carboxypeptidase B treatment. Another factor which may affect the Hardy-Weinberg equilibrium of alleles is if a population size is too small. However, the size of the East Cape York population was relatively large (75 donors), and therefore not likely to be the source of this discrepancy. In the case of the West Cape York Aboriginal population, the chi-square value for comparing total observed and expected numbers of C4A alleles was not significantly different when tested for Hardy-Weinberg equilibrium. However, analysis of the single C4A 3,4 allotype, did detect a significant difference between observed and expected numbers in this population. The coastal PNG population was the only other population to have 'A 4' alleles. In this population the expected and observed numbers of allotypes were not significantly different, including C4A 3,4 heterozygotes. Misinterpretation of 'A 3,4' heterozygous alleles, by C4 protein gel electrophoresis in the coastal PNG population would be less evident compared to the two Aboriginal Australian populations, as the C4A gene locus is dominated by the 'A 3' allele, and the 'A 4' allele occurs with a much lower frequency. Whereas, misinterpretation of the 'A 3,4' heterozygote would be more evident in the Aboriginal Australian populations, where both 'A 3' and 'A 4' alleles have high frequencies.

Observed deviations from the Hardy-Weinberg equilibrium are rare. Cavalli-Sforza and Bodmer, (1971b), have suggested that heterozygous allele deficiency, in addition to technical error, may result from one or more of five events: (1) Heterogeneity of the population, which may consist of a number of independent sub-
populations; (2) Inbreeding; (3) Selection against the heterozygotes; (4) The presence of a 'silent' allele which masks heterozygotes, making them indistinguishable from a homozygous type; and (5) Positive assortive mating, where individuals of like genotype mate with each other. It has been suggested that for most of the rare cases in which the Hardy-Weinberg equilibrium does not fit the observed data, this may be a result of a recent change in the constitution of a population, for example, the admixture of two different populations. As each Aboriginal population studied here, and those studied by Ranford et al. (1987), have relatively high frequencies of 'A 3' and 'A 4' alleles, it is unlikely that each has undergone a recent change where two populations have come together creating an imbalance of C4 alleles.

Analysis of HLA-extended haplotypes including HLA-A, B, and DR alleles, of donors from the East Cape York Aboriginal population, found a number had foreign admixture of haplotypes more commonly found in Caucasians (Gao et al., 1992a and b). However, the distribution of C4A allele frequencies was not affected by the exclusion of these donors from the population. It is thus unlikely that allelic heterogeneity from different ethnic groups exists in the East Cape York population. The most probable explanation for the discrepancy between expected and observed allele numbers when testing for Hardy-Weinberg equilibrium, is the problem associated with resolving proteins with the same or similar net charge by protein gel electrophoresis and immunofixation. This problem suggests that the traditional method of protein gel electrophoresis is unreliable, and that a more accurate technique for C4 allotyping is required. The discrepancy between observed and expected numbers of 'A 3' and 'A 4' alleles in the East Cape York Aboriginal population, was another reason which prompted further molecular research.

Even though the observed number of C4A alleles from the East Cape York Aboriginal population did not match those expected when testing for Hardy-Weinberg equilibrium, the estimated C4A gene frequencies for this population were not significantly different to those of the West Cape York population. Also, as found here, Ranford
et al. (1987), found high frequencies of both 'A 3' and 'A 4' alleles in two Aboriginal Australian populations. These C4A allele frequencies were consequently used for statistical analysis in studies for this thesis.

Ranford (1989), found populations from the western Pacific region could be grouped according to the distribution of C4A gene frequencies. The C4A gene locus was dominated by 'A 3' alleles, while 'A 4' and 'A Q0' alleles were the next most frequently occurring alleles. This distribution pattern is most prominent in the Micronesian and Melanesian populations where the most frequently occurring allele is nine times more frequent than the second most common allele (Ranford, 1989). In other western Pacific populations such as Japanese, Chinese, Korean, and Thai, the most frequent 'A 3' allele is usually three to four times more frequent than the 'A 4' and 'A Q0' alleles. Analysis of Aboriginal Australian and PNG populations here, also found that these populations, as for those described above, could be grouped according to the distribution of C4A gene frequencies.

The distribution of C4A alleles of different geographic PNG populations studied here and by Bhatia et al. (1991), were similar to that reported for Micronesian and Melanesian populations (Ranford, 1989). The coastal PNG population was the most similar except for having a lower frequency of C4A Q0 alleles than found for Micronesians and Melanesians. In the highland and lowland PNG populations, virtually all of the C4A alleles were 'A 3', whilst other C4A alleles had low frequencies.

Studies of Aboriginal Australians here, and by Ranford et al. (1987), found them to be quite different from other surrounding indigenous populations of the Pacific Islands, in that their most frequently occurring C4A allele occurs at only twice the frequency of the next most frequently occurring C4A allele. Slight differences in the frequency of 'A 3' and 'A 4' alleles between different Aboriginal Australian populations were found here, and by Ranford et al. (1987). These results show heterogeneity exists between different regionally located Aboriginal Australian populations. Thus, it may be
possible to distinguish between different geographically located Aboriginal Australian populations based on the frequency of 'A 3' and 'A 4' alleles, and in turn, distinguish these from other western Pacific indigenous populations. The difference in distribution of the most frequently occurring C4A alleles in Aboriginal populations is not unexpected, considering that many tribes have been isolated from each other by land distances similar to those sea distances separating the remote islands of the Pacific.

This study and the analysis of other western Pacific populations (Ranford, 1989), show that the C4B gene locus is dominated by 'B1', 'B 2', and 'B Q0' alleles, the frequencies of which vary between populations. However, it was difficult to group populations according to the distribution of the most frequently occurring C4B alleles. This is because this locus is more heterogeneous (with respect to variation of frequencies of the frequently occurring C4B alleles), than the C4A gene locus. Thus, analysis of the distribution of C4A allele frequencies appears to be a more reliable marker for studies of population origin and movement.

The estimated frequencies of C4A and C4B null alleles in Aboriginal populations in this study are relatively high in comparison to other C4 alleles in these populations, and similar to C4 null allele frequencies reported for many other western Pacific populations. The frequency of C4A null alleles in the West Cape York Aboriginal population was almost as high as that recorded by Ranford et al. (1987) for the northern Aboriginal Australian population. However, the lack of C4A 3,4 heterozygotes in the East and West Cape York Aboriginal populations could artificially inflate the expected number of C4A null alleles in these populations when tested for Hardy-Weinberg equilibrium. A technique based on optical density analysis may be used for the identification of C4 null alleles (Christiansen et al., 1983). This technique could be used to test the number of heterozygous null alleles in these studies which was estimated by computer analysis.

In this study, C4A Q0 allele frequencies in populations from PNG were low in comparison to other western Pacific populations
(Ranford, 1989). C4A Q0 alleles were not found in either the PNG highland or inland populations, whilst the coastal population had the lowest frequency of C4A Q0 alleles so far reported in the western Pacific. Bhatia et al. (1991), found a slightly higher frequency of C4A Q0 alleles in PNG populations which was nethertheless less than that detected in the Aboriginal Australians. At the C4B gene locus, the distribution of 'B Q0' alleles was variable between the PNG populations. The inland PNG population was devoid of C4B Q0 alleles, whilst the highland population had a frequency of C4B Q0 alleles similar to that detected in Melanesians, Chinese and Thai (Ranford et al., 1986). The coastal PNG population had a 'B Q0' frequency similar to that detected in Aboriginal Australians, Caucasians, and Japanese (summarised by Ranford, 1989). A Korean population is the only other indigenous population in the western Pacific where C4B null alleles were not detected, (Ranford et al., 1986). This finding suggests that, like other frequently occuring C4A alleles, the C4A Q0 rather than the C4B Q0 allele, is a more useful genetic marker for population studies.

A rare C4B allele, 'B 12', which might be considered as a genetic marker of Aboriginal Australian populations, was detected in low frequencies in both the East and West Cape York Aboriginal populations. This allele was also found in low frequencies in both the central and northern Aboriginal populations by Ranford et al. (1987). There appears to be some confusion concerning the nomenclature for this allele as Ranford incorrectly cited this allele as 'B 11' in her 1989 paper. To establish if this allele should be named 'B 11' or 'B 12', two of the three Aboriginal donors having a 'B 12' allotype in this study were sent to the Western Australian C4 standard reference typing laboratory (Department of Clinical Immunology, Royal Perth Hospital, Australia). Here, they were confirmed to be 'B 12' alleles. This allele has not been detected in other western Pacific populations which indicates that the 'B 12' allele might be a unique genetic marker for Aboriginal Australians. Bhatia et al. (1991), detected a rarely occurring C4B allele 'B 11' in PNG populations which migrates closely to the 'B 12' allele. These alleles could easily be misinterpreted as the same allele by protein gel electrophoresis, and so require confirmation by a standard
reference typing laboratory, as was sought for the Aboriginal 'B 12' allele. It is possible that the 'B 11' allele found in PNG populations (Bhatia et al., 1991) and the 'B 12' allele found in Aboriginal Australian populations by Ranford et al. (1987), and in this study, are identical. If true, this would be important as it would suggest a genetic link exists between Aboriginal Australians and PNG populations. However, occasionally the same C4B mutations have arisen independently in different populations (Ranford, 1989), and therefore do not necessarily represent close genetic links between populations. It is also possible that the 'B 11' and 'B 12' alleles comprise of a different set of polymorphic amino acid residues but produce the same migration patterns by protein gel electrophoresis. Significant differences of frequently occurring C4 allele frequencies between Aboriginal Australians and PNG populations, as well as absence of the 'B 12' allele from populations of PNG studied here, suggest that the 'B 11' specific polymorphism may have arisen independently in the PNG populations studied by Bhatia et al. (1991).

Other rare C4 alleles, thought to be unique to specific ethnic groups, have been found in other ethnic populations. The C4B 5 allele which is regarded as an Asian gene (Ranford, 1989), was detected here, for the first time, in the inland and coastal regions of PNG. The C4B 96 allele was previously detected only in Chinese (Ranford et al., 1986), and always on a duplicated haplotype (Dunkley and Hawkins, 1986). However, this allele was detected also for the first time in this study in the inland PNG population occurring with a low frequency. In this study, a greater array of alleles, including eight different C4A alleles and eight different C4B alleles, were found at the respective C4A and C4B gene loci in the PNG populations compared to that found in Aboriginal Australian populations. Despite the polymorphism found here in PNG, the majority of alleles detected in the PNG populations (especially inland and highland regions), were 'A 3' and 'B 1' alleles. The highland PNG population is the most monomorphic indigenous population so far recorded in the western Pacific. The Aboriginal Australians showed less polymorphism of C4A and C4B alleles. However, the distribution of the frequency of C4A and C4B alleles
was more varied having a number of frequently occurring alleles. By comparison to other western Pacific populations the distribution of frequency of the most frequently occurring \( C4 \) alleles in Aboriginal Australian populations is most varied. Studies here, detected the highest degree of polymorphism in the coastal region of PNG as did Bhatia et al. (1991). This high degree of polymorphism in the coastal region of PNG is probably a reflection of the interaction of these people with other visiting indigenous people. On the other hand, those situated in the inland and highland regions have been more isolated, and therefore less subjected to genetic admixture with other populations.

Linkage disequilibrium analysis of \( C4 \) allotypes in the Aboriginal Australian and PNG populations reflects the frequency of the most frequently occurring \( C4 \) alleles in each population. By comparison with the Aboriginal Australian populations, \( 'A \ 3' \) and \( 'B \ 1' \) alleles are in strong linkage disequilibrium in both the coastal and inland PNG populations. This linkage disequilibrium is weaker in the highland PNG populations, as many \( 'A \ 3' \) alleles are associated with other \( C4B \) alleles besides the \( 'B \ 1' \) allele. Bhatia et al. (1991), found linkage disequilibrium between \( 'A \ 3' \) and \( 'B \ 1' \) alleles, and \( 'A \ 4' \) and \( 'B \ 2' \) alleles in the lowland regions but not in the highland region of PNG. Linkage disequilibrium of \( 'A \ 4' \) and \( 'B \ 2' \) alleles was found here in each Aboriginal Australian population, which is stronger than that found between \( 'A \ 3' \) and \( 'B \ 1' \) alleles in these populations. This finding is in agreement with linkage disequilibrium analysis of central and northern Aboriginal Australians by Ranford et al. (1987). Results of linkage disequilibrium analysis conducted here support the hypothesis that different ethnic groups have linkage disequilibrium of different frequently occurring \( C4 \) alleles, which provides another means for population studies at the molecular level.

Comparisons of \( C4 \) gene frequencies and allotype distributions between different Aboriginal Australian and PNG populations indicate that the two Aboriginal populations are more closely related to each other, as are the three PNG populations. The coastal population of PNG is most different from the Aboriginal Australians.
An important factor to consider is that analysis of gene frequency compares only one C4A or C4B gene at a time. The distribution of C4 allotypes between populations, on the other hand, compares C4 allotypes carrying both C4A and C4B alleles, as well as comparing C4 allotypes carrying a specific C4A or C4B allele. Even though C4 gene frequency analysis and C4 allotype frequency analysis analyse the distribution of C4 alleles between populations by a different approach, this study found the results from these two types of analyses were in agreement with each other. Gene frequency and allotype frequency analyses reflect differences in the number of C4 alleles with relatively high frequencies.

Genetic distance analysis may provide the best indication of the genetic closeness between two or more populations as it pools together the C4A and C4B gene frequency data and determines the extent to which one population is different from another. However, to obtain the most accurate indication of the genetic distance between Aboriginal Australian and PNG populations other indigenous populations from surrounding countries would need to be included. Genetic distance analysis in these studies came to a different conclusion in comparison to those results obtained by gene frequency and allotype frequency analyses. Genetic distance analysis found the inland PNG population was most genetically distant to the Aboriginal populations, which is not surprising considering the inland PNG population is almost monomorphic at both the C4A and C4B gene loci.

Serological studies by Kirk (1972) and complement components C2 and factor B studies by Ranford (1989), suggest that northern Aboriginal Australians are more similar to highland PNG populations. However, studies of the C4 gene locus in Aboriginal Australians and PNG populations here and by Ranford et al. (1987), and Bhatia et al. (1991), suggest that the Aboriginal Australians are distinctly different from PNG populations. The occurrence of two more frequently occurring alleles at the C4A and C4B gene loci indicates that the Aboriginal Australians may have originated from two different populations. One population would have a high frequency of
the 'A 3 B 1' allotype, whilst the other would have a high frequency of the 'A 4 B 2' allotype. The Aboriginal Australian 'A 3 B 1' allotype may have originated from any of the surrounding Pacific island indigenous populations, as these alleles are common to all Pacific island populations. The Aboriginal Australian 'A 4 B 2' allotype would more likely to have arisen from one of the Asian populations which have relatively high frequencies of these alleles, rather than from a PNG population. In PNG the 'A 4' allele is virtually absent, and the 'B 2' allele only has a low frequency.

Western Pacific populations can be distinguished from each other according to the differences in frequency of the most frequently occuring C4A alleles 'A 3', 'A 4', and 'A Q0'. PNG is distinct from any other western Pacific population, as the A 3 allele is the most frequently occuring C4A allele, whilst all other C4A alleles have very low frequencies. Micronesians and Melanesians are similar to Papuans, except that they have slightly lower frequencies of 'A 3', and higher frequencies of 'A 4' and 'A Q0' alleles. Japanese, Chinese, Korean and Thai, on the other hand, are distinct from other populations mentioned above, by reason of their lower frequency of 'A 3'. Also, the frequency of 'A 3' in Japanese, Chinese, Korean and Thai populations is three to four times higher than the frequencies of 'A 4' and 'A Q0' in these populations. Aboriginal Australians are distinct from other indigenous western Pacific populations with respect to the most frequently occuring C4A alleles, 'A 3' and 'A 4', which occur with almost equally high frequencies. Caucasians can be distinguished from other populations, as their 'A 3' allele occurs four times more frequently than the 'A Q0' allele, which in turn occurs twice as frequently as the 'A 4' allele (Ranford et al., 1986).

This study has identified individuals with C4A and C4B null alleles, as well as rare C4A and C4B alleles, such as the 'B 12' allele, which will be useful for further studies at the molecular level. C4 allotypes identified here may also be useful for subsequent correlation with DNA analysis aimed at developing a molecular C4 allotyping protocol. The design of a molecular-based C4 allotyping protocol, and the application of this molecular approach are dealt with in the following chapters.
Chapter 4
Molecular-based C4 allotyping protocol

4.1 Introduction

Accurate and rapid identification of C4 allotypes is confined to a limited number of specialised laboratories. The currently used C4 allotyping technique of protein gel electrophoresis has limitations in that it requires highly skilled interpretation of C4 allotypes and careful handling of C4 serum samples as this protein degrades rapidly.

Many techniques have been used to distinguish between C4A and C4B allotypes which are based on detecting differences of the polymorphic C4d gene region. These include, differential weight of C4A and C4B α-chains (Roos et al., 1982; Carroll et al., 1989; Rittner and Stradmann-Bellinghausen, 1990); haemolytic activity of C4A and C4B protein (Kaufmann et al., 1990); immunoblotting of C4 using monoclonal and polyclonal antibodies (Doxiadis and Grosse-Wilde, 1990); pulse-field gel electrophoresis (Dunham et al., 1990); Rodgers and Chido typing (Yu et al., 1988; Giles, 1990); RFLP analysis (Schneider, 1990); C4 protein gel electrophoresis (Awdeh and Alper, 1980), and DNA sequencing (Belt et al., 1984 and 1985; Yu et al., 1986; Anderson et al., 1992; Paz-Artal et al., 1993). As reported at the VI Complement Genetics Workshop in Germany 1989, researchers in the complement field used these protocols to reclassify the majority of known C4 allotypes, and as a result the C4 nomenclature was revised. Even though these methods have collectively demonstrated many features of C4 polymorphism, each technique is limited with respect to the amount of information it can provide about a particular C4 allele.

Of the techniques mentioned above, C4 protein gel electrophoresis followed by immunofixation has remained the major technique for C4 allotyping. This method separates proteins according to their size and overall net charge (described in detail in Chapter 2, section 2.5). The sensitivity of this technique has improved since its first application by Awdeh and Alper (1980). Introduction of
Carboxypeptidase B enables slightly degraded C4 samples to be allotyped, and allows resolution of overlapping alleles (Sim and Cross, 1986), and haemolytic overlays are used to identify co-migrating C4A and C4B alleles (Gaither et al., 1974; Kaufmann et al., 1990). Consequently, sub-types of the more frequently occurring C4 alleles have been identified such as the B 12 allele, which migrates closely to B 1 and B 2 alleles (refer to chapter 3 and Mauff et al., 1983).

Despite these improvements, some C4 allotypes remain difficult to identify by this procedure due to overlapping C4 alleles by protein gel electrophoresis (Sim and Cross, 1986). Gene duplication, which occurs frequently at the C4 gene locus, increases the difficulty of C4 allotype interpretation, as even more protein bands are revealed by protein gel electrophoresis. Also, sequencing has revealed sub-types of C4 alleles identified by protein gel electrophoresis, indicating that not all C4 alleles can be detected by the latter (Belt et al., 1985; Yu, 1991; Anderson et al., 1992).

Modern methods of molecular analysis have identified the molecular basis for much of the polymorphism which is characteristic of C4 alleles by sequencing the polymorphic C4d gene region. An allotyping protocol based on known polymorphic nucleotide bases for each C4 allele would be the most accurate means of identification, and would obviate the problems that exist with conventional techniques. Accurate identification of C4 alleles is essential for the advancement of studies in associated diseases, and will facilitate reproducible results by researchers working in this region throughout the world. A molecular-based protocol has the advantage of being easily adopted by any molecular laboratory and routinely used in disease diagnosis without a high degree of difficulty. It also has the advantage of not being reliant on handling of C4 protein which is prone to degradation in transit and storage.
4.2 **Aim**

The main aim of research reported in this chapter was to establish a molecular-based C4 typing protocol based on PCR-RFLP analysis by utilising known published sequences of C4 polymorphic alleles and new polymorphic nucleotide bases discovered in this study. This protocol was then used to identify C4 alleles which could not be identified by C4 protein gel electrophoresis (chapter 3), and to determine the molecular basis of other rarely occurring C4 alleles.

4.3 **Materials and Methods**

4.3.1 **C4 nomenclature**

C4 allotypes as determined by protein gel electrophoresis and allotypes as determined by DNA sequencing, are distinguished in this thesis by, for example, A 3 and A *3, respectively.

4.3.2 **Samples for C4 PCR-RFLP allotyping**

Nineteen Aboriginal Australian samples from Darwin (H) and East and West Cape York (ECY and WCY), two IDDM patients from Canberra (ASD) and Melbourne (RCH), three European workshop cell lines (EW), and three Caucasians from JCSMR (JC), were used in this study. Specific details of these samples are given in Chapter 2, section 2.3 and 2.4.

4.3.3 **C4 protein gel electrophoresis and immunofixation**

Plasma samples of Aboriginal Australians from Cape York, and Caucasians from JCSMR, were C4 allotyped by C4 protein gel electrophoresis followed by immunofixation, using the protocol outlined in section 2.5. Other samples from IDDM donors (RCH and ASD) were typed by Pam Ranford, JCSMR, and Darwin Aboriginal samples (H) were typed by Heather Dunckley, JCSMR, by using the same method of protein gel electrophoresis.
4.3.4 Genomic DNA extraction

Genomic DNA extraction from buffy coats and cell lines was performed using the technique described in section 2.7.

4.3.5 Taq I restriction analysis of genomic DNA, and C4 specific hybridisation

Approximately 3 μg of genomic DNA was restricted in a total volume of 50 μl, with 30 units of Taq I restriction enzyme, and 1x reaction buffer (provided by the enzyme manufacturer), at 65 °C for 2 hours. Reactions were stopped by adding 5 μl of 0.5 M EDTA, and heating to 65 °C for 10 min. Restricted DNA products were precipitated by adding 25 μl of 7.5 M ammonium acetate and 160 μl of absolute ethanol. Samples were dried and resuspended in 10 μl of distilled water and 4 μl of electrophoresis buffer, and then electrophoresed as described in section 2.9. Samples were transferred to a nylon membrane following the protocol in section 2.11. C4 pAT-A, a full length C4-specific probe, was radioactively labelled and hybridised to Taq I restricted DNA fragments following the methods described in sections 2.12.1 and 2.12.2.

4.3.6 Primer design for PCR amplification of polymorphic C4 gene regions

Primers were designed to amplify nucleotide base regions which encode polymorphic amino acid residues within the β-chain, α-chain and C4d region. Where possible, entire exons were amplified to ensure sequencing of the complete coding region. Primers designed to amplify a region for sequencing, contained an enzyme site at the 5' end which was accompanied by a few additional nucleotide bases (3-5) to enable binding of the enzyme. Primers were tested for primer self-hybridisation, sense and antisense primer hybridisation, and nucleotide 'G' and 'C' base content by using an IBM computer programme called 'PCR'. All primers were synthesised by the biomolecular resource facility at JCSMR, ANU, Canberra. Primer pairs, and polymorphic amino acid residues encoded by nucleotide bases
which were targeted for PCR amplification are listed in Table 4.1, and their location within the α- and β- regions of the C4 gene are shown in Figure 4.1.

4.3.7 PCR amplification of nucleotide bases encoding polymorphic amino acid residues

Polymorphic amino acid residues, their position within the C4 amino acid sequence, and PCR amplification specificity for C4A, C4B, and long or short C4 gene type, are listed in Table 4.2. PCR amplification of polymorphic gene regions followed the general description described in section 2.16. Nucleotide bases encoding amino acid residues #707 and #1267 were amplified using C4A and C4B isotypic primers (see primer pairs 4, 5, 12 and 13). However, the resulting PCR fragment sizes were too large (2703 bp for primer pairs 4 and 5, and 985 bp for primer pairs 12 and 13), and therefore contained too many restriction sites within the sequence to enable easy interpretation of PCR-RFLP banding patterns. As a result, samples were subjected to a second round of amplification using nested primers (pairs 6 and 14 for nucleotides encoding amino acid residue #707 and #1267 respectively), which were situated closer to the polymorphic nucleotide base of interest. This approach resulted in smaller PCR products (212 bp and 238 bp for amplification of nucleotide bases encoding residue #707 and #1267, respectively), containing less restriction sites. Specific PCR reaction conditions for each primer pair, and the resulting PCR product size, are given in Table 4.3. Detection of PCR amplification products was determined by gel electrophoresis as described in sections 2.9 and 2.10.

4.3.8 PCR-RFLP analysis of nucleotide bases encoding C4 polymorphic amino acid residues

Restriction endonuclease digestion conditions for each PCR product (approximately 300 ng), were as described by the manufacturer of each enzyme, with the exception of Aci I. Digestion with this enzyme was optimised by the addition of Buffer #4 (provided by the manufacturer), 2 µl of 1 % gelatin, and 5 units of enzyme in a total
Table 4.1  

<table>
<thead>
<tr>
<th>sense primer</th>
<th>antisense primer</th>
<th>amino acid residue #</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 5'-CAGCTGGTGAAATGGACAGAGC-3'</td>
<td>5'-CCATCTCCCCACCTGCAAGAC-3'</td>
<td>306, 309, 328</td>
</tr>
<tr>
<td>2 5'-CAGCTGGTGAAATGGACAGAGC-3'</td>
<td>5'-GAGGACAGAGGTTAGTCAG-3'</td>
<td>306, 309, 328</td>
</tr>
<tr>
<td>3 5'-CTCCAGGTATCTGCAGGCTCC-3'</td>
<td>5'-CAGTCTCATGCAACCCATGTAG-3'</td>
<td>458</td>
</tr>
<tr>
<td>4 5'-GCAGTGGGTCAGTATCTCC-3'</td>
<td>5'-CACGTCCACACCTGCTCTTGCTCTACAC-3'C4A</td>
<td>707</td>
</tr>
<tr>
<td>5 5'-GCAGTGGGTCAGTATCTCC-3'</td>
<td>5'-CACGTCCACACCTGCTCTTGCTCTATCAC-3'C4B</td>
<td>1054</td>
</tr>
<tr>
<td>6 5'-GCAGTGGGCAGGTGATCTCC-3'</td>
<td>5'-AGCTATGCCAGAATTTCCCCCTACCTCCAGTCTCCT-3'</td>
<td>1101-1106</td>
</tr>
<tr>
<td>7 5'-CTAGAATTCTCTCCAAGGCTACATGC-3'</td>
<td>5'-AGCTATGCCAGAATTTCCCCCTACCTCCAGTCTCCT-3'</td>
<td>1157, 1182</td>
</tr>
<tr>
<td>8 5'-CTAGAATTCTCTCCAAGGCTACATGC-3'</td>
<td>5'-AGCTATGCCAGAATTTCCCCCTACCTCCAGTCTCCT-3'</td>
<td>1186, 1188</td>
</tr>
<tr>
<td>9 5'-GACGGGCCAGCTGACGAGCAGCGCTGACGGCTCG-3'</td>
<td>5'-CTACTTGGGTTAGCTGCAGAATC-3'</td>
<td>1191</td>
</tr>
<tr>
<td>10 5'-GACGGGCCAGCTGACGAGCAGCGCTGACGGCTCG-3'</td>
<td>5'-CTACTTGGGTTAGCTGCAGAATC-3'</td>
<td>1267</td>
</tr>
<tr>
<td>11 5'-GACGGGCCAGCTGACGAGCAGCGCTGACGGCTCG-3'</td>
<td>5'-CTACTTGGGTTAGCTGCAGAATC-3'</td>
<td>1267</td>
</tr>
<tr>
<td>12 5'-GACGGGCCAGCTGACGAGCAGCGCTGACGGCTCG-3'</td>
<td>5'-CTACTTGGGTTAGCTGCAGAATC-3'</td>
<td>1267</td>
</tr>
<tr>
<td>13 5'-GACGGGCCAGCTGACGAGCAGCGCTGACGGCTCG-3'</td>
<td>5'-CTACTTGGGTTAGCTGCAGAATC-3'</td>
<td>1267</td>
</tr>
<tr>
<td>14 5'-GACGGGCCAGCTGACGAGCAGCGCTGACGGCTCG-3'</td>
<td>5'-CTACTTGGGTTAGCTGCAGAATC-3'</td>
<td>1267</td>
</tr>
</tbody>
</table>

Primer pair 1 amplified the short C4 genes, whereas primer pair 2 amplified both the long and short C4 genes simultaneously. Analysis of nucleotide bases encoding amino acid residue #707 and #1267 involved two rounds of PCR. The first, where primer pairs 4 and 12 amplified from the C4A genes, and 5 and 13 amplified from the C4B genes. The second, involved amplification from those C4A and C4B specific PCR products by using internal primers 6 and 14. Primer pair 3 and 9 amplified from both C4A and C4B genes. Those primers denoted with C4A or C4B are C4 isotypic primers.
Figure 4.1 Schematic representation of the molecular based $C_4$ typing protocol based on PCR-RFLP
Table 4.2 *C4* polymorphic amino acid residues, their isotypic-specificity and *C4* gene size specificity

<table>
<thead>
<tr>
<th>amino acid #</th>
<th>amino acid polymorphism</th>
<th><em>C4A/C4B</em> specific</th>
<th><em>C4A/C4B</em> non-specific</th>
<th><em>C4 16 kb</em></th>
<th><em>C4 16 kb &amp; 23 kb</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>306</td>
<td>Lys (K)/Met (M)</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>309</td>
<td>Ile (I)/Met (M)</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>328</td>
<td>Tyr (Y)/Ser (S)</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>458</td>
<td>Arg (R)/Trp (W)</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>707</td>
<td>Leu (L)/Pro (P)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1054</td>
<td>Asp (D)/Gly (G)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1101</td>
<td>Pro (P)/Leu (L)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1102</td>
<td>Cys (C)/Ser (S)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1105</td>
<td>Leu (L)/Ile (I)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1106</td>
<td>Asp (D)/His (H)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1157</td>
<td>Ser (S)/Asn (N)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1182</td>
<td>Tyr (Y)/Ser (S)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1186</td>
<td>Ala (A)/Ala (A)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1188</td>
<td>Ala (A)/Val (V)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1191</td>
<td>Arg (R)/Leu (L)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1267</td>
<td>Ala (A)/Ser (S)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For amino acid residue #1186 the nucleotide bases are polymorphic encoding 'gcc' in *C4A* alleles, and 'gcc' in *C4B* alleles. Some PCR reactions amplified the short *C4* alleles (16 kb) in one product, and short and long *C4* alleles (16 kb and 23 kb) in another product. Amplification of nucleotide bases encoding amino acid #458 is neither *C4A* nor *C4B* specific, nor *C4* gene size specific.
<table>
<thead>
<tr>
<th>primer pair</th>
<th>annealing temperature (°C)</th>
<th>denaturation (sec)</th>
<th>annealing (sec)</th>
<th>extension (sec)</th>
<th>PCR (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>58</td>
<td>60</td>
<td>60</td>
<td>120</td>
<td>565</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>120</td>
<td>204</td>
</tr>
<tr>
<td>3</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>203</td>
</tr>
<tr>
<td>4&amp;5</td>
<td>58</td>
<td>90</td>
<td>90</td>
<td>300</td>
<td>2703</td>
</tr>
<tr>
<td>6</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>212</td>
</tr>
<tr>
<td>7&amp;8</td>
<td>56</td>
<td>60</td>
<td>60</td>
<td>120</td>
<td>422</td>
</tr>
<tr>
<td>9</td>
<td>56</td>
<td>60</td>
<td>60</td>
<td>120</td>
<td>588</td>
</tr>
<tr>
<td>10&amp;11</td>
<td>58</td>
<td>60</td>
<td>60</td>
<td>120</td>
<td>561</td>
</tr>
<tr>
<td>12&amp;13</td>
<td>57</td>
<td>70</td>
<td>70</td>
<td>220</td>
<td>985</td>
</tr>
<tr>
<td>14</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>90</td>
<td>238</td>
</tr>
</tbody>
</table>

For each PCR reaction samples were denatured at 96 °C, and the extension temperature was 72 °C. All PCR reactions were subjected to 35 cycles except for primer pairs 6 and 14 where 2 μl of total reaction volume 100 μl were reamplified using an internal primer for 25 cycles.
reaction volume of 15 µl. Restriction endonucleases, digestion conditions, and electrophoresis parameters are listed in Table 4.4. PCR-RFLP products were separated by agarose and polyacrylamide gel electrophoresis, and detected by gel staining as described in sections 2.9 and 2.10.

4.3.9 Cloning and sequencing of nucleotide bases encoding amino acid residues #286 to #330 in short C4 alleles

PCR products from amplification of nucleotide bases encoding amino acid residues #286 to #330 in short C4 genes, were sequenced for Aboriginal donors H14, ECY178, WCY1158, ECY028 and Caucasian RCH517. All samples were blunt-end cloned according to the method described in section 2.18.2. Recombinant clones were sequenced by automated sequencing as outlined in section 2.19. A minimum of three clones of the sense strand, and three clones of the antisense strand of DNA were sequenced for H14, WCY1158, ECY028 and RCH517. The short C4 gene PCR product of donor ECY178 (approximately 1 µg), was subjected to restriction analysis with enzyme Ple I. The restricted fragments were separated on a 2.5 % agarose gel by electrophoresis as described in section 2.9. Restriction fragments encoding nucleotide bases for amino acid residue #286 to intron 9 nucleotide base #102 (total of 237 bp), were eluted from the gel and concentrated to a total volume of 10 µl, as described in section 2.17 Chapter 2, prior to cloning and sequencing.

4.3.10 Allele specific oligonucleotide (ASO) hybridisation to nucleotides encoding polymorphic amino acid residue #1182

4.3.10.1 Samples for ASO typing

C4A alleles were selectively amplified from 72 donors chosen at random, from the Cape York Aboriginal Australian populations used for analysis in Chapter 3. In addition, Caucasians having homozygous C4 allotypes A 3 (JC002), and A 4 (ASD97), as well as heterozygous allotypes A 3,4 (ASD49), as determined by C4 protein gel electrophoresis, were chosen as controls.
Table 4.4 Experimental details for PCR-RFLP analysis of C4 nucleotide bases encoding polymorphic amino acid residues

<table>
<thead>
<tr>
<th>amino acid #</th>
<th>primer pair</th>
<th>restriction endonuclease</th>
<th>% and type of gel</th>
<th>electrophoresis conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>306</td>
<td>1 &amp; 2</td>
<td><em>Alu I</em></td>
<td>8% polyacrylamide</td>
<td>35mA, 2 hours</td>
</tr>
<tr>
<td>328</td>
<td>1 &amp; 2</td>
<td><em>Ple I</em></td>
<td>8% polyacrylamide</td>
<td>35mA, 2 hours</td>
</tr>
<tr>
<td>458</td>
<td>3</td>
<td><em>Fnu4HI</em></td>
<td>8% polyacrylamide</td>
<td>35mA, 3 hours 30min</td>
</tr>
<tr>
<td>707</td>
<td>6</td>
<td><em>Pvu II</em></td>
<td>12% polyacrylamide</td>
<td>40mA, 3 hours 30min</td>
</tr>
<tr>
<td>1054</td>
<td>7 &amp; 8</td>
<td><em>Fnu4HI</em></td>
<td>15% polyacrylamide</td>
<td>40mA, 1 hour 30min</td>
</tr>
<tr>
<td>1101-1106</td>
<td>9</td>
<td><em>Mnl I</em></td>
<td>12% polyacrylamide</td>
<td>40mA, 3 hours</td>
</tr>
<tr>
<td>1157</td>
<td>10 &amp; 11</td>
<td><em>Alu I</em></td>
<td>8% polyacrylamide</td>
<td>35mA, 2 hours</td>
</tr>
<tr>
<td>1186</td>
<td>10 &amp; 11</td>
<td><em>Hha I</em></td>
<td>2.5% agarose</td>
<td>70 volts, 6 hours</td>
</tr>
<tr>
<td>1188</td>
<td>10 &amp; 11</td>
<td><em>Aci I</em></td>
<td>2.5% agarose</td>
<td>70 volts, 6 hours</td>
</tr>
<tr>
<td>1191</td>
<td>10 &amp; 11</td>
<td><em>Aci I</em></td>
<td>8% polyacrylamide</td>
<td>35mA, 2 hours</td>
</tr>
<tr>
<td>1267</td>
<td>14</td>
<td><em>Aci I</em></td>
<td>8% polyacrylamide</td>
<td>35mA, 2 hours</td>
</tr>
</tbody>
</table>

DNA fragments having C4 polymorphic amino acid residues were amplified with specific primer pairs, and digested with the appropriate restriction endonucleases.
4.3.10.2 PCR amplification of nucleotides encoding amino acid residue #1182 and Southern blotting

The $C4$ gene region encoding amino acid #1182 was amplified with the $C4A$ isotypic primer pair 10 (Table 4.1), using PCR conditions set out in Table 4.3. PCR fragments were visualised by gel electrophoresis as described in sections 2.9, and 2.10, and transferred to a nylon membrane by Southern blotting, as described in section 2.11.

4.3.10.3 End-labelling of allele-specific oligonucleotides (ASO), and hybridisation to membranes

Two ASO's were used for analysis of nucleotide bases encoding the amino acid residue #1182; namely 5'-ATGCCCTGACACTGACCAA-3', which detects nucleotide bases encoding amino acid 'Thr', and 5'-ATGCCCTGTCACTGACCAA-3', which detects nucleotide bases encoding amino acid 'Ser'. Both ASO's extend from nucleotide bases encoding amino acid residues #1180 to #1183. End-labelling and hybridisation of oligonucleotides to PCR products were as described in sections 2.12.3 and 2.12.4.

4.3.11 Sequencing of $C4A^{*}3$ and $C4A^{*}4$ alleles to confirm ASO hybridisation to nucleotides encoding amino acid residue #1182

4.3.11.1 Donor samples

Donor samples used for this study were Aboriginal Australian samples ECY743 ($C4A$ 3 allele), and ECY519 ($C4A$ 4 allele), whose $C4$ allotypes were detected by $C4$ protein gel electrophoresis, as described in section 2.5.
4.3.11.2 PCR amplification of the gene region surrounding nucleotide bases which encode amino acid residue #1182

Primers used to amplify the $C_4$ gene region encoding amino acid residues #1091 to #1206 of donors ECY743 and ECY519 were sense strand primer 5'-GACGGCCAGTGTCGACAGCAGGCTGACGGCTCG-3', which has nucleotide bases encoding amino acid residues #1091 to #1097 (same as sense primer in pair 9, Table 4.1), and antisense primer 5'-AGCTATGACCGAATTCCCCCCTACCTCCAGTCTCCT-3', which has nucleotide bases for amino acid residues #1203 to #1206 (same as antisense primer in pair 9, Table 4.1).

A second set of primers were used to amplify the $C_4$ gene region encoding amino acid residues #1091 to #1189, of samples ECY743 and ECY519. These primers were sense strand primer 5'-GACGGGCC AGTGTGACAGCAGGCTGACGGCTCG-3' encoding amino acid residues #1091 to #1097 (same as sense primer in pair 9, Table 4.1), and antisense primer 5'-GTCGAATTCGTCCGCAGGCGCCTTGGTC-3' encoding amino acid residues #1184 to #1189.

Samples were subjected to PCR as described in section 2.16, with specific conditions for each primer set as follows: The first set of primers which were designed to amplify nucleotide bases encoding amino acid residues #1091 to #1206, were amplified by 35 cycles of 96 °C for 60 sec, 56 °C for 60 sec, and 72 °C for 120 sec. The second set of primers used to amplify nucleotide bases encoding amino acid residues from #1091 to #1189, were amplified using the same conditions as the first set of primers, however the annealing temperature was reduced to 49 °C. PCR products were visualised by gel electrophoresis and EtBr staining as described in section 2.9 and 2.10.

4.3.11.3 Cloning of PCR products and automated sequencing

PCR products having nucleotide bases which encode amino acid residues #1091 to #1206, were cloned into M13 using the blunt-end method outlined in section 2.18.2 to 2.18.7, whereas PCR products having nucleotides which encode residues #1091 to #1189, were
cloned into M13 using sticky-end cloning (see section 2.18.1, and 2.18.3 to 2.18.7). Four M13 blunt-end clones from ECY743, and two M13 blunt-end clones from ECY 519, were chosen for sequencing. Three M13 sticky-end clones from ECY 743, and three M13 sticky-end clones from ECY 519, were chosen for sequencing. Samples were sequenced by automated sequencing as described in section 2.19.

4.3.12 Predicted Isoelectric Point (PI) of derived protein sequences for each C4 allele

Computer analysis of the derived protein sequence was used to determine the predicted isoelectric point (PI) of each C4 allele. Details concerning this computer programme are given in section 2.20 (Chapter 2).

4.4 Results

Results of this chapter are reported in three major sections. The first section deals with the approach taken to design a C4 molecular allotyping protocol based on PCR-RFLP, using published information describing the molecular genetics of C4. The second section records the amino acid residues encoded by polymorphic nucleotide bases of C4 alleles within each donor. The final section details the molecular basis of each C4 allele including some alleles which were not accurately identified by C4 protein gel electrophoresis. This was carried out by combining and analysing the data from C4 protein gel electrophoresis (see Chapter 3), PCR-RFLP, Taq I RFLP analysis, and the predicted isoelectric points (PI), which were determined by using a protein analysis computer programme. An account of the approach taken, and consideration of results and those of other workers are necessary to relate here as an integral part of the development of the C4 allotyping protocol.

4.4.1 Design of the C4 PCR-RFLP molecular allotyping protocol

A major factor distinguishing C4A from C4B alleles is a set of closely situated polymorphic amino acid residues encoded by
nucleotide bases in exon 26 of the C4d region. These amino acid residues, 'Pro,Cys-Leu,Asp' for C4A genes, and 'Leu,Ser-Ile,His' for C4B genes, have become known as the C4 isotypic or class-specific amino acid residues (Palsdottir et al., 1987a; Yu, 1991; Anderson et al., 1992). The C4, PCR-RFLP typing protocol was designed using C4A and C4B isotypic primers for the amplification of other nucleotide bases encoding polymorphic amino acid residues in each allele. This approach permitted C4A and C4B alleles to be typed individually.

Only polymorphic amino acid residues previously detected by nucleotide base sequencing were chosen for these studies, as they were thought to be more reliable than those detected by protein sequencing. Evidence elsewhere supported this approach. For example, the polymorphic amino acid residue #708 reported by Moon et al. (1981), may be a protein sequencing artifact as other more recent studies (Yu, 1991), did not detect such polymorphism.

A schematic representation of the molecular-based C4 typing protocol based on PCR-RFLP is shown in Figure 4.1. C4 polymorphic amino acid residues #306 (identified in these studies), and residue #328, located in exon 9 of the β-chain, were situated too far from the C4 isotypic region (5 kb in short C4 genes and 12 kb in long C4 genes), to implement the strategy of only C4A or C4B isotypic amplification. Knowing that the majority of C4A alleles reported are of the long (23 kb) type, whereas both long and short (16 kb) C4B alleles have been found (Schneider, 1990), an alternative approach to characterise amino acid residues #306 and #328, by utilising the differential size of intron 9 was designed. Two primer pairs were used, one which amplified only short C4 genes from the gene region encoding amino acid residues #286 to #330. The second primer pair amplified nucleotide bases encoding amino acid residues #286 to #329, such that both long and short C4 genes were amplified. Those amino acid residues detected in the short C4 gene were assigned to C4B alleles. Residues detected by the second set of primers were compared to those detected in the short C4 genes. Those which were different were assigned to either C4A or C4B alleles depending on: (1). The allotype identified by C4 protein gel electrophoresis; (2).
Other \(C4\) alleles of the donor, and (3). PI values obtained using the protein analysis computer programme. It is important to note that even though \(C4\) protein migration patterns are derived differently by isoelectric focusing and gel electrophoresis, they will ultimately result in the same pattern of migration, as both techniques separate the proteins according to their size and net charge. This thesis often refers to predicted isoelectric point (PI) (according to computer analysis, using the programme 'isoelectric'), and migration patterns resulting from \(C4\) protein gel electrophoresis. It is the pattern resulting from migration of the subject protein in relation to other migrating \(C4\) proteins which is important to this study, and not the technical procedures used to obtain these patterns of migration.

Polymorphic amino acid residue #458 was the only other residue situated too far from the \(C4\) isotypic region to allow \(C4A\) and \(C4B\) specific amplification. As a result, PCR-RFLP analysis of nucleotide bases encoding amino acid residue #458 was performed on PCR product containing both \(C4A\) and \(C4B\) alleles.

4.4.2 Amino acid residues encoded by polymorphic nucleotide bases as determined by PCR-RFLP analysis

4.4.2.1 Amino acid residue #328

Nucleotide bases encoding amino acid residue #328 in both long (23 kb) and short (16 kb) \(C4\) genes, were identified by PCR-RFLP analysis using the restriction enzyme 'Ple I'. PCR-RFLP banding patterns resulting from restriction analysis of nucleotide bases encoding amino acid residue #328 are shown in Figure 4.2. Polymorphic amino acid residues for each Aboriginal and Caucasian donor, and their \(C4\) allotypes, according to \(C4\) protein gel electrophoresis, are listed in Tables 4.5 and 4.6, respectively.

PCR analysis showed that the majority of donors possessed both long and short \(C4\) genes. Short \(C4\) genes were detected by successful PCR amplification of DNA using PCR primers which were specific for short \(C4\) genes. Long \(C4\) genes were detected by
Figure 4.2 PCR-RFLP banding patterns of nucleotide bases encoding polymorphic amino acid residues #306 and #328 from short (16 kb) and long (23 kb)/short (16 kb) C4 PCR fragments. Examples of each banding pattern are listed as lane numbers for the respective amino acid residue. Where homozygote examples of each banding pattern could not be shown, a heterozygous example was referred to.

Molecular weight markers consisted of pBR322 restricted with *Msp I* in lane 10 and 13 for gels corresponding to amino acid residue #306, and lane 10 of the gel corresponding to amino acid residue #328. Resulting fragment sizes (bp) of pBR322 *Msp I* shown in each photograph include 67, 76, 90, 110, 123, 147, 160, 180, 190, 201, 217, 238, 242, 309, 404, 527, and 622 base pair fragments.

Unrestricted PCR DNA fragments are shown in lanes 11 and 12 of gels corresponding to amino acid residue #306, and lane 11 for amino acid residue #328. Photographs of gels where PCR fragments are black represent gels stained with silver, and gels where fragments are white were stained with EtBr. Small PCR fragments 16 bp or less had migrated from the gel during electrophoresis and were therefore not visible.
<table>
<thead>
<tr>
<th>Amino Acid Residue #</th>
<th>Short C4 Gene (16kb)</th>
<th>Long and Short C4 Gene (16kb and 23kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Alu 1 PCR-RFLP (bp)</td>
<td>Lane # Example</td>
</tr>
<tr>
<td>306</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>3+118+185+259</td>
<td>1,2</td>
</tr>
<tr>
<td>K</td>
<td>3+63+118+122+259</td>
<td>7,8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Amino Acid Residue #</th>
<th>Short C4 Gene (16kb)</th>
<th>Long and Short C4 Gene (16kb and 23kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ple I PCR-RFLP (bp)</td>
<td>Lane # Example</td>
</tr>
<tr>
<td>328</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>57+237+271</td>
<td>12,13</td>
</tr>
<tr>
<td>S</td>
<td>57+102+135+271</td>
<td>18,19</td>
</tr>
</tbody>
</table>
Table 4.5 β-chain, exon 9 polymorphic amino acid residues in Aboriginal Australians

<table>
<thead>
<tr>
<th>Donor</th>
<th>C4 allotype#</th>
<th>C4 allotype*</th>
<th>amino acid # 306</th>
<th>amino acid # 328</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>short C4</td>
<td>long and short C4</td>
</tr>
<tr>
<td>ECY504</td>
<td>A 2,3 B 1</td>
<td>A *2,*3 B *1</td>
<td>K</td>
<td>K</td>
</tr>
<tr>
<td>ECY751</td>
<td>A 3 B 1</td>
<td>A *3 B *1</td>
<td>K</td>
<td>K</td>
</tr>
<tr>
<td>ECY858</td>
<td>A 3 B 1</td>
<td>A *3 B *1</td>
<td>K</td>
<td>K&amp;M</td>
</tr>
<tr>
<td>ECY248</td>
<td>A 4 B 1,2</td>
<td>A *4 B *2,*22</td>
<td>K&amp;M</td>
<td>K&amp;M</td>
</tr>
<tr>
<td>ECY302</td>
<td>A 2,4 B 2,4</td>
<td>A *2,*4 B *2,*4</td>
<td>K</td>
<td>K</td>
</tr>
<tr>
<td>WCY1207</td>
<td>A 3 B 1</td>
<td>A *3 B *1</td>
<td>K</td>
<td>K</td>
</tr>
<tr>
<td>WCY1158</td>
<td>A 3,6 B 1</td>
<td>A *3,*58 B *1</td>
<td>K</td>
<td>K&amp;M</td>
</tr>
<tr>
<td>ECY432</td>
<td>A 2,4 B 1,2</td>
<td>A *2,*4 B *2,*22</td>
<td>K&amp;M</td>
<td>K&amp;M</td>
</tr>
<tr>
<td>WCY1179</td>
<td>A 3 B 12</td>
<td>A *3 B *12</td>
<td>no product</td>
<td>K&amp;M</td>
</tr>
<tr>
<td>WCY1204</td>
<td>A 3 B 1,12</td>
<td>A *3 B *1,12</td>
<td>no product</td>
<td>K&amp;M</td>
</tr>
</tbody>
</table>

# allotype determined by protein gel electrophoresis. * allotype determined by DNA analysis.

'no product' refers to lack of PCR amplification of short C4 alleles, indicating only long C4 genes were encoded by these donors. All amino acid residues are represented by their one letter code as listed in Table 4.2.
### Table 4.6 \( \beta \)-chain, exon 9 polymorphic amino acid residues

<table>
<thead>
<tr>
<th>Donor</th>
<th>C4 allotype#</th>
<th>short C4</th>
<th>long and short C4</th>
<th>C4 allotype*</th>
<th>short C4</th>
<th>long and short C4</th>
<th>amino acid # 306</th>
<th>amino acid # 328</th>
</tr>
</thead>
<tbody>
<tr>
<td>JC001</td>
<td>A ( 4 ) B 2</td>
<td>K</td>
<td>K</td>
<td>A ( 4 ) B*2,*22</td>
<td>K</td>
<td>no product</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>JC002</td>
<td>A ( 3 ) B *1</td>
<td>K</td>
<td>K</td>
<td>A ( 3 ) B*1</td>
<td>K</td>
<td>no product</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>JC003</td>
<td>A ( 3.6 ) B *1</td>
<td>K</td>
<td>K</td>
<td>A ( 3.6 ) B*1</td>
<td>K</td>
<td>no product</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>RCH 517</td>
<td>A ( 3 ) B *3</td>
<td>K</td>
<td>K</td>
<td>A ( 3 ) B *3</td>
<td>K</td>
<td>no product</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>EW-L0081785</td>
<td>A ( 3 ) B *Q0</td>
<td>K</td>
<td>K</td>
<td>A ( 3 ) B *Q0</td>
<td>K</td>
<td>no product</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>EW-Ducaf</td>
<td>A ( 3 ) B *Q0</td>
<td>K</td>
<td>K</td>
<td>A ( 3 ) B *Q0</td>
<td>K</td>
<td>no product</td>
<td>S</td>
<td>S</td>
</tr>
</tbody>
</table>

*allotype determined by protein gel electrophoresis. \# allotype determined by DNA analysis. 'no product' refers to lack of PCR amplification of short C4 alleles, indicating only long C4 genes were encoded.

Based on these results, it is observed that the presence of short C4 alleles is associated with the absence of certain specific amino acid residues, which in turn may have implications for the functional properties of the \( \beta \)-chain.
heterozygosity of nucleotide bases encoding amino acid residue #328 in the PCR product where primers had amplified all C4 genes both long and short. However, this approach was only possible if homozygosity was detected in the short C4 gene at residue #328. Taq I RFLP studies of 13 Aboriginal donors (some of which were included in PCR-RFLP analysis), and two Caucasian donors, support the finding by PCR-RFLP analysis, that the majority of donors have both long and short C4 genes (Table 4.7). From PCR-RFLP analysis, donors WCY1179, WCY1204, EW-L0081785 and EW-Ducaf, have only long C4 genes, as short C4 genes could not be amplified (Tables 4.5 and 4.6). Taq I RFLP analysis of donor WCY1204 confirmed the analysis by PCR-RFLP, that only long C4 genes were found at locus I and II in this donor (Table 4.7).

For the majority of alleles in this study, nucleotide bases encoding 'Tyr' for amino acid #328 were associated with long C4 genes, whereas nucleotide bases which encode 'Ser' for residue #328 were associated with short C4 genes (Tables 4.5 and 4.6). PCR-RFLP analysis of nucleotides encoding residue #328 in WCY1179 and WCY1204 (both donors only had long C4 genes), supported this finding, as each allele had residue 'Tyr'. Donors ECY 858, EW-L0081785, EW-Ducaf and ECY 178, on the other hand, were exceptions. Residue 'Ser' was encoded by nucleotides in both long and short C4 alleles of donor ECY 858, and in the long C4 alleles of donors EW-L0081785 and EW-Ducaf, whereas residues 'Ser' and 'Tyr' were detected in the short C4 alleles of donor ECY 178. Taq I RFLP analysis was necessary to establish these associations, as nucleotides encoding residue #328 in both long and short C4 alleles of these donors were homozygous. In the case of donor ECY 858, it should be noted that Taq I RFLP analysis data was not available to confirm that both long and short C4 genes were present.

A search was undertaken for nucleotide differences between long and short C4 genes which may identify these types of alleles. Yu (1991), reported the sequence for the large 7 kb intron 9 of a C4A *3a allele, where two copies of a mobile genetic element encoding nucleotide bases 5'-acagaca-3' at the beginning and end of the intron
Table 4.7 *Taq I* RFLP analysis of the *C4* gene region for Aboriginal Australians, and Caucasians

<table>
<thead>
<tr>
<th>Donor</th>
<th><em>C4</em> allotype#</th>
<th>Locus 1</th>
<th>Locus 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECY006</td>
<td>A 3 B 1,Q0</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>ECY218</td>
<td>A 3 B 2</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>ECY277</td>
<td>A 4 B 1,Q0</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>ECY720</td>
<td>A 4 B 1</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>ECY740</td>
<td>A 3 B 1</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>WCY1159</td>
<td>A 3,Q0 B 1</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>ECY704</td>
<td>A 3,Q0 B 1</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>WCY1176</td>
<td>A 4 B 1,2</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>WCY1199</td>
<td>A 4,Q0 B 2</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>WCY1204</td>
<td>A 3 B 1,12</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>ECY028</td>
<td>A 4 B 2</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>ECY248</td>
<td>A 4 B 1,2</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>ECY504</td>
<td>A 2,3 B 1</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>JC002</td>
<td>A 3 B 1</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>JC003</td>
<td>A 3,6 B 1</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

# allotype determined by protein gel electrophoresis. "Taq I" RFLP *C4* locus I and II long (23 kb) genes are represented by 7 kb and 6 kb fragments, respectively. *C4* locus II short (16 kb) genes are represented by 5.4 kb fragments. Locus I gene deletion and short (16 kb) locus II gene is represented by 6.4 kb fragments. RFLP fragments represent 21-hydroxylase A gene by 3.7 kb and 21-hydroxylase B gene by 3.2 kb.
9 sequence were detected. To establish if intron 9 of the short \( C4 \) gene encoded one copy of the mobile genetic element, and to search for any other sequence differences which could be used to distinguish long from short \( C4 \) genes, the short \( C4 \) gene in the region of intron 9 was sequenced. If an association between specific nucleotide bases and the size of \( C4 \) alleles was found, this information could be used to analyse polymorphic residues separately in long and short alleles, and thereby be of use to the PCR-RFLP allotyping protocol.

Various donors which were found to have nucleotide bases encoding amino acid residues 'Ser' and/or 'Tyr' for residue #328, by PCR-RFLP analysis of the short \( C4 \) genes, were chosen for sequencing analysis. Sequencing results were compared with long \( C4 \) allele '\( C4A *3a \)' nucleotide base sequence reported by Yu, (1991), and are shown in Figure 4.3. Only nucleotide bases of the long \( C4A *3a \) allele which are also found in the short \( C4 \) alleles are shown. The remaining nucleotides of this allele, which comprise approximately 7 kb of intron 9 sequence, are not shown, and if included would appear at the site of the mobile genetic element.

The short \( C4B \) null allele of Aboriginal donor H14 (denoted as \( C4A \ ^{CAN1} \), reported in Chapter 5), unlike the majority of other short \( C4 \) alleles (Tables 4.5 and 4.6), had 'Tyr' for residue #328. A Caucasian donor RCH517 (\( A *3 B *3 \)), was heterozygous for residue #328 in the short \( C4 \) alleles (RCH517 a and b), having nucleotide bases encoding 'Ser' and 'Tyr' (Table 4.6). Aboriginal donors WCY1158 and ECY 028 were homozygous for short \( B *1 \) and \( B *2 \) alleles, respectively), both having residue 'Ser' for residue #328. Aboriginal donor ECY 178 (heterozygous short \( B *2,*5 \) alleles) was heterozygous having residues 'Ser' and 'Tyr' encoded by nucleotides for residue #328. Only the short \( B *5 \) allele of donor ECY 178 having residue 'Tyr' was sequenced, as the two short alleles in this donor were digested using restriction enzyme '\( P_{le} I \)' and then size fractionated by gel electrophoresis and electro-eluted prior to cloning and sequencing.
Figure 4.3 Sequencing results of short C4 genes from nucleotide bases encoding residues #286 to #330, which extends across the variable sized intron 9. Each short C4 gene sequence is aligned with the corresponding long C4A *3a gene sequence reported by Yu (1991). C4 alleles are from the following donors, A *CAN1 from H14, B *5 from ECY178, B *1 from WCY1158, B *2 from ECY028, RCH517(a) and RCH517(b) are from donor RCH517. Upper-case sequence represent exon sequence, whereas lower-case represents intron 9 sequence. Numbers in bold print represent the position of a code of nucleotide bases for an amino acid residue. Intron 9 nucleotid bases are numbered #1 to #420, and numbers denoted '#' represent a particular intron base. Nucleotides encoding residue #286 to intron 9 base #107 were sequenced in C4 allele 'B *5', as sequencing was of a restricted PCR fragment, not a full length PCR fragment, as were the other sequences. The large 7 kb intron present in 23 kb C4 genes is not shown, however has been reported by Yu (1991), to be positioned between intron 9 base #280 and #287 (mobile genetic element) in this figure.
amino acid # 286 (exon 9)
A *3a  cag CTG GTG AAT GGA CAG AGC CAC ATT TCC CTC TCA AAG GCA GAG TTC
A *CAN 1  cag CTG GTG AAT GGA CAG AGC CAC ATT TCC CTC TCA AAG GCA GAG TTC
B *5  cag CTG GTG AAT GGA CAG AGC CAC ATT TCC CTC TCA AAG GCA GAG TTC
B *I&B *2  cag CTG GTG AAT GGA CAG AGC CAC ATT TCC CTC TCA AAG GCA GAG TTC
RCH517(a)  cag CTG GTG AAT GGA CAG AGC CAC ATT TCC CTC TCA AAG GCA GAG TTC
RCH517(b)  cag CTG GTG AAT GGA CAG AGC CAC ATT TCC CTC TCA AAG GCA GAG TTC

amino acid # 306 309
A *3a  CAG GAC GCC CTG GAG AAG CTG AAT ATG GGC ATT ACT GAC CTC CAG
A *CAN 1  CAG GAC GCC CTG GAG ATG CTG AAT ATA GGC ATT ACT GAC CTC CAG
B *5  CAG GAC GCC CTG GAG ATG CTG AAT ATA GGC ATT ACT GAC CTC CAG
B *I&B *2  CAG GAC GCC CTG GAG AAG CTG AAT ATG GGC ATT ACT GAC CTC CAG
RCH517(a)  CAG GAC GCC CTG GAG ATG CTG AAT ATA GGC ATT ACT GAC CTC CAG
RCH517(b)  CAG GAC GCC CTG GAG AAG CTG AAT ATG GGC ATT ACT GAC CTC CAG

amino acid # 328 intron 9
A *3a  GGG CTG CGC CTC TAC GTT GGT GCT GCA GCC ATC ATT GAG TAT CCA G gtgggtg
A *CAN 1  GGG CTG CGC CTC TAC GTT GGT GCT GCA GCC ATC ATT GAG TCT CCA G gtgggtg
B *5  GGG CTG CGC CTC TAC GTT GGT GCT GCA GCC ATC ATT GAG TCT CCA G gtgggtg
B *I&B *2  GGG CTG CGC CTC TAC GTT GGT GCT GCA GCC ATC ATT GAG TCT CCA G gtgggtg
RCH517(a)  GGG CTG CGC CTC TAC GTT GGT GCT GCA GCC ATC ATT GAG TCT CCA G gtgggtg
RCH517(b)  GGG CTG CGC CTC TAC GTT GGT GCT GCA GCC ATC ATT GAG TCT CCA G gtgggtg

intron 9 bp#
A *3a  acttccccctttggaaccccagacctttgactctgagctacaaccctcagcacaacacca
A *CAN 1  acttccccctttggaaccccagacctttgactctgagctacaaccctcagcacaacacca
B *5  acttccccctttggaaccccagacctttgactctgagctacaaccctcagcacaacacca
B *I&B *2  acttccccctttggaaccccagacctttgactctgagctacaaccctcagcacaacacca
RCH517(a)  acttccccctttggaaccccagacctttgactctgagctacaaccctcagcacaacacca
RCH517(b)  acttccccctttggaaccccagacctttgactctgagctacaaccctcagcacaacacca

intron 9 bp#
A *3a  ccccc actttcactcactcactcagactcaaaacc aggaacaccccagagacagctgtttcttcacaactc
A *CAN 1  ccccc actttcactcactcactcagactcaaaacc aggaacaccccagagacagctgtttcttcacaactc
B *5  ccccc actttcactcactcactcactaca
B *I&B *2  ccccc actttcactcactcactcagactcaaaacc aggaacaccccagagacagctgtttcttcacaactc
RCH517(a)  ccccc actttcactcactcactcagactcaaaacc aggaacaccccagagacagctgtttcttcacaactc
RCH517(b)  ccccc actttcactcactcactcagactcaaaacc aggaacaccccagagacagctgtttcttcacaactc
A

*3a

tgtctctgtgactcggcccttttccctggctgagtttatttatttctttgctcgttctgctcattccttcactcctcca

A

*CAN 1

tgtctctgtgactcggcccttttccctggctgagtttatttatttctttgctcgttctgctcattccttcactcctcca

B *1&B *2

tgtctctgtgactcggcccttttccctggctgagtttatttatttctttgctcgttctgctcattccttcactcctcca

R C H 5 1 7 ( a ) tgtctctgtgactcggcccttttccctggctgagtttatttatttctttgctcgttctgctcattccttcactcctcca
R C H 5 1 7(b) tgtctctgtgactcggcccttttccctggctgagtttatttatttctttgctcgttctgctcattccttcactcctcca

intron

9 bp# 238

#280

#287

#295

#304

A *3a

gtggacatgtgttgttcaatgccccgtgctaggcctcagcatgcacagacaggaatacggcagcctgtgccctg

A *CAN 1

gtggacatgtgttgttcaatgccccgtgctaggcctcagcatg'lacagacaggaatacagcagcctgcgccctg

B *1&B *2

glggacatgtgttgttcaatgccccgtgctaggcctcagcatgtacagacaggaatacagcagcctgcgccctg

RC H5 17(a) ggggacatgtgttgttcaatgccccgtgctaggcctcagcatgtacagacaggaatacagcagcctgcgccctg
R C H 5 1 7(b) ggggacatgtgttgttcaatgccccgtgctaggcctcagcatgiacagacaggaatacagcagcctgcgccctg
intron 9 bp#

#354

A *3a

ggagctcactgtcttgtgggagggaaccactcaagccactccccacttgtcctcctgtccctctcttcttgggctc

A *CAN 1

ggagctcactgtcttgtgggagggaaccactcaagccactccctacttgtcctcctgtccctctcttcttgggctc

B *1&B *2

ggagctcactgtcttgtgggagggaaccactcaagccactccctacttgtcctcctgtccctctcttcttgggctc

R C H 5 17(a) ggagctcactgtcttgtgggagggaaccactcaagccactccctacttgtcctcctgtccctctcttcttgggctc
R C H 5 1 7 ( b ) ggagctcactgtcttgtgggagggaaccactcaagccactccctacttgtcctcctgtccctctcttcttgggctc
intron

9 bp#

#4 2 0

330

(exon 10)

A *3a

tgtcccccacctctctctgtcctttgtcttgcag

GT GGG GAG ATG G

A *CAN 1

tgtcccccacctctctctgtcctttgtcttgcag

GT GGG GAG ATG G

B *1&B *2

tgtcccccacctctctctgtcctttgtcttgcag

GT GGG GAG ATG G

R C H 5 1 7 ( a ) tgtcccccacctctctctgtcctttgtcttgcag

GT GGG GAG ATG G

R C H 5 1 7 ( b ) tgtcccccacctctctctgtcctttgtcttgcag

GT GGG GAG ATG G


Sequencing data (Figure 4.3) showed that the \( C4A \) \(*CAN1\) allele from donor H14, the \( B *5 \) allele from donor ECY 178, and one allele from donor RCH517(a), possessed nucleotide bases which encoded amino acid residue 'Met' for residue #306. On the other hand, published nucleotide base sequence from the long \( C4 \) allele '\( C4A *3a \)' (Yu, 1991), and sequencing results of donors WCY1158 (\( B *1 \)), ECY 028 (\( B *2 \)) and the other short \( C4 \) allele of donor RCH517(b), showed nucleotide bases encoding amino acid residue 'Lys' for residue #306. Sequencing revealed differences in nucleotide bases between the same alleles for residue #309 (Figure 4.3). The \( C4A *CAN1 \) and \( B *5 \) alleles, and donor RCH517(a) allele, had nucleotide bases for amino acid residue 'Ile', whereas the published long \( C4A *3a \) allele, as well as donors WCY1158 (\( B *1 \)), ECY 028 (\( B *2 \)) and RCH517(b) alleles, had nucleotides which encode residue 'Met'. In addition, heterogeneity was detected for nucleotide bases encoding residue #328. The \( C4A *CAN1 \) and \( B *5 \) alleles, and donor RCH517(a) allele, and the published \( A *3a \) alleles had nucleotide bases which encode residue 'Tyr'. Donors WCY1158 (\( B *1 \)), ECY 028 (\( B *2 \)) and RCH517(b) alleles had nucleotides which encode residue 'Ser'. These results were the same as those obtained by PCR-RFLP analysis of nucleotide bases encoding residue #328 (section 4.4.2.1).

Sequencing the short \( C4 \) genes of the above donors, found intron 9 to be 420 bp compared with approximately 7 kb in the long type of \( C4 \) gene reported by Yu (1991). Polymorphism between the short \( C4 \) alleles sequenced in this study and the published sequence of '\( A*3a \)' (Yu, 1991) was detected of nucleotide bases in intron 9. For convenience, intron 9 has been numbered from 1 to 420 bp in this study. Yu (1991), detected two copies of the mobile genetic element 5'-acagaca-3', at the beginning and end of the long \( C4 \) gene '\( C4A *3a \)'. In these studies a single copy of the mobile genetic element was detected in the short \( C4 \) alleles which were sequenced from intron 9 base #281 to #287 (Figure 4.3). A polymorphic nucleotide base #280, which may distinguish long from short \( C4 \) alleles, was found closely situated to the mobile genetic element in intron 9. Yu, (1991), found this base was a 'cytosine' in the long \( C4 \) allele '\( A3a \)'. In these studies however, the base was found to be a 'thymine' in the short \( C4 \) alleles.
Other polymorphic nucleotide bases of intron 9 in short $C4$ alleles were detected in this study (Figure 4.3), which are: (1) The 70th nucleotide base of intron 9 in each short $C4$ allele which encoded 'guanine', whereas the long $A^{*3a}$ sequence (Yu, 1991) encoded 'adenine'; (2) The short $C4$ alleles from this study had 'adenine' at nucleotide base #295 of intron 9, whereas Yu (1991) reported 'guanine' for this nucleotide base in the long $A^{*3a}$ allele; (4) Intron 9 nucleotide base #304 in the short $C4$ alleles was 'cytosine', whereas Yu (1991) reported 'thymine' for this base in the long $A^{*3a}$ allele; (5) Nucleotide base #354 of intron 9 in short $C4$ alleles of this study was 'thymine', whereas Yu (1991) reported 'cytosine' for this base in the long $A^{*3a}$ allele. Yu (1991), reported nucleotide base 'cytosine', at base positions #87 and #119 in intron 9 of the long $C4A^{*3a}$ allele. However, these nucleotide bases were deleted in the short $C4$ alleles of this study (Figure 4.3). One other polymorphism was detected in intron 9 at nucleotide base #238 (Figure 4.3). The long $A^{*3a}$ allele reported by Yu (1991), and short $C4$ alleles from donors H14 ($C4A\ CAN^{*1}$) and WCY1158 ($B^*I$) and ECY028 ($B^*2$) in this study, had nucleotide base 'thymine', whereas both types of short $C4$ alleles of donor RCH517 in this study, had 'guanine'.

A number of these polymorphic sequences can be detected by restriction analysis, and therefore could potentially be useful in the PCR-RFLP protocol. Of those polymorphisms detected in intron 9, the following can be detected by various enzymes; nucleotide base #70 by $Hgi\ CI$ and $Nla\ 4$; base #238 by $Bsr\ I$, $Sec\ I$, $EcoR\ II$, $Bss\ KI$, $Aea\ I$, and $Scr\ FI$; #280 by $Rsa\ I$, $Bsp\ 140$, $Csp\ 6I$, and $Sph\ I$; and #304 by $Hin\ 6I$ and $Hha\ I$. The deletion which was detected in intron 9 of base #119, can be detected by restriction enzyme $Sec\ I$. Polymorphic nucleotide bases encoding amino acid residues #306 and #328 can be detected by restriction enzymes $Alu\ I$ and $Ple\ I$, respectively. $Alu\ I$ restriction analysis of nucleotides encoding residue #306 has been included in the PCR-RFLP allotyping protocol. A number of polymorphic nucleotide bases including those which encode residue #309, nucleotide bases #295 and #354 in intron 9, and the deletion of base #87 in intron 9, are not recognised by restriction enzymes which are commercially available.
4.4.2.2 Amino acid residue #306

PCR-RFLP analysis of nucleotide bases encoding amino acid residue #306 for each donor, and resulting restriction fragment banding patterns are shown in Tables 4.5 and 4.6, and Figure 4.2. The majority of C4 alleles had nucleotide bases which encode amino acid residue 'Lys' for residue #306, in both long and short C4 genes. Nucleotides encoding amino acid residue 'Met' were mainly found in donors having rare alleles, and were not necessarily associated with either long or short C4 alleles.

4.4.2.3 Amino acid residue #458

PCR-RFLP banding patterns for polymorphic amino acid residues #458 are represented in Figure 4.4, and a summary of results for each donor is given in Tables 4.8 and 4.9. In this study, only a single Caucasian donor 'JC003' having the allotype C4A *3,*6 B *1, showed polymorphism of residue #458 encoding 'Trp' (see Table 4.9). It is possible the C4A *6 protein in this donor is haemolytically inactive, as the same polymorphism has been reported for the haemolytically inactive A *6 allele described by Anderson et al. (1992).

4.4.2.4 Amino acid residue #707

PCR-RFLP banding patterns from restriction of nucleotide bases encoding amino acid residue #707 are shown in Figure 4.4. Polymorphism of nucleotide bases encoding this amino acid residue was not detected amongst the C4 alleles represented in this study (see Tables 4.8 and 4.9). Amino acid residue 'Pro' was encoded by nucleotides in each allele. Nucleotides of C4A*3a allele described by Yu (1991), and Belt et al. (1984), however, were found to encode 'Leu'. Yu (1991), reported that residue 'Leu' was not detected in a B *5 or B *3 allele, nor in ten other randomly chosen donors. This amino acid residue was therefore most probably the result of a rare mutation.
Figure 4.4 PCR-RFLP banding patterns of nucleotide bases encoding C4 allele polymorphic amino acid residues #458, #707, #1054 and #1101-#1106. Examples of each banding pattern are listed as lane numbers for the respective amino acids. Where homozygote examples of each banding pattern could not be shown, a heterozygous example was referred to. Molecular weight markers consisted of pBR322 restricted with either Msp I or Hae III, and resulting fragment sizes (bp) of each are listed in Figure 2.1, Chapter 2. pBR322 restricted with Msp I resulting in 67, 76, 90, 110, 123, 147, 160, 180, 190, 201, 217, 238, 242, 309, 404, 527, and 622 base pair fragments are shown in lane 9 in the gel for amino acid residue #458 as well as the 34 bp fragment, lane 1 for residue #707, and lane 2 for the gel corresponding to residue #1101-#1106. pBR322 restricted with Hae III resulting in 18, 21, 51, 57, 64, 80, 89, 104, 123, 124, 184, 192, 213, 234, 267, 434, 458, 504, 540, and 587 base pair fragments are shown in lane 2 for residue #1054. Unrestricted PCR DNA fragments are shown in lanes 2 and 1 of the gels corresponding to amino acid residues #707 and #1101-#1106, respectively. Unrestricted PCR fragments for PCR-RFLP analysis of residues #458 and #1054 are not shown in this figure. Small PCR fragments less than 19 bp were lost by gel electrophoresis for residue #1054, and fragments less than 40 bp for residues #1101-#1106.
<table>
<thead>
<tr>
<th>amino acid residue #</th>
<th>Fnu4HI PCR-RFLP (bp)</th>
<th>lane # example</th>
</tr>
</thead>
<tbody>
<tr>
<td>458</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>41+53+109</td>
<td>1,2</td>
</tr>
<tr>
<td>W</td>
<td>53+150 (R/W heterozygote)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>amino acid residue #</th>
<th>Fnu4HI PCR-RFLP (bp)</th>
<th>lane # example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1054</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>3+5+22+60+63+269</td>
<td>9</td>
</tr>
<tr>
<td>G</td>
<td>3+3+5+19+60+63+269</td>
<td>3,4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>amino acid residue #</th>
<th>Pvu II PCR-RFLP (bp)</th>
<th>lane # example</th>
</tr>
</thead>
<tbody>
<tr>
<td>707</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>104+108</td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>212</td>
<td>3,4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>amino acid residue #</th>
<th>Mnl I PCR-RFLP (bp)</th>
<th>lane # example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1101-1106</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCPVLD</td>
<td>3+5+9+22+28+31+31+40+69+103+108+139</td>
<td>7,8</td>
</tr>
<tr>
<td>LSPVIH</td>
<td>3+5+9+22+28+31+31+40+53+55+69+103+139</td>
<td>3</td>
</tr>
<tr>
<td>Donor</td>
<td>C4 allelotype#</td>
<td>C4 allelotype*</td>
</tr>
<tr>
<td>--------</td>
<td>----------------</td>
<td>---------------</td>
</tr>
<tr>
<td>ECY504</td>
<td>A 2,3 B 1</td>
<td>A*2,<em>3 B</em>1</td>
</tr>
<tr>
<td>WCY1174</td>
<td>A 2,4 B 2</td>
<td>A<em>4 B</em>2,*6</td>
</tr>
<tr>
<td>WCY1207</td>
<td>A 3 B 1</td>
<td>A<em>3 B</em>1</td>
</tr>
<tr>
<td>ECY751</td>
<td>A 3 B 1</td>
<td>A<em>3 B</em>1</td>
</tr>
<tr>
<td>ECY858</td>
<td>A 3 B 1</td>
<td>A<em>3 B</em>1</td>
</tr>
<tr>
<td>ECY028</td>
<td>A 4 B 2</td>
<td>A<em>4 B</em>2</td>
</tr>
<tr>
<td>ECY248</td>
<td>A 4 B 1,2</td>
<td>A<em>4 B</em>2,*22</td>
</tr>
<tr>
<td>WCY1158</td>
<td>A 3,6 B 1</td>
<td>A*3,<em>58 B</em>1</td>
</tr>
<tr>
<td>ECY178</td>
<td>A 4,6 B 2</td>
<td>A*4,<em>58 B</em>2,*5</td>
</tr>
<tr>
<td>ECY432</td>
<td>A 2,4 B 1,2</td>
<td>A*2,<em>4 B</em>2,*22</td>
</tr>
<tr>
<td>WCY1179</td>
<td>A 3 B 12</td>
<td>A<em>3 B</em>12</td>
</tr>
<tr>
<td>WCY1204</td>
<td>A 3 B 1,12</td>
<td>A<em>3 B</em>1,*12</td>
</tr>
<tr>
<td>ECY302</td>
<td>A 2,4 B 2,4</td>
<td>A*2,<em>4 B</em>2,*4</td>
</tr>
</tbody>
</table>

# allotype determined by protein gel electrophoresis. * allotype determined by DNA analysis. Amino acid residue #1186 is alanine 'A' for both C4A and C4B alleles, however the nucleotide bases encoding the amino acid are different, 'gcg' for C4A alleles, and 'gcc' for C4B alleles. 'gcg' is represented by 'a' and 'gcc' by 'b.'
Table 4.9 PCR-RFLP results for C4 polymorphic amino acid residues of Caucasians

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>JC001</td>
<td>A 4 B 2</td>
<td>A<em>4 B</em>2,22</td>
<td>R P P P</td>
<td>D</td>
<td>G&amp;D</td>
<td>PCPVLD</td>
<td>LSPVIH</td>
<td>N S a b</td>
<td>V A</td>
<td>L R</td>
<td>A A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JC002</td>
<td>A 3 B 1</td>
<td>A<em>3 B</em>1</td>
<td>R P P P</td>
<td>D</td>
<td>G</td>
<td>PCPVLD</td>
<td>LSPVIH</td>
<td>N S a b</td>
<td>V A</td>
<td>L R</td>
<td>A A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JC003</td>
<td>A 3,6 B 1</td>
<td>A*3,<em>6 B</em>1</td>
<td>W P P D</td>
<td>G</td>
<td>PCPVLD</td>
<td>LSPVIH</td>
<td>N N a b</td>
<td>V A</td>
<td>R&amp;L R</td>
<td>A&amp;S A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCH517</td>
<td>A 3 B 3</td>
<td>A *3 B *3</td>
<td>- - - G</td>
<td>PCPVLD</td>
<td>LSPVIH</td>
<td>N S a b</td>
<td>V - L -</td>
<td>uncut</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EW-L0081785</td>
<td>A 3 B 0</td>
<td>A *3 B *Q0</td>
<td>- - - D</td>
<td>PCPVLD</td>
<td>deleted</td>
<td>S - a - V - L -</td>
<td>uncut</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EW-Ducaf</td>
<td>A 3 B 0</td>
<td>A *3 B *Q0</td>
<td>- - - G&amp;D</td>
<td>PCPVLD</td>
<td>deleted</td>
<td>S - - V - L - A -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

# allotype determined by protein gel electrophoresis. * allotype determined by DNA analysis. Amino acid residue #1186 is the same for both C4A and C4B alleles, however the nucleotide bases encoding the amino acid are different, 'g' for C4A alleles, and 'c' for C4B alleles. 'g' is represented by 'a' and 'c' by 'b':'-' represents not determined. Donors EW-L0081785 and EW-Ducaf had B*Q0 alleles estimated to result from deletion of the C4B genes. PCR-RFLP analysis of these donors showed that only C4A specific nucleotide bases encoded amino acid residues #1101 to #1106. Therefore C4B isotypic analysis of other nucleotides encoding polymorphic amino acid residues in C4B alleles for these donors were not determined.
4.4.2.5 Amino acid residue #1054

PCR-RFLP banding patterns from restriction analysis with enzyme Fnu4HI are shown in Figure 4.4, and a summary of the results for each donor is given in Tables 4.8 and 4.9. Amino acid residue #1054 is polymorphic, having nucleotide bases encoding either residue 'Gly' or 'Asp'. It is isotypic for the majority of C4A alleles, and allele-specific for C4B alleles (Tables 4.8 and 4.9). The A *2 alleles of donors ECY504 and ECY302 were found to be exceptions by having nucleotide bases which encode residue 'Gly' for residue #1054. This was also reported by Paz-Artal et al. (1993). Other researchers have found residue 'Gly' for residue #1054 in C4A *1 alleles and also in C4A *2 (Yu et al., 1986; Anderson et al., 1992). In this study, residue #1054 in C4B *1, B *12, B *22, B *3, B *4 and B *5 alleles was found to be 'Gly', whereas in C4B *2 and B*6 it was 'Asp', (Table 4.8 and 4.9). Others have found the same association of 'Gly' with C4B *1 and 'Asp' with C4B *2 (Belt et al., 1984, and 1985; Yu et al., 1986; Yu., 1991; Anderson et al., 1992). Consistent with results here, Yu et al. (1986), found B *3 to have residue 'Gly' for residue #1054, whereas contrary to results found here, the C4B *5 allele was reported to have 'Asp'. Belt et al. (1985), detected another C4B *1 allele named C4B *1b, which had residue 'Asp' encoded by nucleotides for residue #1054. A number of individuals found to be heterozygous for residue #1054 were homozygous by C4 protein gel electrophoresis. For example, C4A alleles of donor EW-Ducaf, and C4B alleles of donors ECY 178 and WCY1174.

4.4.2.6 Amino acid residues #1101-#1106

Amino acid residues #1101-#1106 were isotypic in nature. This region of the C4 gene is conserved and distinguishes C4A from C4B alleles. In this study, individual's C4A alleles had residues 'Pro,Cys,Pro,Val,Leu,Asp', whilst their C4B alleles had 'Leu,Ser,Pro, Val,Ile,His' (Figure 4.4 and Tables 4.8 and 4.9). Other researchers have reported the same finding with C4 alleles (Belt et al., 1984 and 1985; Yu et al., 1986; Yu, 1991; Anderson et al., 1992; Paz-Artal et al., 1993). Both EW-L0081785 and EW-Ducaf are Caucasian
donors with deleted \( C4B \) genes, and therefore only \( C4A \) isotypic residues were detected (Figure 4.4 and Table 4.9).

4.4.2.7 Amino acid residue \#1157

PCR-RFLP banding patterns from restriction analysis using enzyme \( Alu I \), and results for each donor are shown in Figure 4.5, and Tables 4.8 and 4.9. As with residue \#1054, amino acid residue \#1157 was generally isotypic, \( C4A \) alleles encoding 'Asn' and \( C4B \) alleles encoding 'Ser'. Exceptions were detected in donors WCY1174, ECY 178 and JC003, where residue 'Asn' was found in the nucleotide sequence of the \( C4B \) alleles (Table 4.8 and 4.9). Similarly, other researchers have found 'Asn' in the nucleotide sequence of a \( C4B \) *1a allele (Belt et al., 1985). Another exception was detected in \( C4A \) alleles of donors EW-L0081785 and EW-Ducaf, where residue 'Ser' was present instead of 'Asn'. The \( C4A \) *1 allele has also been reported to have 'Ser' at this residue (Yu et al., 1986; Anderson et al., 1992).

4.4.2.8 Amino acid residue \#1182

Polymorphism of residue \#1182 has only been detected in a \( C4A \) *4 allele, where nucleotide bases were found to encode amino acid residue 'Ser' by cDNA sequencing (Belt et al., 1984). All other reported nucleotide base sequence studies for both \( C4A \) and \( C4B \) alleles found amino acid residue 'Thr' encoded for residue \#1182 (Belt et al., 1985; Yu et al., 1986; Yu., 1991; Kawaguchi et al., 1992b). To determine if residue \#1182 was polymorphic in \( C4A \) alleles, allele specific oligonucleotide (ASO) hybridisation of \( C4A \) specifically amplified PCR products from Aboriginal Australians and three Caucasian controls (described in section 4.3.10.1), was carried out. This study detected nucleotide bases encoding amino acid 'Thr' for \#1182. Sequencing the gene region encoding this residue in donor ECY743 having an A *3 allele, and another donor ECY519 having an A *4 allele, found only nucleotide bases encoding residue 'Thr'. The presence of nucleotide bases encoding amino acid residue 'Asn' for residue \#1157 confirmed that only the \( C4A \) gene, and not the \( C4B \)
**Figure 4.5** PCR-RFLP banding patterns of nucleotide bases encoding polymorphic amino acid residues #1157, #1186, #1188, #1191, and #1267 in C4 PCR fragments. Examples of each banding pattern are listed as lane numbers for the respective amino acids. Restriction analysis by enzyme *Aci I* of the same PCR product identified which nucleotide bases encoded residues #1188 and #1191. Four photographs of gels representing the larger (429 and 510 bp) and smaller (73 and 51 bp) PCR fragment banding patterns of nucleotides encoding residues #1188 and #1191 surround the table showing the respective amino acid residues. Molecular weight marker *pBR322* restricted with *Msp I* was used for each gel and is represented by 67, 76, 90, 110, 123, 147, 160, 180, 190, 201, 217, 238, 242, 309, 404, 527, and 622 base pair fragments in lane 2 and 8 for residue #1157, lanes 17, and 21 for residues #1188 and #1191, and lane 2 for residue #1267. Molecular weight markers 622, 527, 404 and 309 base pair fragments are shown in lane 1 for residue #1186 and lane 7 for residues 1188 and 1191, and an additional fragment of 242 base pairs is shown in lane 17 for residue 1186 and lane 14 for residues 1188 and 1191. Unrestricted PCR DNA fragments are shown in lane 1 and 7 for gels corresponding to residue #1157, lane 2 and 18 for gels corresponding to residue #1186 table, lane 8, 15, 16, and 22 for gels corresponding to residues #1188 and #1191 table, and lane 1 for the gel corresponding to residue #1267 table. Only the larger (396 and 481 bp) fragments are shown in the gels corresponding to residue #1186 table. Fragments less than 41 bp were lost during gel electrophoresis.
gene, had been sequenced.

4.4.2.9 Amino acid residue #1186

Results of PCR-RFLP analysis of nucleotide bases encoding residue #1186 for each donor, and examples of the resulting restricted fragment banding patterns are shown in Tables 4.8 and 4.9, and Figure 4.5. Polymorphism of nucleotide bases encoding this amino acid residue was only detected at the nucleotide base level, 'gcg' for C4A alleles, and 'gcc' for C4B alleles. This study showed that these bases were isotypic in nature as were nucleotide bases encoding amino acid residues #1101 - #1106, and therefore were useful for distinguishing C4A from C4B alleles. Kawaguchi et al. (1992b), found that the A *7 allele, unlike other C4A alleles, encoded nucleotide bases 'gcc' at this position. Studies here, found heterozygosity of the nucleotide bases encoding amino acid residue #1186 in C4B alleles of donors WCY1174, ECY 178 and ECY 302 (Table 4.8).

4.4.2.10 Amino acid residues #1188 and #1191

Results of PCR-RFLP analysis of nucleotide bases encoding residues #1188 and #1191 for each donor, and examples of the resulting restricted fragment banding patterns, are shown in Tables 4.8 and 4.9, and Figure 4.5. For the majority of donors, nucleotide bases encoding amino acid residue #1188 were isotypic in nature, encoding 'Val' for C4A and 'Ala' for C4B alleles. Similarly, nucleotide bases encoding residue #1191 were isotypic in nature, encoding 'Leu' for C4A, and 'Arg' for C4B alleles (Table 4.8 and 4.9). C4B alleles of donors WCY1174, ECY 178, and ECY 302 were an exception, having a combination of residues encoded by nucleotides which are common to C4A and C4B isotypes, for residues #1188 and #1191. Anderson et al. (1992), similarly found the C4A *1 allele had nucleotide bases which encoded 'Ala' for residue #1188 and 'Arg' for residue #1191, and in the case of the C4B *5 allele, Yu et al. (1986) found 'Val' and 'Leu' were the respective amino acid residues. Of interest also here was the detection of heterozygosity for residue #1191 in donor JC003.
C4A alleles, whereas residue #1188 was homozygous in this donor. This combination of C4A and C4B residues for residues #1188 and #1191 in the same allele has not previously been reported. Usually a C4 allele will have either C4A- or C4B-type amino acid residues for residues #1188 and #1191 and not a combination of C4A and C4B residues in the one haplotype.

4.4.2.11 Amino acid residue #1267

Results of PCR-RFLP analysis of nucleotide bases encoding residue #1267 for each donor are given in Tables 4.8 and 4.9, and examples of the resulting restricted fragment banding patterns are shown in Figure 4.5. Nucleotide bases encoding this amino acid residue were relatively non-polymorphic, encoding residue 'Ala' for both C4A and C4B alleles of most donors. Nucleotide bases encoding amino acid residue #1267 were polymorphic for the C4A *3 alleles of donors ECY 858 and A *3, *6 alleles of donor JC003. The association of amino acid residue 'Ser' with A *3 alleles (Tables 4.8 and 4.9) is not surprising considering nucleotide bases encoding 'Ser' have previously been detected in cDNA sequence of both A *3b and A *4 alleles (Belt et al., 1984, and 1985). However, it is not certain if residue 'Ser' is associated with the A *3 or A *6 allele in donor JC003. The C4A alleles of donor EW-L0081785 remained unrestricted, suggesting there may be another sequence of nucleotide bases for residue #1267.

4.4.3 C4 allele-specific polymorphism

Identification of heterozygous C4A and C4B alleles by C4 protein gel electrophoresis was often difficult (Chapter 3). It was possible to assign polymorphic amino acid residues to specific C4A and C4B alleles by the analysis of: (1). Data from C4 protein gel electrophoresis (Figures 4.6, 4.7 and 4.8); (2). PCR-RFLP results from donors listed in Tables 4.8 and 4.9; (3). C4 nucleotide sequences, both from this study and that published by other researchers (Belt et al., 1984 and 1985; Yu et al., 1986; Yu, 1991; Anderson et al., 1992; Paz-Artal et al., 1993); and (4). Computer protein analysis for predicting a protein's isoelectric point (PI) according to its derived
Figure 4.6 C4B 12 and C4B 22 allotypes identified by protein gel electrophoresis. Schematic diagram showing the relative migration distances of C4 allotypes (VI complement genetics workshop reference typing), is reproduced with permission from WHO Bulletin, 70: 531-535, (1992). C4 allotypes for donor serum samples are as follows: Lanes 1, A 3 B 1 standard; 2, A 3 B 1,12 WCY1204; 3, A 4 B 1,2 ECY024; 4, A 3 B 1 standard; 5, A 3 B 1,12, WCY1204; 6, A 3,Q0 B 1 ECY851; 7, A 3 B 1 standard; 8, A 3 B 12 WCY1179; 9, A 4 B 2 (A *4 B *22,*2) JC001; 10, A 4 B 1,2 WCY1213. C4 allotypes which were different by PCR-RFLP are enclosed in brackets after the allotypes detected by protein gel electrophoresis. Samples in lane 4, 5, 6, 9, and 10 had been treated with Carboxypeptidase B prior to gel electrophoresis.
Figure 4.7 C4 allotypes identified by protein gel electrophoresis. Schematic diagram showing the relative migration distances of C4 allotypes (VI complement genetics workshop reference typing), is reproduced with permission from WHO Bulletin, 70: 531-535, (1992). C4 allotypes for donor serum samples are as follows: 1, A 3 B 1 standard; 2, A 3,Q0 B 1 WCY1159; 3, A 2,4 B 2 (A *4 B *2,*6) WCY1174; 4, A 3 B 1 standard; 5, A 2,4 B2 (A *4 B *2,*6) WCY1174; 6, A 3 B 1 standard; 7, A 2,4 B 2,4 ECY302; 8, A 3,6 B 1,2 ECY556; 9, A 3 B 1 standard; 10, A 2,4 B2,4 ECY302. C4 allotypes which were different by PCR-RFLP are enclosed in brackets after allotypes detected by protein gel electrophoresis. Samples in lane 4, 5, 8, and 10 had been treated with Carboxypeptidase B, and therefore only the most anodally travelling protein band of each C4 allele can be seen.
Figure 4.8 $C4$ allotypes identified by protein gel electrophoresis. Schematic diagram showing the relative migration distances of $C4$ allotypes (VI complement genetics workshop reference typing), is reproduced with permission from *WHO Bulletin*, 70: 531-535, (1992). $C4$ allotypes for donor serum samples are as follows: 1, $A_{4,6}B_{2}$ ($A_{*4,*58}B_{*2,*5}$) ECY178; 2, $A_{4}B_{1,2}$ ECY046; 3, $A_{3}B_{1}$ standard; 4, $A_{2,4}B_{1,2}$ ($A_{*2,*4}B_{*2,*22}$) ECY432; 5, $A_{3}B_{1}$ standard; 6, $A_{3}B_{1}$ standard; 7, $A_{4}B_{2}$ ($A_{*4}B_{*22,*2}$) JC001; 8, $A_{4}B_{2,4}$ ECY168. $C4$ allotypes which were different by PCR-RFLP are enclosed in brackets after allotypes detected by protein gel electrophoresis. Samples in lane 6, 7, and 8 had been treated with *Carboxypeptidase B* prior to gel electrophoresis.
amino acid sequence. Absence of C4A DNA contaminating the PCR reaction during C4B amplification, and C4B DNA contaminating the PCR reaction during C4A amplification, was shown by the lack of amplification of C4A and C4B isotypes DNA in homozygous C4B and C4A allele donors, respectively. PCR-RFLP results of donors with polymorphic nucleotide bases for a particular amino acid residue were repeated for confirmation. Results of these analyses are shown in Tables 4.10 and 4.11.

4.4.3.1 C4A alleles

Previous studies found the A *1 allele to be different from other C4A alleles by encoding amino acid residue 'Gly' for residue #1054, 'Ser' for #1157, 'gcc' nucleotide bases for amino acid residue #1186, as well as residues 'Ala' and 'Arg' for residue #1188 and #1191, respectively (Yu et al., 1986). Studies here found the derived amino acid sequence for the A *1 allele to have a PI of 7.16 (Table 4.10). Amino acid residues #1054 and #1191 of the A *1 allele have a different charge to these residues in other C4A alleles, and therefore were thought to be responsible for the A *1 alleles unique migration pattern by C4 protein gel electrophoresis.

This study, and that of Paz-Artal et al. (1993), found the A *2 allele to be different from all other C4A alleles by encoding residue 'Gly' for residue #1054. Examples include C4A alleles of donors ECY 504 and ECY 302 (Table 4.8). It should be noted that the A 2 allotype identified by protein gel electrophoresis was confirmed by the C4 standard reference typing laboratory in Western Australia. Three other donors in this study, namely ECY 481 with A *2,*3, ECY 581 with A*2,*4, and Caucasian 'JC004' with A *2, had amino acid residue 'Gly' for residue #1054, but these are not shown in Table 4.8. The A *2 allele's combination of amino acid residues results in a protein with a PI of 7.11, which corresponds with its migration pattern by C4 protein gel electrophoresis (Figure 4.7).
Table 4.10 Polymorphic amino acid residues for specific C4A alleles, and their predicted isoelectric points

<table>
<thead>
<tr>
<th>C4A allele</th>
<th>306</th>
<th>328</th>
<th>458</th>
<th>707</th>
<th>1054</th>
<th>1101-1106</th>
<th>1157</th>
<th>1186</th>
<th>1188,1191</th>
<th>1267</th>
<th>example donor</th>
<th>PI</th>
</tr>
</thead>
<tbody>
<tr>
<td>A*1</td>
<td>-</td>
<td>-</td>
<td>R</td>
<td>G</td>
<td>PCPVL</td>
<td>S</td>
<td>b</td>
<td>AR</td>
<td>A</td>
<td>-</td>
<td></td>
<td>7.16</td>
</tr>
<tr>
<td>A*2-</td>
<td>K</td>
<td>Y</td>
<td>R</td>
<td>P</td>
<td>PCPVL</td>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7.11</td>
</tr>
<tr>
<td>A*3a-</td>
<td>K</td>
<td>Y</td>
<td>R</td>
<td>L</td>
<td>PCPVL</td>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7.05</td>
</tr>
<tr>
<td>A*3a-</td>
<td>K</td>
<td>Y</td>
<td>R</td>
<td>P</td>
<td>PCPVL</td>
<td>N</td>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7.05</td>
</tr>
<tr>
<td>A*3b-</td>
<td>K</td>
<td>S</td>
<td>R</td>
<td>D</td>
<td>PCPVL</td>
<td>N</td>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7.05</td>
</tr>
<tr>
<td>A*4</td>
<td>K</td>
<td>S</td>
<td>R</td>
<td>P</td>
<td>PCPVL</td>
<td>N</td>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7.05</td>
</tr>
<tr>
<td>A*3b-</td>
<td>K</td>
<td>S</td>
<td>R</td>
<td>P</td>
<td>PCPVL</td>
<td>N</td>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7.05</td>
</tr>
<tr>
<td>A*3c-</td>
<td>K</td>
<td>S</td>
<td>-</td>
<td>D</td>
<td>PCPVL</td>
<td>S</td>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td>EW-L0081785</td>
<td>7.05</td>
</tr>
<tr>
<td>A*4-</td>
<td>K</td>
<td>Y</td>
<td>R</td>
<td>P</td>
<td>PCPVL</td>
<td>N</td>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7.05</td>
</tr>
<tr>
<td>A*58-</td>
<td>M</td>
<td>YorS</td>
<td>R</td>
<td>P</td>
<td>PCPVL</td>
<td>N</td>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7.05</td>
</tr>
<tr>
<td>A*6-</td>
<td>K</td>
<td>Y</td>
<td>W</td>
<td>P</td>
<td>PCPVL</td>
<td>N</td>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7.05</td>
</tr>
</tbody>
</table>

Nucleotide bases encoding the amino acid residue 'A' were 'gcc' for C4A alleles, and 'gcc' for C4B alleles. 'gcc' is represented by 'a' and 'gcc' by 'b'. C4A*1 was from genomic sequencing by Anderson et al. (1992), and Yu et al. (1986). C4A*3a and A*4 were from genomic sequencing by Yu (1991), and cDNA sequencing by Belt et al. (1984), C4A*3b and A*6 were from genomic sequencing by Anderson et al. (1992). PI represents the predicted isoelectric point by computer protein analysis. '-' indicates not determined for the respective amino acid residue.
<table>
<thead>
<tr>
<th>C4B allele</th>
<th>306</th>
<th>328</th>
<th>458</th>
<th>707</th>
<th>1054</th>
<th>1101-1106</th>
<th>1157</th>
<th>1186</th>
<th>1188,1191</th>
<th>1267</th>
<th>example donor</th>
<th>PI</th>
</tr>
</thead>
<tbody>
<tr>
<td>B*1a</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>G</td>
<td>LSPVIH</td>
<td>N</td>
<td>b</td>
<td>AR</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>B*1b</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>G</td>
<td>LSPVIH</td>
<td>S</td>
<td>b</td>
<td>AR</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>B*1cos31</td>
<td>-</td>
<td>S</td>
<td>R</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>B*1-</td>
<td>K</td>
<td>S</td>
<td>R</td>
<td>P</td>
<td>G</td>
<td>LSPVIH</td>
<td>S</td>
<td>b</td>
<td>AR</td>
<td>A</td>
<td>ECY858</td>
<td>7.23</td>
</tr>
<tr>
<td>B*12-</td>
<td>M</td>
<td>Y</td>
<td>R</td>
<td>P</td>
<td>G</td>
<td>LSPVIH</td>
<td>S</td>
<td>b</td>
<td>AR</td>
<td>A</td>
<td>WCY1179</td>
<td>7.18</td>
</tr>
<tr>
<td>B*2</td>
<td>K</td>
<td>S</td>
<td>R</td>
<td>P</td>
<td>D</td>
<td>LSPVIH</td>
<td>S</td>
<td>b</td>
<td>AR</td>
<td>A</td>
<td>-</td>
<td>7.18</td>
</tr>
<tr>
<td>B*2-</td>
<td>K</td>
<td>S</td>
<td>R</td>
<td>P</td>
<td>D</td>
<td>LSPVIH</td>
<td>S</td>
<td>b</td>
<td>AR</td>
<td>A</td>
<td>ECY028</td>
<td>7.18</td>
</tr>
<tr>
<td>B*22-</td>
<td>M</td>
<td>S</td>
<td>-</td>
<td>-</td>
<td>G</td>
<td>LSPVIH</td>
<td>S</td>
<td>b</td>
<td>AR</td>
<td>A</td>
<td>ECY248</td>
<td>7.18</td>
</tr>
<tr>
<td>B*3</td>
<td>-</td>
<td>-</td>
<td>P</td>
<td>G</td>
<td>LSPVIH</td>
<td>S</td>
<td>b</td>
<td>AR</td>
<td>A</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B*3-</td>
<td>M</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>G</td>
<td>LSPVIH</td>
<td>S</td>
<td>b</td>
<td>AR</td>
<td>-</td>
<td>RCH517</td>
<td>7.18</td>
</tr>
<tr>
<td>B*4-</td>
<td>K</td>
<td>SorY</td>
<td>R</td>
<td>-</td>
<td>G</td>
<td>LSPVIH</td>
<td>S</td>
<td>a</td>
<td>VL</td>
<td>A</td>
<td>ECY302</td>
<td>7.18</td>
</tr>
<tr>
<td>B*5</td>
<td>-</td>
<td>-</td>
<td>P</td>
<td>D</td>
<td>LSPVIH</td>
<td>S</td>
<td>a</td>
<td>VL</td>
<td>A</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B*5-</td>
<td>M</td>
<td>Y</td>
<td>R</td>
<td>P</td>
<td>G</td>
<td>LSPVIH</td>
<td>N</td>
<td>a</td>
<td>VL</td>
<td>A</td>
<td>ECY178</td>
<td>7.12</td>
</tr>
<tr>
<td>B*6-</td>
<td>M</td>
<td>S</td>
<td>-</td>
<td>D</td>
<td>LSPVIH</td>
<td>N</td>
<td>a</td>
<td>VL</td>
<td>A</td>
<td>WCY1174</td>
<td>7.07</td>
<td></td>
</tr>
</tbody>
</table>

Nucleotide bases encoding amino acid residue 'A' were 'gcg' for C4A alleles, and 'gcc' for C4B alleles. 'gcg' is represented by 'a' and 'gcc' by 'b'. C4B*1a and B*1b were from cDNA sequencing by Belt et al. (1985). B*1cos31 was from genomic sequencing by Anderson et al. (1992). B*3 and B*5 genomic sequence data was from Yu et al. (1986). C4B*2 information was from cDNA sequencing by Belt et al. (1984 and 1985). '-' indicates not determined for the respective amino acid residue.
Studies for this thesis, and other researchers, have identified a number of A *3 alleles, including A *3a~, A *3b~, and A *3c~ (found here); A *3a (Yu et al., 1986), and A*3b (Belt et al., 1984 and 1985). Published A *3a has amino acid residue 'Leu' for residue #707 (Yu, 1991), whereas A *3a~ in this study, has residue 'Pro' (Table 4.10). This study detected the published A *3b allele in donor ECY 858 (Table 4.8), which was denoted here as A *3b~. The C4A 3 allotype, determined by protein gel electrophoresis for donor ECY 858, was confirmed by the C4 standard reference typing laboratory in Western Australia. In this study the A *3c~ allele was detected in donors EW-Ducaf and EW-L0081785. This allele was different to other previously described A *3 alleles, encoding amino acid residue 'Ser' for residue #1157. The combination of amino acid residues encoded by nucleotides for each C4A allele of donor EW-Ducaf were difficult to interpret as polymorphism of residue #1054 was detected. Polymorphism of residue #1157 was only detected in A *1 and A *2 alleles. Neither of these alleles were detected by protein gel electrophoresis for these donors. Also, unrestricted DNA fragments resulting from PCR-RFLP analysis of nucleotide bases encoding residue #1267 in donor EW-L0018785, suggests other nucleotide bases encode this residue. These results suggest other types of A *3 alleles may exist.

The published A *4 allele sequence was essentially the same as published A *3b, except A *4 had amino acid residue 'Ser' instead of 'Thr' for residue #1182 (polymorphic amino acid residues of these alleles, with the exception of residue #1182, are shown in Table 4.10). Alleles A *4~ (Donor ECY 028), and A *3a~ (Donor ECY 751), in these studies, had an identical composition of amino acid residues (Table 4.10), but different migration patterns were observed by protein gel electrophoresis. The C4 allotypes of these donors were confirmed by the C4 standard reference typing laboratory in Western Australia. Computer protein analysis of the derived amino acid sequence of each type of A *3 and A *4 allele in this and other studies, found the same PI of 7.05, despite their amino acid differences.
An A *6 allele, which was previously described by Anderson et al. (1992), was found in Caucasian donor JC003 in this study (Table 4.9). Another type of C4A allele was detected in Aboriginal Australian donors WCY1158 and ECY 178 (Table 4.8), which produced a similar migration pattern as the A *6 allele by protein gel electrophoresis. This novel C4A allele, unlike the A *6 allele, had amino acid residue 'Arg' for residue #458, and 'Met' for residue #306, and was therefore denoted as an A *58~ allele. The C4A allotype of donor ECY 178, as determined by protein gel electrophoresis, was confirmed by the C4 standard reference typing laboratory in Western Australia, and was found to migrate between the A 5 and A 6 alleles. Unfortunately, the serum sample of donor WCY1158 upon arrival in Western Australia was degraded, and therefore could not be allotyped by protein gel electrophoresis. In this study, the derived amino acid sequence of A *6~ and A *58~ alleles produced the same PI value of 7.0, which corresponded to the migration patterns of these alleles by protein gel electrophoresis (Figure 4.7 and 4.8).

4.4.3.2 C4B alleles

The B *1~ allele in Aboriginal Australian and Caucasian donors (Table 4.11), was found here to have the same amino acid residues as the published B *1b allele (Belt et al., 1985). B *1a, described by Belt et al. (1985), was different to the B *1~ alleles studied here, as nucleotide bases from Belt's study encoded 'Asn' instead of 'Ser' for residue #1157. Computer protein analysis of the B *1~ allele found a PI value of 7.23. Substituting amino acid residue 'Asn' for 'Ser' for residue #1157 did not alter the PI value of this allele. This result is not unexpected considering both types of amino acid residues #1157 were neutral-polar, and therefore were not expected to result in a differently charged protein or to affect the migration of the protein by gel electrophoresis.

The B *2~ allele in this study (Table 4.11), had an identical polymorphic amino acid residue composition as the published B *2 allele (Belt et al., 1984, and 1985; Yu, 1991; Anderson et al., 1992). The B *2~ allele was different to the B *1~ allele by having nucleotide bases for amino acid residue 'Asp' instead of 'Gly' for
residue #1054. This amino acid residue change, results in a charge change from an acidic to a neutral, non-polar residue, and therefore is thought to be responsible for the \( B^{*2-} \) allele having a PI of 7.18, which corresponds to its migration pattern by protein gel electrophoresis. Examples of the electrophoresis migration pattern of the 'B 2' allele are shown in Figures 4.6, 4.7, and 4.8.

The rare \( C4B \) allele, \( B^{12} \) (Chapter 3; Ranford et al., 1987), for which the molecular basis had not previously been described, was found to migrate between the \( B^{1} \) and \( B^{2} \) alleles by protein gel electrophoresis (Figure 4.6). This migration pattern for donors WCY1179 and WCY1204 was confirmed by the \( C4 \) standard reference typing laboratory, Western Australia. In these studies, this allele was found to have amino acid residue 'MET' for residue #306 in long \( C4 \) alleles, unlike \( B^{*1-} \) and \( B^{*2-} \) alleles which have 'Lys' for this residue. The \( B^{*12-} \) allele had amino acid residue 'Gly' for residue #1054 as did the \( B^{*1-} \) allele. The PI value for \( B^{*12-} \) was found to be the same as the \( B^{*2-} \) allele (7.18). Donor WCY1204, who was heterozygous at the \( C4B \) locus (\( B^{*1,*12} \)), could be distinguished by gel electrophoresis. However, the resulting \( C4B \) protein bands merged into each other making interpretation of the \( C4 \) allotype by this method difficult (Figure 4.6).

Another \( C4 \) novel allele denoted as \( B^{*22-} \), was detected by PCR-RFLP analysis in donors ECY 248, ECY 432 and JC001. Donors ECY248 and ECY432 were allotyped as \( B^{1,2} \), and JC001 as \( B^{2} \) by protein gel electrophoresis. Donor's ECY 248 and ECY 432 allotypes were confirmed by the \( C4 \) standard reference typing laboratory in Western Australia. The polymorphic amino acid residue composition of the \( B^{*22-} \) allele was found to be similar to the \( B^{*12-} \) allele. However, unlike the \( B^{*12-} \) allele, \( B^{*22-} \), had 'Met' for residue #306 in the short \( C4 \) allele instead of the long \( C4 \) allele (Table 4.5). All other residues in the \( B^{*22-} \) allele were identical to those of the \( B^{*12-} \) allele, except for residue #328 where \( B^{*12-} \) had 'Tyr', and \( B^{*22-} \) had 'Ser'. Polymorphic residues for #328 do not result in a differently charged \( C4 \) protein. However, amino acid 'Tyr' is larger than 'Ser', and may slightly affect the mobility of the \( B^{*22-} \) protein migrating more closely to \( B^{2} \) than \( B^{12} \) by protein gel
electrophoresis. The $B^{*22-}$ allele was identified by the detection of heterozygosity of nucleotides encoding residue #306 in the short $C4$ genes PCR product. This indicated that two different short $C4$ alleles were present in donors ECY248 and ECY432. Both short $C4B$ alleles have 'Ser' for residue #328, whereas the long alleles in these donors have 'Tyr'. Taq I RFLP analysis of donor ECY 248 (Table 4.7), found long $C4$ genes at locus I and II, as well as short $C4$ genes at locus II. Taq I RFLP analysis together with PCR-RFLP analysis of the same donor suggests there may be two types of short $C4$ genes, and a long $C4$ gene at locus II on the same haplotype. $C4$ protein gel electrophoresis of donor ECY 248 found $A^{4B^{1,2}}$. It is possible that another $C4B$ allele, such as $B^{*1}$ or $B^{*12-}$, is masked by the other $C4B$ alleles by PCR-RFLP analysis. Unfortunately, Taq I RFLP analysis data is not available for donors ECY 432 and JC001 to determine if a similar gene organisation to ECY 248 is present. Allotypes $A^{2,4B^{1,2}}$ of donor ECY432, $A^{4B^{1,2}}$ of donor ECY248, and $A^{4B^{2}}$ of donor JC001, as determined by protein gel electrophoresis, suggest that the $B^{*22-}$ allele is not visible by gel electrophoresis and co-migrates with $B^{2}$ (Figure 4.6 and 4.8, donor ECY 248 not shown).

Donor RCH517 was found by protein gel electrophoresis to have a $B^{3}$ allele which had the same polymorphic residues (#1054 to #1186) as those of the published $B^{*3}$ allele (Yu et al., 1986). Both PCR-RFLP and sequence analysis of $B^{*3-}$ from donor RCH517 found 'Met' and 'Tyr' were encoded by nucleotides for residues #306 and #328, respectively. By assuming the published $B^{*3}$ allele and the $B^{*3-}$ allele described in this study were the same allele, the PI of this allele was estimated to be 7.18, which is the same as for the $B^{*2-}$ allele. However, there is evidence suggesting that these two types of $B^{*3}$ alleles are not the same. The $B^{*3-}$ allele described in this study, has 'Tyr' for residue #328, and was found to be a short $C4$ allele by PCR and sequence analysis. The published $B^{*3}$ allele, however, was a long $C4$ allele (Yu et al., 1986), and nucleotide bases encoding residue #328 were not reported. Those nucleotide bases which have been analysed for the $B^{*3-}$ allele in this study, have the same composition as the $B^{*12-}$ allele. However, their migration
patterns are different by protein gel electrophoresis. Also, PCR-RFLP analysis has identified two types of short $C4B$ alleles in donor RCH517, whereas protein gel electrophoresis only detected one type of $C4B$ allele.

Donor 'ECY 302', was found to be heterozygous at the $C4B$ locus by protein gel electrophoresis, having both $B4$ and $B2$ alleles (Figure 4.7). The $B4$ allele was difficult to identify by having a similar migration pattern to the $A1$ allele by protein gel electrophoresis. The $C4$ standard reference typing laboratory, Western Australia, confirmed that donor ECY302 had a $B2$ allele, however, the second $C4B$ allele of this donor was reported to be haemolytically active (indicating a $C4B$ allele and not $C4A$ allele), and migrates more anodally than the $B3$ allele by protein gel electrophoresis. PCR-RFLP analysis of $C4A$ and $C4B$ alleles separately, of donor ECY302, enabled the $B*4~$ and $B*2~$ alleles to be easily identified (Table 4.8). Polymorphic residues, which were different to the well-described $B*2$ allele, were assumed to be possessed by the $B*4~$ allele. The polymorphic amino acid residue composition of the $B*2~$ and $B*4~$ alleles was different. However, the PI value from the derived protein sequence for each allele was the same. The same PI value for these two alleles was not consistent with their different migration patterns by gel electrophoresis (Figure 4.7).

Donor ECY 178 was found to have the allotype $B2$ using the traditional method of protein gel electrophoresis (Figure 4.8). This was confirmed by the $C4$ standard reference typing laboratory in Western Australia. However, PCR-RFLP analysis, found heterozygosity of polymorphic residues at the $C4B$ locus. Polymorphic residues from the well-described $B*2$ allele were assigned as $B*2$ residues. The remainder were assigned to a second $C4B$ allele, were collectively analysed, and found to have a PI value of 7.12 (Table 4.11). According to the PI of other $C4$ alleles, this allele with the PI value of 7.12 would migrate between an $A1$ and an $A2$ allele, and therefore was denoted as $B*5~$. In comparison with the $B*5$ allele described by Yu et al. (1986), the $B*5~$ allele from this study was different, encoding amino acid residues 'Gly' and 'Asn' for residues #1054 and #1157, respectively (Table 4.11).
Donor 'WCY1174', was thought to have the $C4$ allotype $A\, 2,4\, B\, 2$, as determined by protein gel electrophoresis (Figure 4.7). However, PCR-RFLP allototyping detected heterozygosity of residues at the $C4B$ locus (Table 4.8). Nucleotide bases encoding amino acid residues common to the $B\,*2$ allele were assigned as $B\,*2~$ residues, whilst the remaining residues were collectively analysed and found to have a PI of 7.07. This PI value corresponded to the migration pattern produced by a $C4B$ allele migrating between an $A\, 2$ and an $A\, 3$ allele, and was therefore denoted as a $B\,*6~$ allele. By re-analysing the allotype of donor WCY1174 as determined by protein gel electrophoresis (Figure 4.7), together with PCR-RFLP allotyping and the PI value, it was possible to determine the $C4$ allotype of this individual as $A\,*4\, B\,*2,\,*6$.

4.4.3.3 Long and short $C4$ genes

For some alleles in this study, it was possible to identify whether they were of the long or short $C4$ gene type. Where possible, this information together with published data regarding $C4$ gene size of different alleles was used to establish the amino acid composition of some $C4$ alleles. Results of PCR-RFLP analysis in this study suggested there was an association between long $C4$ genes (23 kb) and nucleotide bases encoding the amino acid 'Tyr' for residue #328. For the majority of donors studied, those alleles amplified using the PCR primers which only amplified short $C4$ genes were found to have nucleotide bases for amino acid 'Ser' for residue #328. For those alleles amplified using the primers which amplify both long and short $C4$ genes, amino acid residues 'Tyr' and 'Ser' were usually encoded by nucleotides for residue #328. It was therefore not possible to determine if residue 'Ser' was also encoded by the long $C4$ genes.

Donor ECY 178 was found to have two short $C4B$ alleles ($B\,*2~,*5~$) by PCR-RFLP analysis and one $C4B$ allele ($B\,2$) as determined by protein gel electrophoresis. Heterozygosity of residues #306 and #328 were detected in the short $C4$ gene PCR product (Table 4.5). Nucleotide bases encoding amino acid 'Tyr' were thought to be those of the $B\,*5~$ allele, as only nucleotides encoding amino acid 'Ser' had
been detected in \( B^{*2-} \) alleles in this study and elsewhere (Belt et al., 1984 and 1985). The \( B^{*5} \) allele described by Yu et al. (1986), was a short allele. The short \( B^{*5-} \) allele in this study was different to other short \( C4 \) alleles by having amino acid 'Tyr' for residue #328. The only other short \( C4 \) allele encoding amino acid residue 'Tyr' was the \( C4A^{*CAN1-} \) allele described in Chapter 5. The \( C4 \) gene region encoding polymorphic amino acid residues #306 and #328 had not been sequenced for the majority of published \( C4B \) alleles (Belt et al., 1985; Yu et al., 1986), and therefore few comparisons could be made with the literature (see Table 4.11 as a summary).

Short \( C4A \) alleles have been reported for allotypes \( C4A^{*35} B^{*5} \) (Schneider, 1990), and \( C4A^{*11,*6} B^{*Q0} \) (Giles et al., 1987), by RFLP analysis of the \( C4 \) gene region. Short \( C4A \) alleles were found to occur in only one of 103 donors (Schneider, 1990). Two published \( C4A \) alleles, namely \( A^{*3b} \) and \( A^{*4} \) (Table 4.10), had amino acid 'Ser' for residue #328. Whether these alleles were long or short has not been reported. The \( A^{*3b-} \) allele detected in donors ECY 858 and JC003 in these studies, was found to encode amino acid residue 'Ser' for residue #328 and 'Ser' for residue #1267. \textit{Taq I} RFLP analysis would be required to determine if the \( A^{*3b-} \) allele is a long or short type of \( C4 \) gene. This hypothesis has not yet been tested due to a shortage of donor genomic DNA. It has been reported that the \( A^{*l} \) and \( A^{*3a} \) alleles are of the long \( C4 \) gene type (Yu et al., 1986).

Donors WCY1204 and WCY1179 had only long \( C4 \) genes, as amplification of the short \( C4 \) gene using primer pair 1 (Table 4.1), was unsuccessful, and \textit{Taq I} RFLP analysis of donor WCY1204 genomic DNA found only long \( C4 \) genes at both loci I and II (Table 4.7). Each \( C4 \) allele in these donors had amino acid 'Tyr' for residue #328 (Table 4.5).

4.5 Discussion

\( C4 \) allotypes (defined in Chapter 3), as determined by \( C4 \) protein gel electrophoresis, provided a useful resource for the development of a \( C4 \) molecular allotyping protocol. This study found more polymorphism of \( C4 \) allotypes by PCR-RFLP analysis in comparison
to protein gel electrophoresis. However, as a result of gene duplication, it is possible that some $C4$ alleles are overlooked by PCR-RFLP analysis, and therefore one must rely on either Taq I RFLP or PFGE to determine the gene organisation in these donors. $C4A$ alleles $A^{*}1$, $A^{*}2$, $A^{*}3a$, $A^{*}3c$, $A^{*}58$, $A^{*}6$, and $C4B$ alleles $B^{*}1$, $B^{*}12$, $B^{*}22$, $B^{*}2$, $B^{*}4$, $B^{*}5$ and $B^{*}6$, can now be allotyped individually using PCR amplification and restriction analysis by utilising the isotypic region of $C4$ genes and differential size of $C4$ intron 9. This procedure enables heterozygous alleles at loci I and II of the $C4$ gene region to be easily identified, which had previously been misinterpreted or undetected by the traditional method of protein gel electrophoresis. Although this new molecular $C4$ allotyping protocol cannot distinguish between all $C4$ alleles, it has proven to be useful in the identification of a number of rare and common $C4$ alleles whose molecular basis has not previously been established. These include, $B^{*}12$, $B^{*}22$, $B^{*}4$, $B^{*}5$, $B^{*}6$, and $A^{*}58$.

A number of notable results were found by sequence analysis of the short $C4$ alleles, including: (1). The identification of two additional polymorphic nucleotide bases encoding residues #306 and #309; (2). Identification of polymorphic nucleotide bases in intron 9 which were unique to short $C4$ alleles; and (3). Location of a single copy of the mobile genetic element previously reported by Yu, (1991), which borders the long intron 9.

Of particular significance from sequencing analysis was the association of the polymorphic residue #306 with the $C4A^{*}CAN1$ allele of donor H14 and the rare $B^{*}3$ (RCH517a) and $B^{*}5$ alleles. As a result, this residue which could be identified by restriction analysis with enzyme $Alu I$, was included in the PCR-RFLP $C4$ allotyping protocol. Polymorphism of residue #306 proved to be very important in the identification of other rare $C4$ alleles including $A^{*}58$ and $B^{*}6$. Alleles which had nucleotide bases encoding the residue 'Met' for residue #306, migrated more anodally than those alleles with a similar composition of amino acid residues but had 'Lys' for this residue, for example, $B^{*}1$ and $B^{*}12$ alleles. This observation is not unexpected, as nucleotide base differences of residue #306
resulted in a charge change from a basic 'Lys' residue to a non-polar 'Met' residue, thereby altering the overall charge of the C4 allele and therefore its migration pattern by protein gel electrophoresis.

Polymorphism of nucleotide bases encoding amino acid residue #309 resulted in a silent alteration of the amino acid residue, and therefore was not expected to greatly affect the migration pattern of each C4 allele. Unfortunately this polymorphic residue could not be recognised by restriction enzymes which are available. Residue #309 is closely situated to the polymorphic residue #306, and sequencing results indicate that the two residues are linked. As a consequence, the polymorphic residue #309 was not included in the PCR-RFLP allotyping protocol.

Of significance also, sequence analysis of the short C4 alleles, identified a number of polymorphic nucleotide bases in intron 9 which distinguished between short and long C4 alleles. One of these polymorphic bases is situated alongside the mobile genetic element in intron 9, namely base #280. Short C4 alleles in this study had nucleotide base 'cysteine' for base #280, compared to the long C4A *3a allele described by Yu, (1991), which had 'thymine' for base #280. Studies for this thesis have shown that the polymorphic intron 9 base #280 is not universally linked with nucleotide bases encoding polymorphic amino acid residue #328. This is because short C4 genes (with nucleotide base 'cysteine' for base #280), which were sequenced here, encoded both amino acid residues 'Ser' and 'Tyr' for residue #328. However, intron 9 base #280 may be specifically associated with the size of a C4 allele, as nucleotide base 'thymine' was consistently found at base #280 in each short C4 gene.

PCR-RFLP, Taq I RFLP and sequencing analysis of short and long C4 genes described above, suggest that the mutation creating polymorphism of nucleotide base #280 predated the mutation creating polymorphism of nucleotides encoding residue #328 in short C4 genes. The polymorphic base #280 could potentially be used for specific amplification of either long or short C4 alleles. In turn, it could be used for linkage analysis of polymorphic nucleotide bases encoding amino acid residues #328 and #306 to either long or short
C4 alleles. Nucleotide base #280 would be the least likely polymorphic nucleotide base, situated in intron 9, to undergo recombination with a long C4 allele. This is because other polymorphic nucleotide bases are located further from the mobile genetic element in which the large 7 kb intron inserts. Therefore, intron 9 base #280 would be the most reliable marker for long and short C4 alleles. A study to substantiate these findings should be given high priority in further development of this molecular C4 allotyping protocol based on PCR-RFLP.

From this study it can be seen that each allele is a composition of polymorphic nucleotide bases encoding amino acid residues which have been derived from ancestral C4 alleles. It is this combination of residues which is responsible for each allele's net charge, and unique migration pattern produced by protein gel electrophoresis.

A number of C4 alleles having the same predicted PI but a different combination of polymorphic amino acid residues were identified. C4A alleles, A *3a, A*3a~, A *3b, A *3c~, and A *4, all had the same predicted PI of 7.05, and produced indistinguishable migration patterns by protein gel electrophoresis, with the exception of A *4 which migrated slightly more anodally than did the A 3 allele. C4B alleles B *12~, B *2, B *22~, B *3~, and B *4, on the other hand, had predicted PI values of 7.18 but migrated to different positions by protein gel electrophoresis, except for B *2 and B *22~ which co-migrated. These results suggest additional, undetected polymorphisms exist in C4 alleles with the same PI. Also, it is possible that the unique combination of polymorphic amino acid residues, in the C4 alleles having the same PI, result in different structural conformations of the protein. In turn, the different conformation may affect the migration patterns of these alleles. A number of researchers have found that different alleles with the same net charge can be resolved by isoelectric focusing. For example, different mouse haemoglobins for which the α-chain composition is well known, were resolved with a high-voltage isoelectric focusing system, even though their proteins had the same isoelectric point (Whitney III et al., 1979). It is quite possible that proteins with the
same net charge but different size may be resolved by protein gel electrophoresis as this system, like isoelectric focusing, resolves proteins according to their net charge and size.

In the case of the C4A group of alleles mentioned above, each allele had a different combination of polymorphic amino acid residues. However, these differences were only between small, neutral, non-polar amino acid residues which do not alter the overall structure and charge of a protein, and therefore would not cause a large shift in the migration of a protein. The only factor which may alter the migration of these proteins by protein gel electrophoresis is if additional polymorphic amino acid residues were identified which were different in size and/or charge. This could affect the overall conformational structure and migration of these proteins. For example, amino acid 'Arg' has a side chain consisting of three hydroxyl groups terminating in a basic group, compared to amino acid 'Gly' which is the smallest and simplest neutral, non-polar amino acid residue (Watson et al., 1987).

The structure and net charge of amino acid residues #306, #1054 and #1191 were different for each C4B allele described above. However, despite these differences the predicted PI (7.18) of each allele was the same although the proteins from these alleles were found to have different migration patterns by protein gel electrophoresis. It is possible that the different structure of some polymorphic amino acid residues encoded by nucleotides of these alleles are solely responsible for the unique migration pattern of these alleles despite their similar PI. However, as the entire sequence of each allele has not been determined, it is also possible that other polymorphic amino acid residues which will alter the PI of these alleles have not as yet been identified.

Summarising the above results and observations, there are two groups of C4 alleles which have the same predicted PI. The first group has a PI of 7.05, and different polymorphic amino acid residues which are similar in size and charge, and migrate to the same or similar position by protein gel electrophoresis. The second group of alleles has the same PI which is 'basic' in charge, and
different amino acid compositions which are different in size and charge, and can be resolved by protein gel electrophoresis. It is most likely, that the different electrophoretic mobility patterns of these alleles with the same PI is probably a reflection of the different sizes of amino acid residues which they have.

Considering there are other $C4$ alleles which have not been studied at the molecular level and that other polymorphic amino acid residues may exist, in this study, a close association was often found between $C4$ allotypes identified by protein gel electrophoresis and $C4$ allotypes identified by PCR-RFLP. In addition, a close association was found between the PI of many $C4$ alleles compared to these alleles migration patterns produced by protein gel electrophoresis. These close associations suggest that the majority of polymorphic amino acid residues which are responsible for the unique migration patterns of the different $C4$ alleles have most probably been analysed in these studies.

A number of $C4$ allotypes such as $C4A\ 51, 14, 13, 11$ and $C4B\ 51, 31, 29, 21, 91$, which were identified by protein gel electrophoresis, and described in the 1983 nomenclature statement (Mauff et al., 1983), are being reviewed. Some of these $C4A$ and $C4B$ alleles are thought to be one and the same. For example, $C4B\ 31$ is now considered to be identical to $C4A\ 91$. Also, some alleles such as $B\ 29$ and $B\ 3$ are now recognised as the same $C4B$ allele (WHO-IUIS Nomenclature Sub-Committee, 1993). Application of the $C4$ molecular typing protocol based on PCR-RFLP will help overcome these inconsistencies.

Development of the $C4$ molecular allotyping protocol described in this thesis is at an early stage, where polymorphism between $C4$ alleles has been detected, and some new alleles have been identified. However, additional sequencing is required to reveal other polymorphic nucleotide bases which will distinguish between those $C4$ alleles which have the same combination of polymorphic residues.
The C4 PCR-RFLP allotyping protocol has been designed in this study to type the C4A and C4B alleles separately, and thereby decrease the probability of misinterpreting heterozygous alleles. However, gene duplication causes problems with the interpretation of heterozygotes at a particular C4A or C4B locus. The use of Taq I RFLP analysis and pulsed-field gel electrophoresis (PFGE), may be used to determine the number of C4 genes present in a donor, prior to PCR-RFLP analysis.

Another problem which will need to be resolved with the PCR-RFLP typing protocol is the masking effect of multiple alleles having the same amino acid residues for residues #306 and #328. These residues are situated too far from the isotypic region to enable analysis of C4A and C4B alleles separately. Therefore, it appears in these cases that in addition to PCR-RFLP analysis, protein gel electrophoresis allotyping is required to ensure all C4 alleles are identified. For example, in donor ECY 178, the A *58 allele is masked by the B *5 allele for residue #306. However, for future use of the PCR-RFLP typing protocol it would be more appropriate to link the polymorphic residues #306 and #328 to the C4 isotypic region and intron nucleotide base 280 which determines the different size of C4 alleles.

Other techniques such as allele specific oligonucleotide (ASO) hybridisation (ASO), and single stranded conformational polymorphisms (SSCP), may have been used to develop a C4 molecular allotyping protocol. However, for the majority of C4 alleles described in Tables 4.10 and 4.11, only A *6, and A *3a could be identified by hybridisation of a single ASO. For other C4 alleles described in Tables 4.10 and 4.11, multiple ASO's would need to be used to enable the C4 allotype of an individual to be determined.

It would be possible to develop a protocol similar to the PCR-RFLP protocol described here, which is based on polymorphism of nucleotide bases found in introns of the C4 gene. This study has identified a number of nucleotide base differences within intron 9. However, many of these base differences distinguish between the long and short form of this intron in C4 genes, and are not C4
allele-specific. The majority of published C4 nucleotide base sequences have been obtained by sequencing cDNA, which excludes intron sequence, and therefore little information concerning the polymorphism of nucleotide bases in C4 introns is known. Sequencing of introns from different C4 alleles is required to identify polymorphic nucleotide bases which could then be used to develop a molecular-based C4 allotyping protocol based on intron polymorphic nucleotide bases. However, as the C4 gene is very large (either 23 kb or 16 kb), it is highly likely that many researchers will continue to sequence cDNA in preference to genomic DNA (containing both intron and exon sequence) where the entire C4 gene is encoded by 5.2 kb of DNA sequence.

An alternative to ASO and PCR-RFLP analysis would be the use of SSCP for a C4 molecular typing technique. Amplified PCR fragments which are heterogeneous have different nucleotide base sequence, and would therefore migrate differently in their single stranded forms. The main disadvantage of this technique is that it does not determine which particular nucleotide base is polymorphic as does the PCR-RFLP typing protocol. However, new polymorphic nucleotide bases can be detected by SSCP which may not be detected by PCR-RFLP analysis. PCR-RFLP is directed at identification of a specific sequence of nucleotide bases. However, it is possible for PCR-RFLP to identify other polymorphic bases by heteroduplex formation between restricted fragments having different nucleotide base sequence.

A number of other known nucleotide base sequences encoding polymorphic amino acid residues (Yu., 1991), were not included for PCR-RFLP analysis in this study. Belt et al. (1984), found that amino acid residue #1478 was polymorphic in the C4A *4 allele. Each residue encoded by nucleotides for residue #1478 was differently charged and could therefore play an important role in establishing different migration patterns between the A *4 and A *3 alleles. Amino acid residue #1478 was not included in this study due to the unavailability of a restriction enzyme which could identify the nucleotide base sequence encoding this polymorphic residue. Also, C4 isotypic amplification of the nucleotide base sequence encoding
this residue would be difficult as it is situated 3.36 kb from the isotypic region. An oligonucleotide, specific for nucleotide bases encoding this residue, would be required for hybridisation to different C4 alleles to establish if this residue is polymorphic.

Nucleotide bases encoding amino acid residue #399, could potentially be useful in the identification of the A *4 allele by PCR-RFLP analysis. Nucleotide bases encoding this residue were found by Yu (1991), to be polymorphic in the C4A *4 allele, encoding 'Ala' instead of 'Val' which is seen in other C4 alleles. Both of these amino acid residues are non-polar. They would therefore not be expected to cause a large shift in the migration pattern of this allele by protein gel electrophoresis, but rather may be responsible for a smaller shift similar to that observed between A 3 and A 4 alleles.

PCR-RFLP analysis, using primers specific for the sequence region surrounding intron 9, enabled identification of the size (either 23 kb or 16 kb) of some C4 alleles where the alleles were homozygous. For example, the length of the C4B *12 allele, and C4A and C4B alleles of donor ECY858, was determined by the presence or absence of PCR amplification using primer pairs 1 and 2. When alleles were in a heterozygous form, the length of each C4 allele could not be determined with any degree of certainty. Sequencing from polymorphic nucleotide bases of exon 9, through intron 9, and to the isotypic nucleotide bases of the C4d region, would be the most accurate way to determine the length of each specific allele. If intron 9 nucleotide base #280 is specifically associated with the length of a C4 allele, polymorphism of this nucleotide base may be useful for establishing the association of other polymorphic amino acid residues in this region, with specific C4 alleles.

The C4 PCR-RFLP allotyping protocol is advantageous as resulting DNA fragments are clearly interpreted, and the protocol may be easily implemented in any laboratory environment avoiding the problems associated with plasma samples containing unstable C4 protein.
In conclusion, development of the molecular C4 allotyping protocol based on PCR-RFLP analysis has proven to be valuable by facilitating the determination of the molecular basis of a number of rare C4 alleles (B *12~, B *22~, B *4~, B *5~, B *6~ and A *58~). These alleles together with more commonly occurring C4 alleles can be detected by PCR-RFLP analysis without the more technically difficult methods of C4 protein gel electrophoresis and haemolytic overlays. Alleles which are in a heterozygous state at a single C4 gene locus, can be easily and accurately identified. In time, when other polymorphic nucleotides are identified, other C4 alleles may be allotyped without the need for traditional protein typing techniques.
Chapter 5
Molecular basis of C4 null alleles in Aboriginal Australians

5.1 Introduction

Characterisation of C4 null alleles at the molecular level will advance the understanding of their association with disease susceptibility. Their identification will also assist in providing a more rapid and accurate method of C4 allotyping these alleles at the molecular level compared to the traditional method of protein gel electrophoresis and immunofixation. Earlier in this thesis (Chapter 4), it was found that the traditional method of protein gel electrophoresis does not always accurately identify the C4 allotype of an individual. However, polymorphism within the C4 gene has enabled the identification of many of these C4 alleles at the molecular level by PCR-RFLP analysis (Chapter 4). It is important that C4 null alleles be included in the PCR-RFLP typing protocol due to their association with disease and their high allelic frequency within the general population.

There is a high frequency of the C4 null allotypes C4A Q0 (5-15 %) and C4B Q0 (10-20 %), in the Caucasian population (Schendel et al., 1984; Partanen and Koskimies, 1986a and b). Frequencies of C4 null alleles vary between different ethnic populations, with the highest frequency being detected in an Aboriginal Australian population, where C4A Q0 alleles approached 30 % and C4B Q0 alleles approached 25 % (Ranford et al., 1987). These unusually high frequencies of C4 null alleles suggest that either partial C4 deficiency is not a major health risk, or that C4 null alleles detected by protein gel electrophoresis and immunofixation are incorrectly defined. The second hypothesis proposes that a number of alleles designated as C4 null alleles, are functional alleles but their true identity has been misinterpreted by C4 protein typing.

As previously described in Chapter 1, C4 null alleles have been found to eventuate from: (1) Deletion of a C4 locus which usually occurs together with a 21-hydroxylase gene (Carroll et al., 1985a, c); (2) Non-expression of a defective C4A mutant gene, resulting from a 2 bp insertion into exon 29 (Barba et al., 1993); and
(3) Isoexpression (Raum et al., 1984; Rittner et al., 1984; Uring-Lambert et al., 1984), or homoexpression of identical C4 isotypes on the same haplotype (Yu and Campbell, 1987).

There appears to be a common type of C4 null allele in the Caucasian population, where approximately 60% of null alleles were found to result from deletion of a C4 gene together with an adjacent 21-hydroxylase gene (Schneider et al., 1986). C4A null alleles in Caucasian populations have been found to be strongly associated with SLE (Ng and Walport, 1988). A worthwhile study would be to establish if C4A Q0 alleles in Aboriginal Australians, (who have the highest recorded frequency of this allele), result from large deletions of the C4 gene region, and whether these alleles are associated with SLE or other disease states.

Where C4 null allele allotypes result from homoexpression and/or isoexpression, protein gel electrophoresis detects C4 protein in either the anodal or cathode region of the gel, as these alleles migrate to the same region. As a result of the common migration pattern produced by each allele, this type of C4 allotype simultaneously results in a C4 null allotype at the adjoining region of a gel, even though both C4 gene loci contain functional alleles. Examples of this phenomenon have been reported, such as, isoexpression of the C4 allotype C4 A *2,*3 B *Q0, found in Norway and Germany (Fredrikson et al., 1991). This allotype was found to be most commonly a part of the extended haplotype HLA-A 3, Cw 4, B 35, C4A 3, C4B Q0, Bf F, C2 C, DR 1. Isoexpression and homoexpression have only been detected with C4A alleles, where a single haplotype has two long C4A alleles but no C4B allele (Partanen and Campbell, 1989; Welch et al., 1989; Braun et al., 1990).

C4 alleles which have a combination of C4A and C4B type polymorphisms have been associated with the C4 null allotype (Rittner et al., 1984; Raum et al., 1984; Uring-Lambert et al., 1984; Partanen and Campbell, 1989). One individual for example, had the allotype A 3 B Q0, and had long C4 genes at both loci I and II (Partanen and Campbell, 1989). At the molecular level, this individual was found to have a gene at locus II which had a C4A
isotype but antigenic determinants more typical of \textit{C4B} genes. This unusual allele was associated with one of the following haplotypes HLA-\textit{B 44}, \textit{Bf S}, \textit{C4A 3}, \textit{C4B Q0}, \textit{DR 4}, or HLA-\textit{B 62}, \textit{Bf S}, \textit{C4A 3}, \textit{C4B Q0}, \textit{DR 4}. The specific haplotype having this allele could not be determined as other family members were not informative for determining the inheritance of each allele. Thomson \textit{et al.} (1988), detected the same haplotype HLA-\textit{B 44}, \textit{C4A 3}, \textit{C4B Q0}, \textit{DR 4} in association with Felty's syndrome. The \textit{C4} gene region of this individual was found to have a duplicated set of \textit{C4A} alleles which was shown to produce functional \textit{C4A} protein. The \textit{C4A} alleles were only detected using monoclonal antibodies specific for the isotypic region of the gene. Therefore, whether the antigenic determinants, of each of these \textit{C4A} alleles, were \textit{C4A}- or \textit{C4B}-like were unknown. Both haplotypes described by Partanen and Campbell (1989), and Thomson \textit{et al.} (1988), may have the same unusual \textit{C4A}/\textit{C4B} hybrid-type allele, which was shown to be functional by Thomson \textit{et al.} (1988). This unusual type of \textit{C4}-hybrid allele is associated with the \textit{C4} null allotype as determined by protein gel electrophoresis, however, the molecular basis of this association has not been resolved and requires further investigation.

Barba \textit{et al.} (1993), found three types of nondeleted \textit{C4A} null alleles which were described in Chapter 1 (Section 1.11.2). In summary, the first type of \textit{C4A} null allele had a 2 bp insertion into exon 29 which leads to a termination codon during transcription. This type of \textit{C4A} null allele was found to be associated with the haplotype HLA-\textit{B 60}, \textit{DR 6}. Barba \textit{et al.} (1993) found evidence for gene conversion to the \textit{C4B} isotype in the second type of \textit{C4A} null allele. The molecular basis for the third type of nondeleted \textit{C4A} null allotype has not yet been determined.

Summarising some of the more significant findings of these earlier studies described above; (1) The \textit{C4} null allotype is quite common in the general population, and occurs with relatively high frequencies in some specific ethnic groups such as Aboriginal Australians; (2) The molecular basis for a number of \textit{C4} null allotypes has been determined, and other types of \textit{C4} null alleles have been identified which require further investigation; (3) These earlier studies have
provided evidence which questions the association of certain types of $C4$ null alleles with the $C4$ null allotype as determined by protein gel electrophoresis; these null alleles possibly being functional alleles with a $C4$ null allele allotype.

5.2 Aim

To determine the molecular basis of $C4$ null allele allotypes in Aboriginal Australians, and to use this knowledge for identification of $C4$ null alleles which can be incorporated into the molecular-based $C4$ allotyping protocol.

5.3 Materials and Methods

5.3.1 $C4$ null allele donors for genetic analysis

Aboriginal Australian $C4$ null allele donors were used in this study due to the high frequency of this allotype, as well as the high prevalence and severity of SLE in Aboriginal Australian populations. Also, the molecular-basis of $C4$ null alleles in this ethnic group is unknown. Peripheral blood samples from donors H11, H14 and H19 were collected from Darwin, Northern Territory; donors denoted 'ECY' and 'WCY' were Aboriginal samples from the East and West Cape York region of Australia, respectively. Sample numbers 30017 to 40045 were Aboriginal samples from Groote Island. A series of 50 SLE Caucasian patients numbered SLE502 to SLE686 were from Canberra. IDDM patient cell lines were provided by the Royal Childrens Hospital (RCH) in Melbourne, and Woden Valley Hospital (ASD), Canberra. 10th International Histocompatibility Workshop cell line samples were denoted as 'EW'.

5.3.2 $C4$ protein gel electrophoresis and immunofixation

Samples from Aboriginal Australians denoted as 'ECY' and 'WCY' were $C4$ allotyped as described in section 2.5. RCH and ASD patient samples were $C4$ allotyped using the same technique by Pam Ranford, John Curtin School of Medical Research (JCSMR). Darwin Aboriginal samples and Groote Island Aboriginal Australians having
null allele allotypes were $C4$ allotyped by Heather Dunckley, JCSMR. SLE samples were $C4$ allotyped by Heather Dunkley and Elizabeth Jazwinska, JCSMR.

5.3.3 Genomic DNA extraction

Genomic DNA from donor buffy coats or cell lines were extracted according to the method described in section 2.7.

5.3.4 $Taq$ I restriction analysis of genomic DNA, and $C4$-specific hybridisation

Approximately 3 μg of genomic DNA was restricted with 30 units of $Taq$ I restriction enzyme as previously described in Chapter 4, section 4.3.4.

5.3.5 Gel electrophoresis, Southern blotting and hybridisation of PCR amplified products containing $C4$ isotypic nucleotide bases

Nucleotide bases encoding the $C4$ isotypic amino acid residues #1101-#1106 were amplified by primer pair number nine (see Chapter 4, Table 4.1), then visualised by electrophoresis, using the method described in section 2.9, and EtBr stained as outlined in section 2.10. Samples were transferred to nylon membrane by Southern blotting following the protocol in section 2.11.

The following $C4$ oligonucleotides were designed for hybridisation to nucleotide bases encoding the $C4$ isotypic amino acid residues #1101-#1106. The $C4A$-specific oligonucleotide for nucleotide bases encoding amino acid residues #1100 to #1107 was 5'-ACCCCTGTCGGA GTGTTAGACAGG-3'. The $C4B$-specific oligonucleotide for nucleotide bases encoding amino acid residues #1100 to #1107 was 5'-ACCTCTC TCCAGTGATACAGG-3'. End-labelling and hybridisation of oligonucleotides to PCR products were as described in sections 2.12.3 and 2.12.4.
5.3.6 C4A- and C4B-specific PCR, and PCR-RFLP analysis of C4 polymorphic regions

PCR primer pairs 10 and 11 were used for C4A- and C4B-specific amplification, respectively. PCR primers used for analysing nucleotide bases which encode amino acid residues #306, #328, #1054, #1101-#1106, #1157, #1186, #1188, #1191, and #1267 are listed in Chapter 4, Table 4.1, and their position within the C4 gene sequence is shown in Figure 4.1. PCR conditions for each set of primers can be found in Table 4.3. Polymorphic amino acid residues encoded by nucleotide bases from the C4 gene region are listed in Table 4.2. Restriction enzymes, reaction conditions, and gel electrophoresis parameters for each PCR-RFLP are given in Table 4.4.

5.3.7 Screening by PCR-RFLP analysis, for a 2 bp insertion in exon 29 which creates a C4A Q0 allele

PCR-RFLP analysis was used to screen for a 2 bp insertion into the nucleotide sequence of exon 29 in C4A Q0 allele Aboriginal donors. Sense primer 5'-GACGGCCAGTGTCGACGATAACCTGTACTGGGGCTC-3', and antisense primer 5'CTACTTGGGTACTGCGGAATC-3', were used to amplify a 254 bp C4 nucleotide sequence encoding amino acid residues #1207 to #1284, in exon 29. PCR reaction conditions were 35 cycles of 60 sec at 96 °C denaturation, 60 sec at 58 °C annealing, and 90 sec at 72 °C extension. The resulting PCR product was restricted with Fok I enzyme, using conditions recommended by the manufacturer, as outlined in section 2.8. Samples were prepared for gel electrophoresis as described in section 2.9. PCR-RFLP products were separated by gel electrophoresis in an 8 % polyacrylamide gel, at 35 mA, for 2 hours. Separated PCR-RFLP products were visualised by staining with silver as described in section 2.10.
5.3.8 Total RNA extraction from Aboriginal donors H11, H14, and H19. Preparation of cDNA, and $C4$ null allele specific PCR analysis

Total RNA was extracted from cell lines of Aboriginal donors carrying $C4$ null alleles as described in section 2.13. RNA was transferred to a nylon membrane by Northern blotting, and hybridised with a $C4$ specific probe following the protocol in section 2.14. Total cDNA was prepared from total RNA for each donor by using a Poly 'T' tail non-specific 3' primer which adheres to all 3' nucleotide coding region of genes. $C4$-specific cDNA was prepared by using a $C4$-specific 3' primer which adheres to exon 30 of the $C4$ gene region. This technique has been described in detail in section 2.15.

PCR primers for $C4$ cDNA analysis of donors H11, H14, H19 and genomic controls are listed in Table 5.1, and PCR reaction conditions for each primer pair are given in Table 5.2. Primer pairs '1' and '2', '3' and '4', '5' and '6', were used to amplify $C4A$ and $C4B$ cDNA separately, whereas primer pair '7' amplified each $C4$ allele from the $C4$ gene region. Primer pairs '1' and '2' amplified nucleotide bases encoding amino acid residues #1091 to #1097, '3' and '4' amplified nucleotide bases encoding amino acid residues #1033 to #1106, '5' and '6' amino acid residues #1101 to #1284, and '7' amplified nucleotides encoding residues #1091 to #1284.

5.3.9 Predicted Isoelectric Point (PI) of derived protein sequences for each $C4$ allele

Computer analysis of the derived protein sequence was used to determine the predicted isoelectric point (PI) of each $C4$ allele. Details concerning this computer programme are given in section 2.20 (Chapter 2).
<table>
<thead>
<tr>
<th>primer</th>
<th>sense primer</th>
<th>antisense primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5'-GACGGCCAGTGTGACACGACGAGGGCTGACGGCTCG-3'</td>
<td>5'-CTGCTTAAGGTGGGGAACACCCCA-3' C4A</td>
</tr>
<tr>
<td>2</td>
<td>5'-GACGGCCAGTGTGACACGACGAGGGCTGACGGCTCG-3'</td>
<td>5'-CTGCTTAAGGTGGGGAACACCCCA-3' C4B</td>
</tr>
<tr>
<td>3</td>
<td>5'-CTAGAATTCTCTCCAAGGCTACATGCG-3'</td>
<td>5'-CACGTCGACCACCTTGATCTCTGCTCTACAC-3' C4A</td>
</tr>
<tr>
<td>4</td>
<td>5'-CTAGAATTCTCTCCAAGGCTACATGCG-3'</td>
<td>5'-CACGTCGACCACCTTGATCTCTGCTCTATATC-3' C4B</td>
</tr>
<tr>
<td>5</td>
<td>5'-GACGGCCAGTGTGACACCCCTGACACGAGGG-3'C4A</td>
<td>5'-CTACTTGGGTACACTGCGGAATC-3'</td>
</tr>
<tr>
<td>6</td>
<td>5'-GACGGCCAGTGTGACACCCCTGACACGAGGG-3'C4B</td>
<td>5'-CTACTTGGGTACACTGCGGAATC-3'</td>
</tr>
<tr>
<td>7</td>
<td>5'-GACGGCCAGTGTGACACGAGGCTGACGGCTCG-3'</td>
<td>5'-CTACTTGGGTACACTGCGGAATC-3'</td>
</tr>
</tbody>
</table>
**Table 5.2** PCR conditions for each pair of C4-specific primers in cDNA synthesis

<table>
<thead>
<tr>
<th>primer</th>
<th>pair</th>
<th>annealing °C</th>
<th>denaturation (sec)</th>
<th>annealing (sec)</th>
<th>extension (sec)</th>
<th>PCR (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&amp;2</td>
<td></td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>120</td>
<td>342</td>
</tr>
<tr>
<td>3&amp;4</td>
<td></td>
<td>56</td>
<td>60</td>
<td>60</td>
<td>120</td>
<td>262</td>
</tr>
<tr>
<td>5&amp;6</td>
<td></td>
<td>60</td>
<td>70</td>
<td>70</td>
<td>220</td>
<td>574</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>60</td>
<td>70</td>
<td>70</td>
<td>220</td>
<td>601</td>
</tr>
</tbody>
</table>

For each PCR reaction samples were denatured at 96 °C, and extended at 72 °C. All PCR reactions were subjected to 35 cycles.
5.4 Results

Comparisons with certain results from other workers and a brief discussion is necessary here to show the relevance of these results in an applied context and the rationale behind the step-wise progression of research conducted to achieve the stated aim.

5.4.1 Detection of $C4$ null alleles by protein gel electrophoresis and immunofixation

$C4$ null allele allotypes were detected by the absence of $C4$ protein in the acidic and/or basic region of the gel for a given donor's serum sample. $C4A$ null alleles were detected in homozygous form in Aboriginal donors H19, ECY509 and Caucasian EW-Vavy. Homozygous $C4B$ null alleles were detected in Aboriginal donor's H11, H14, ECY519, ECY743 and Caucasian EW-L0081785. $C4A$ and $C4B$ null alleles in a heterozygous state with a functional $C4$ allele were also detected, but were not included in this analysis. Examples of homozygous $C4B$ null allele and heterozygous $C4A$ and $C4B$ null alleles are shown in Figure 5.1.

5.4.2 Taq I RFLP analysis of the $C4$ gene region

To establish the molecular basis of Aboriginal $C4$ null allotypes, it was first necessary to determine if the $C4$ genes of each donor were intact and of the expected $C4$ gene size (23 kb for long and 16 kb for short $C4$ genes), or whether they resulted from a large deletion, as previously detected in $C4$ null alleles from Caucasian populations.

Taq I RFLP analysis of the $C4$ gene region for $C4 Q0$ allele

Aboriginal Australian donors are shown in Figure 5.2. A summary of these results, the $C4$ allotypes, and their HLA- $DRB1$, $DQB1$, and $DQA1$ extended haplotypes are given in Table 5.3. Even though Taq I RFLP fragments are not directly associated with $C4A$ and $C4B$ genes (Tokunaga et al., 1991), it is possible to establish if donors carrying $C4$ null alleles have $C4$ genes at both $C4$ gene loci I and
Figure 5.1 C4 allotypes of C4 Q0 Aboriginal donors determined by protein gel electrophoresis and immunofixation. Schematic diagram showing the relative migration distances of C4 allotypes (VI complement genetics workshop reference typing), is reproduced with permission from WHO Bulletin, 70: 531-535, (1992). C4 allotypes for donor serum samples are as follows: Lanes #1, A 3 B 1,Q0, ECY006; #2, A 3 B 1 standard; #3, A 3,Q0 B 1, WCY1159; #4, A 3 B 1, standard; #5, A 3 B Q0, ECY743; and #6, A 3,Q0 B 1, ECY704. Samples from lane #4, 5, and 6 had been treated with Carboxypeptidase B prior to protein gel electrophoresis.
Figure 5.2 Taq I RFLP and C4-specific probe hybridisation analysis of the C4 gene region for C4 Q0 Aboriginal Australian donors. Taq I RFLP of C4 loci I and II long (23 kb) genes are represented by 7 kb and 6 kb fragments, respectively. C4 locus II short (16 kb) genes is represented by 5.4 kb fragments. Locus I gene deletion together with a short (16 kb) locus II gene are represented by 6.4 kb fragments. 21-hydroxylase A and 21-hydroxylase B genes are represented by RFLP fragments 3.7 kb and 3.2 kb, respectively. Aboriginal donors are represented in the following figures: A. lane #2 ECY509 (A Q0 B 1); #7 ECY525 (A 4 B Q0)B. lane #1 ECY743 (A 3 B Q0); #2 H19 (A Q0 B 1); #3 H14 (A3 B Q0); #4 H11 (A 3 B Q0); #5 ECY519 (A 4 B Q0); C. lane #2 ECY704 (A 3,Q0 B 1). Taq I RFLP results for each of the C4 Q0 allele donors is shown in Table 5.3. Molecular weight marker λ restricted with Hind III and Hind III/Eco RI is represented in A, lane #5 and #10, and in C, lane #1.
**Table 5.3** C4 allotypes according to protein electrophoresis, HLA DR, DQB and DQA genotypes and C4 Taq I RFLP results for C4 null allele Aboriginal Australian donors

<table>
<thead>
<tr>
<th>Donor</th>
<th>C4 allotype#</th>
<th>HLA-DRB1</th>
<th>HLA-DQB1</th>
<th>HLA-DQA1</th>
<th>Taq 1 RFLP (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Locus 1</td>
</tr>
<tr>
<td>H11</td>
<td>A 3 B Q0</td>
<td>*04, *04</td>
<td>*0402, *0402</td>
<td>*0301, *0301</td>
<td>7</td>
</tr>
<tr>
<td>H14</td>
<td>A 3 B Q0</td>
<td>*0803, *04</td>
<td>*0503, *0402</td>
<td>*0103, *0301</td>
<td>6</td>
</tr>
<tr>
<td>ECY743</td>
<td>A 3 B Q0</td>
<td>*0405, *1301</td>
<td>*0401, *0603</td>
<td>*0301, *0103</td>
<td>6</td>
</tr>
<tr>
<td>ECY519</td>
<td>A 4 B Q0</td>
<td>*0803, *0803</td>
<td>*0503, *0503</td>
<td>*0103, *0103</td>
<td>5.4</td>
</tr>
<tr>
<td>ECY525</td>
<td>A 4 B Q0</td>
<td>*0301, *0803</td>
<td>*0201, *0503</td>
<td>*0501, *0103</td>
<td>5.4</td>
</tr>
<tr>
<td>H19</td>
<td>A Q0 B1</td>
<td></td>
<td></td>
<td></td>
<td>3.2</td>
</tr>
<tr>
<td>ECY509</td>
<td>A Q0 B1</td>
<td>*0404, *1402</td>
<td>*0302, *0301</td>
<td>*0301, *0501</td>
<td></td>
</tr>
<tr>
<td>ECY704</td>
<td>A 3 Q0 B1</td>
<td>*0301, *1405</td>
<td>*0201, *0503</td>
<td>*0501, *0101</td>
<td></td>
</tr>
</tbody>
</table>

# allotype determined by protein gel electrophoresis. Taq 1 RFLP of C4 locus I and II long (23 kb) genes are represented by 7 kb and 6 kb fragments, respectively. C4 locus II short (16 kb) genes are represented by 5.4 kb fragments. Locus I gene deletion together with a short (16 kb) locus II gene is represented by 6.4 kb fragments. RFLP fragments represent 21-hydroxylase A gene by 3.7 kb and 21-hydroxylase B gene by 3.2 kb fragments. The HLA DRB, DQB and DQA genotypes were not determined. HLA Class II typing was performed by the Human Genetics Group, JCSMR.
II. \textit{C4A} genes are usually present at locus I and \textit{C4B} genes at locus II (Schneider, 1990).

Each donor with a \textit{C4B} \textit{Q0} allotype had long \textit{C4} genes at locus I. Aboriginal donor H11 had short \textit{C4} genes at locus II, whereas Aboriginal donors ECY743 and ECY519 had long \textit{C4} genes at locus II. Aboriginal donor H14 was a heterozygote at locus II, with both long and short \textit{C4} genes. Each donor had a \textit{CYP21A} and \textit{CYP21B} gene which were represented by 3.7 kb and 3.2 kb fragments, respectively.

It was hypothesised that two types of \textit{C4B} *\textit{Q0} alleles were present in this population of Aboriginals as both long and short \textit{C4} genes were found in association with the \textit{C4B} null allotype by \textit{Taq I} RFLP analysis. This hypothesis was supported by analysis of each donor's HLA-extended haplotype, results of which are given in Table 5.3. Donor H11 was thought to be a homozygote for the short type of \textit{C4B} *\textit{Q0} allele which had the HLA-extended haplotype \textit{DRB} *04, \textit{DQB} *0402, \textit{DQA} *0301, whereas donor ECY519 appeared to be a homozygote for the long type of \textit{C4B} *\textit{Q0} allele having the HLA-extended haplotype \textit{DRB} *0803, \textit{DQB} *0503, \textit{DQA} *0103. Donor H14 appeared to be heterozygous for each type of \textit{C4B} *\textit{Q0} allele, having one of each of the HLA-extended haplotypes found in H11 and ECY519. Donor ECY743 had a different HLA-extended haplotype more typical of that seen in Caucasian populations (Dawkins et al., 1983), and therefore was expected to have different \textit{C4B} \textit{Q0} alleles.

\textit{Taq I} RFLP analysis of Aboriginal donors H19 and ECY509 having a \textit{C4A} \textit{Q0} allotype, found long (7 kb) \textit{C4} genes at locus I, and both long (6 kb) and short (5.4 kb) \textit{C4} genes at locus II, (Figure 5.2 and Table 5.3). The HLA-extended haplotype for donor H19 was not resolved, and therefore requires further investigation. The HLA-extended haplotype for donor ECY509 indicated heterozygosity of \textit{C4A} null alleles.

Only one Aboriginal donor in this study, namely, 'ECY704' with the allotype \textit{A 3,Q0 B 1}, was found to have \textit{C4A} deficiency resulting from a large deletion of the \textit{C4} gene at locus I, which was
represented by a 6.4 kb Taq I RFLP fragment (Figure 5.2 and Table 5.3). This donor had foreign ethnic admixture of a DRB1 *0301 allele (Table 5.3), which has been found in linkage disequilibrium with C4A null alleles resulting from a large deletion in Caucasian populations (Schur et al., 1990; Kumar et al., 1991).

Aboriginal donor ECY525, which has the C4 allotype A 4 B Q0, showed unusual results of Taq I RFLP and HLA-extended haplotype analysis. Only a long (7 kb) C4 fragment was detected at locus I by Taq I RFLP analysis. The HLA-extended haplotype for this donor showed the same alleles on one haplotype as that of donor ECY704, (having partial C4A deficiency resulting from a large deletion of the C4A gene at locus I). The second haplotype of donor ECY525 had the same alleles as the HLA-extended haplotype found in donor ECY519, (having C4B deficiency of the long (6 kb) C4 gene type). Donor ECY525 may have gene duplication at the C4 gene region, as Taq I RFLP analysis resulted in a strong 7 kb DNA fragment and weaker 6.4 kb, 6 kb and 5.4 kb DNA fragments. The 6.4 kb fragment corresponds to a deletion of a long (7 kb) gene at locus I and the presence of a short C4 gene at locus II on one chromosome, however the strong 7 kb fragment suggests two long loci I C4 genes. C4 gene copy number of this donor could be clarified by PFGE analysis with Bss HII or Sac II restriction enzymes and a C4 gene specific probe.

These results show that the molecular-basis of C4 null allotypes in Aboriginal Australian donors do not result from large C4 gene deletions. The next important question is whether these non-deleted C4 null alleles are functional and can produce messenger RNA (mRNA), which will result in C4 protein. The C4 gene locus is a duplicated gene region having a minimum of two highly homologous alleles on each haplotype. Therefore, prior to RNA analysis it was first necessary to find a means of identifying the C4 null alleles as distinct from other C4 alleles on each haplotype. Donors ECY525 and ECY704 were not included in any further molecular analysis as donor ECY525 was suspected of having a duplicated C4 gene loci and ECY704 was heterozygous for the C4A alleles. Both of these features may interfere with the interpretation of molecular results for the C4 null alleles in these donors.
5.4.3 C4A- and C4B- specific oligonucleotide hybridisation analysis of Aboriginal donors' having C4 null alleles

C4-specific PCR primers were used to amplify the C4 gene region encoding nucleotide bases for the C4A and C4B isotypic amino acid residues, and was followed by C4A- and C4B-specific oligonucleotide hybridisation analysis (see Figure 5.3). The C4A-specific oligonucleotide hybridised to each donor including those with C4A null allele allotypes, namely H19 and ECY509, whereas the C4B-specific oligonucleotide hybridised to each donor, except those with C4B null allele allotypes, namely H11, H14, ECY743 and ECY519. This result indicated that nucleotide bases encoding the C4A isotypic amino acid residues were present in all C4 null allele donors including those with a C4A null allele allotype. On the other hand, nucleotide bases encoding the C4B isotypic amino acid residues were absent and/or different in donors with C4B null allele allotypes.

5.4.4 C4A and C4B isotypic PCR amplification of C4 *Q0 alleles

The absence and/or alteration of nucleotide bases encoding C4B isotypic amino acid residues for C4B null alleles was confirmed by lack of PCR amplification using the C4B-specific oligonucleotide (Figure 5.4). Donors H11 and ECY743 produced the same results but are not shown in this figure. The presence of nucleotide bases encoding C4A isotypic amino acid residues for C4A null alleles was confirmed by PCR-RFLP analysis of C4A-specifically amplified PCR, DNA fragments (Figure 5.5, 5.6 and Table 5.4).

5.4.5 PCR-RFLP analysis of nucleotide bases encoding C4A and C4B isotypic amino acid residues in C4B null allele donors

Restriction analysis of the C4 isotypic region in Aboriginal donors having either C4A or C4B null alleles was performed by PCR-RFLP with enzyme Mnl I (Figure 5.5 and Table 5.4). This analysis showed
Figure 5.3 Hybridisation of $C4A$- and $C4B$-specific oligonucleotides to Aboriginal donors with $C4Q0$ alleles. PCR products were obtained using primers which would amplify both $C4A$ and $C4B$ alleles. Hybridisation in A was with a $C4A$-specific oligonucleotide probe, whereas in B the same PCR products were probed with a $C4B$-specific oligonucleotide. Lane numbers in A and B represent the following Aboriginal donors; #1 ECY519 ($A 4 B Q0$); #2 ECY743 ($A 3 B Q0$); #3 H11 ($A 3 B Q0$); #4 H14 ($A 3 B Q0$); #5 H19 ($A Q0 B 1$); #6 WCY1164 ($A 4 B 2$); #7 ECY504 ($A 2,3 B 1$).
Figure 5.4 C4B isotypic PCR amplification of Aboriginal donors with C4B Q0 alleles. Samples include; lane #1 Molecular weight marker pBR322 restricted with *Hae III*; #2 H14 (A 3 B Q0); #3 ECY519 (A 4 B Q0); #4 ECY873 (A 3 B 1,12); #5 WCY1204 (A 3 B 1,12); #6 PCR negative control (PCR reaction with no DNA template added).
Figure 5.5 PCR-RFLP banding patterns of nucleotide bases encoding polymorphic amino acid residues #1054 and #1101-#1106 in C4 Q0 alleles. Lanes for nucleotide bases encoding amino acid residue #1054 are; #3. 'GLY'; #9. 'ASP'; #7. heterozygote 'GLY' and 'ASP'. PCR-RFLP banding patterns for nucleotide bases encoding amino acid residues #1101-#1106 are; lane #19. PCR-RFLP patterns for C4B Q0 donors which have C4A isotypic nucleotide bases, #18. PCR-RFLP pattern for C4A Q0 donors with both C4A and C4B isotypic nucleotides. Restricted fragments less than 19 bp for residue #1054, and fragments less than 53 bp for residues #1101-#1106, were lost during electrophoresis. Unrestricted PCR fragments are shown in lane #1 and #13 for residues #1054 and #1101-#1106, respectively. Molecular weight marker pBR322 restricted with Hae III resulting in 18, 21, 51, 57, 64, 80, 89, 104, 123, 124, 184, 192, 213, 234, 267, 434, 458, 504, 540, and 587 base pair fragments are shown in lane #2 for residue #1054. pBR322 restricted with Msp I represented by 67, 76, 90, 110, 123, 147, 160, 180, 190, 201, 217, 238, 242, 309, 404, 527, and 622 base pair fragments are shown in lane #14 for residues #1101-#1106.
Figure 5.6 PCR-RFLP banding patterns of nucleotide bases encoding polymorphic amino acid residues #1157, #1186, #1188, #1191 and #1267, in C4 Q0 alleles. For gels corresponding to residue #1157, lanes of each PCR-RFLP banding pattern are; lane #2 'ASN'; #11 'SER'; #5 heterozygote for 'ASN' and 'SER', unrestricted fragments are shown in lanes #1 and #7. For gels corresponding to residue #1186, example PCR-RFLP banding patterns for this residue are; lane #9 nucleotide bases 'gCG' for residue 'ALA'; #10 heterozygous for bases 'gCG' and 'gCC', both encoding residue 'ALA', unrestricted fragments are shown in lane #2. PCR-RFLP analysis of nucleotides encoding residues #1188 and #1191 are from digestion of the same PCR product and with the same restriction enzyme. Lanes of PCR-RFLP analysis of residues #1188 and #1191 are; lane #3 'ALA' for residue #1188; #10 heterozygous for 'ALA' and 'VAL' for residue #1188; #18 'ARG' for residue #1191; #23 'LEU' for residue #1191, unrestricted fragments are shown in lane #2, #16 and #22. PCR-RFLP banding patterns for nucleotides encoding residue #1267 are; lane #3 'ALA'; #9 heterozygote for 'ALA' and 'SER', unrestricted fragments are shown in lane #1. Each gel shows molecular weight marker pBR322 restricted with Msp I represented by 67, 76, 90, 110, 123, 147, 160, 180, 190, 201, 217, 238, 242, 309, 404, 527, and 622 base pair fragments in lanes #5 and #8 for residue 1157, lanes #17 and #21 for residues 1188 and 1191, and lane #2 for residue 1267. Only molecular weight fragments of 622, 527, 404, and 309 bp are shown in lane #1 for residue 1186, and 622, 527, and 404 bp are shown in lane #1 for residues 1188 and 1191.
Table 5.4 C4 polymorphic amino acids of Aboriginal Australian donors as derived from analysis of nucleotide sequence by PCR-RFLP analysis

<table>
<thead>
<tr>
<th>Donor</th>
<th>C4 allotype#</th>
<th>C4 allotype*</th>
<th>#1054</th>
<th>#1101-#1106</th>
<th>#1157</th>
<th>#1186</th>
<th>#1188</th>
<th>#1191</th>
<th>#1267</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>C4A</td>
<td>C4B</td>
<td>C4A</td>
<td>C4B</td>
<td>C4A</td>
<td>C4B</td>
<td>C4A</td>
</tr>
<tr>
<td>H11</td>
<td>A 3 B Q0</td>
<td>A *3,*CAN1</td>
<td>D</td>
<td>PCPVLD</td>
<td>N&amp;S</td>
<td>a&amp;b</td>
<td>V&amp;A</td>
<td>R&amp;L</td>
<td>A</td>
</tr>
<tr>
<td>H14</td>
<td>A 3 B Q0</td>
<td>A *3,*CAN1</td>
<td>D</td>
<td>PCPVLD</td>
<td>N&amp;S</td>
<td>a&amp;b</td>
<td>V&amp;A</td>
<td>R&amp;L</td>
<td>A</td>
</tr>
<tr>
<td>ECY519</td>
<td>A 4 B Q0</td>
<td>A *4,*4</td>
<td>D</td>
<td>PCPVLD</td>
<td>N</td>
<td>a</td>
<td>V</td>
<td>L</td>
<td>A</td>
</tr>
<tr>
<td>ECY743</td>
<td>A 3 B Q0</td>
<td>A *3,*3</td>
<td>D</td>
<td>PCPVLD</td>
<td>N</td>
<td>a</td>
<td>V</td>
<td>L</td>
<td>A</td>
</tr>
<tr>
<td>ECY509</td>
<td>A Q0 B 1</td>
<td>A *Q0a,*Q0b B</td>
<td>G&amp;D</td>
<td>G</td>
<td>PCPVLD LSPVIH</td>
<td>N</td>
<td>S</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>H19</td>
<td>A Q0 B 1</td>
<td>A *Q0a,*Q0b B</td>
<td>D</td>
<td>G</td>
<td>PCPVLD LSPVIH</td>
<td>N</td>
<td>S</td>
<td>a</td>
<td>b</td>
</tr>
</tbody>
</table>

# allotype determined by protein gel electrophoresis. * allotype determined by DNA analysis. Amino acid #1186 is alanine 'A' for both C4A and C4B alleles, however the nucleotide bases encoding this amino acid are different, 'gcg' for C4A alleles, and 'gcc' for C4B alleles. 'gcg' is represented by 'a' and 'gcc' by 'b'. Only C4A-specifically amplified PCR product was used for analysis of C4B Q0 allele donors for nucleotides encoding amino acid residues #1054, #1157, #1186, #1188, #1191, and #1267, as analysis of nucleotides encoding isotypic residues #1101-#1106 in these donors showed only C4A bases were present.
null allele donors had both \textit{C4A} and \textit{C4B} isotypic nucleotides, whereas \textit{C4B} null allele donors possessed only \textit{C4A} nucleotide bases in this region. These results confirmed those detected by \textit{C4} isotypic hybridisation and by \textit{C4A} and \textit{C4B} isotypic PCR amplification studies described above.

5.4.6 PCR-RFLP analysis of nucleotide bases encoding \textit{C4} allelic polymorphic amino acid residues in \textit{C4B *Q0} alleles of Aboriginal donors

The studies described above suggest that \textit{C4B} null alleles of Aboriginal donors H11, H14, ECY519 and ECY743, may have two \textit{C4A} alleles on the one haplotype but no \textit{C4B} allele. To test this hypothesis, and identify nucleotides which are unique to the \textit{C4B} null allele, PCR-RFLP analysis of nucleotide bases encoding known polymorphic amino acid residues (Yu, 1991), was performed. Only \textit{C4A} specifically amplified PCR fragments from \textit{C4B *Q0} allele donors were used for PCR-RFLP analysis, as these donors did not have \textit{C4B} isotypic nucleotide bases. PCR-RFLP banding patterns are shown in Figures 5.5, 5.6, and 5.7, and a summary for each donor is presented in Tables 5.4 and 5.5.

PCR-RFLP analysis of nucleotide bases encoding residues #306 and #328 for short (16 kb) \textit{C4} genes of donors H11 and H14, detected nucleotides which encoded amino acid residues 'Met' and 'Tyr', respectively. Analysis of PCR fragments, containing both long (23 kb) and short (16 kb) \textit{C4} genes, in donors H11 and H14, found both residues 'Met' and 'Lys' were encoded by nucleotides for residue #306 in each donor. Amino acid residue 'Tyr' was encoded by nucleotides for residue #328 in donor H11, whereas residues 'Tyr' and 'Ser' were encoded by nucleotides for residue #328 in donor H14. These results indicate that nucleotides encoding 'Tyr' are present in both long and short \textit{C4} genes of donor H11. However, nucleotides of donor H14 encode 'Tyr' in the short and 'Ser' in one or both of the long \textit{C4} genes for residue #328.
Figure 5.7 PCR-RFLP banding patterns from the analysis of nucleotide bases encoding polymorphic amino acid residues #306 and #328 in C4 Q0 allele donors. Lanes representing the analysis of residue #306 are; #1. 'MET' in short C4 genes; #7. 'LYS' in short C4 genes, #3. 'LYS' in long and short C4 genes, #15. heterozygous for 'LYS' and 'MET' in long and short C4 genes. Lanes representing the analysis of nucleotides encoding residue #328 are; lane #12. 'TYR' in short C4 genes; #19. 'SER' in short C4 genes; #14. 'TYR' in long and short C4 genes; #28. 'SER' in long and short C4 genes; #15. heterozygous for 'TYR' and 'SER' in long and short C4 genes. Unrestricted PCR fragment controls for the analysis of residue #306 are shown in lane #11 and #12, and lane #11 and #25 for residue #328. PCR-RFLP fragment sizes less than 57 bp were lost from each gel during gel electrophoresis. Molecular weight marker pBR322 restricted with Msp I represented by 67, 76, 90, 110, 123, 147, 160, 180, 190, 201, 217, 238, 242, 309, 404, 527, and 622 base pair fragments are shown in lane #10 and #13 for residue #306, and lane #10 and #26 for residue #328.
<table>
<thead>
<tr>
<th>Amino Acid Residue #</th>
<th>Short C4 Gene (16kB)</th>
<th>Long and Short C4 Gene (16kB and 23kB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>57+102+135+271</td>
<td>10+61+135</td>
</tr>
<tr>
<td>Y</td>
<td>328</td>
<td>3+63+118+122+229</td>
</tr>
<tr>
<td>K</td>
<td>306</td>
<td>3+16+63+122</td>
</tr>
<tr>
<td>M</td>
<td>PCR-RFLP (bp)</td>
<td>PCR-RFLP (bp)</td>
</tr>
</tbody>
</table>
Table 5.5  β-chain, exon 9 polymorphic amino acid residues in Aboriginal Australians

<table>
<thead>
<tr>
<th>Donor</th>
<th>C4 allotype#</th>
<th>C4 allotype*</th>
<th>amino acid # 306</th>
<th>amino acid # 328</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>short C4</td>
<td>long and short C4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>short C4</td>
<td>long and short C4</td>
</tr>
<tr>
<td>H11</td>
<td>A 3 B Q0</td>
<td>A *3, *CAN1</td>
<td>M</td>
<td>M &amp; K</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>H14</td>
<td>A 3 B Q0</td>
<td>A *3, *CAN1</td>
<td>M</td>
<td>M &amp; K</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>S &amp; Y</td>
</tr>
<tr>
<td>ECY519</td>
<td>A 4 B Q0</td>
<td>A *4, *4</td>
<td>no product</td>
<td>K</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>no product</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>ECY743</td>
<td>A3 B Q0</td>
<td>A *3, *3</td>
<td>no product</td>
<td>K</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>no product</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>S &amp; Y</td>
</tr>
<tr>
<td>ECY509</td>
<td>A Q0 B 1</td>
<td>A *Q0a, *Q0b B *1</td>
<td>K</td>
<td>K</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>S &amp; Y</td>
</tr>
<tr>
<td>H19</td>
<td>A Q0 B 1</td>
<td>A *Q0a, *Q0b B *1</td>
<td>K</td>
<td>K</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>S &amp; Y</td>
</tr>
</tbody>
</table>

# allotype determined by protein gel electrophoresis. * allotype determined by DNA analysis. 'no product' refers to lack of PCR amplification of short C4 alleles, indicating only long C4 genes were encoded by these donors.
These results indicate that the short $C4B\ *Q0$ alleles of donors H11 and H14 are unique by comparison to the majority of other donors as 'Tyr' instead of 'Ser' is encoded by nucleotides for residue #328 of the short $C4$ genes (Chapter 4, Tables 4.5 and 4.6). Also, similar to the amino acid composition of rare $C4$ alleles described in Chapter 4 (such as $A\ *58\sim$, $B\ *12\sim$, $B\ *22\sim$, $B\ *5\sim$, and $B\ *6\sim$), amino acid residue 'Met' is encoded by nucleotides for residue #306 in the $C4B\ *Q0$ alleles of donors H11 and H14. $C4B\ *Q0$ alleles of donors ECY519 and ECY743 were different to those of donors H11 and H14, and similar to those of other long $C4$ genes described in Chapter 4, (Tables 4.5 and 4.6). Donor ECY519 had amino acid residues 'Tyr' and 'Lys' encoded by nucleotides for residues #328 and #306, respectively, whilst donor ECY743 had amino acid residues 'Tyr' and 'Ser' encoded by nucleotides for residue #328, and 'Lys' for residue #306. Heterozygosity of nucleotide bases encoding amino acid residues #1157, #1186, #1188, and #1191, and homozygosity of nucleotides encoding amino acid residues #1054 and #1267 were found in donors H11 and H14 (Table 5.4). These results indicate that more than one type of $C4A$ allele is present in Aboriginal donors H11 and H14. Also, heterozygosity of nucleotides encoding amino acid residue #328 of donor ECY743 suggests more than one type of $C4A$ allele exists for this donor, whereas donor ECY519 had only one type of $C4A$ allele according to PCR-RFLP analysis.

5.4.7 PCR-RFLP analysis of nucleotide bases encoding polymorphic amino acid residues of $C4A$ null alleles of Aboriginal donors

Results of PCR-RFLP analysis of nucleotide bases encoding amino acid residues #306 and #328 of $C4A\ *Q0$ alleles of donors H19 and ECY509, are given in Table 5.5. PCR-RFLP banding patterns for each of these donors are shown in Figure 5.7. Both long and short $C4$ genes of donors H19 and ECY509 had 'Lys' for residue #306. Short $C4$ genes of these donors had 'Ser' for residue #328, whereas both 'Ser' and 'Tyr' amino acid residues were detected in PCR fragments amplified from the long and short $C4$ genes. These results indicate that 'Ser' is encoded by nucleotides of the short $C4$ genes of these donors, whereas 'Tyr' and/or 'Ser' are encoded by the long $C4$ genes.
Results of PCR-RFLP analysis of nucleotide bases encoding amino acid residues #1054, #1157, #1186, #1188, #1191, and #1267 for C4A *Q0 alleles of donors H19 and ECY509 are given in Table 5.4, and the resulting PCR-RFLP banding patterns are shown in Figures 5.5 and 5.6. Each type of C4A allele of these donors had the same nucleotide bases, encoding amino acid residues #1157, #1186, #1188, and #1191. However, heterozygosity was detected for residue #1267, where both donors had amino acid residues 'Ala' and 'Ser'. Donor H19's nucleotide bases encoded 'Asp' for residue #1054, whereas donor ECY509's nucleotide bases showed heterozygosity for this residue encoding 'Asp' and 'Gly'.

5.4.8 Aboriginal C4B *Q0 alleles identified by PCR-RFLP analysis and their predicted isoelectric points (PI)

The amino acid residues derived from PCR-RFLP analysis of nucleotide bases from Aboriginal C4 null alleles were compared with those residues derived for other C4 alleles nucleotide bases described in Chapter 4 (Tables 4.10 and 4.11). PCR-RFLP analysis of C4 null alleles and comparison with other defined C4 alleles enabled nucleotides encoding amino acid residues specific for C4 *Q0 alleles to be identified. A summary of these results and the predicted isoelectric point (PI) of each C4 *Q0 allele is given in Table 5.6.

From these results it can be seen that a new C4B *Q0 allele has been identified in donors H11 and H14. This allele is a short C4 gene which consists of a unique hybrid combination of both C4A- and C4B-specific amino acid residues. This C4B *Q0 allele has not previously been detected, and as it has C4A isotypic residues, was named C4A *CAN1. Nucleotide bases of the C4A *CAN1 allele encoded 'Met' for residue #306, whereas nucleotide bases of the A *3 allele, of donors H11 and H14, encoded 'Lys'. Sequencing of the C4A *CAN1 allele revealed nucleotide bases which encode 'Met' for residue #309, whereas other C4 alleles nucleotide bases encoded 'Ile' (Chapter 4, Figure 4.3). However, this polymorphism could not be detected by PCR-RFLP analysis, and was therefore excluded from this analysis. As with the C4A *3 allele, nucleotide bases of the
### Table 5.6 C4 *Q0* allele amino acid residue polymorphism, derived from nucleotide base PCR-RFLP analysis

<table>
<thead>
<tr>
<th>C4 allotype</th>
<th>#306</th>
<th>#309</th>
<th>#328</th>
<th>#1054</th>
<th>#1101-#1106</th>
<th>#1157</th>
<th>#1186</th>
<th>#1188, #1191, #1267</th>
<th>PI</th>
<th>example</th>
<th>donors</th>
</tr>
</thead>
<tbody>
<tr>
<td>A *3</td>
<td>K</td>
<td>M</td>
<td>Y</td>
<td>D</td>
<td>PCPVLD</td>
<td>N</td>
<td>a</td>
<td>VL</td>
<td>A</td>
<td>7.05</td>
<td>published</td>
</tr>
<tr>
<td>A *CAN1</td>
<td>M</td>
<td>I</td>
<td>Y</td>
<td>D</td>
<td>PCPVLD</td>
<td>S</td>
<td>b</td>
<td>AR</td>
<td>A</td>
<td>7.05</td>
<td>H11, H14</td>
</tr>
<tr>
<td>B *1</td>
<td>K</td>
<td>M</td>
<td>S</td>
<td>G</td>
<td>LSPVIH</td>
<td>S</td>
<td>b</td>
<td>AR</td>
<td>A</td>
<td>7.23</td>
<td>ECY858</td>
</tr>
<tr>
<td>B *2</td>
<td>K</td>
<td>M</td>
<td>S</td>
<td>D</td>
<td>LSPVIH</td>
<td>S</td>
<td>b</td>
<td>AR</td>
<td>A</td>
<td>7.18</td>
<td>published</td>
</tr>
<tr>
<td>A *3</td>
<td>K</td>
<td>M</td>
<td>Y</td>
<td>D</td>
<td>PCPVLD</td>
<td>N</td>
<td>a</td>
<td>VL</td>
<td>A</td>
<td>7.05</td>
<td>published</td>
</tr>
<tr>
<td>A *3 or A *4</td>
<td>K</td>
<td>M</td>
<td>Y</td>
<td>D</td>
<td>PCPVLD</td>
<td>N</td>
<td>a</td>
<td>VL</td>
<td>A</td>
<td>7.05</td>
<td>ECY743, ECY519</td>
</tr>
<tr>
<td>A *Q0Aba</td>
<td>K</td>
<td>-</td>
<td>Y</td>
<td>G</td>
<td>PCPVLD</td>
<td>N</td>
<td>a</td>
<td>VL</td>
<td>A</td>
<td>7.11</td>
<td>ECY509</td>
</tr>
<tr>
<td>A *Q0Abb</td>
<td>K</td>
<td>-</td>
<td>Y</td>
<td>D</td>
<td>PCPVLD</td>
<td>N</td>
<td>a</td>
<td>VL</td>
<td>A</td>
<td>7.05</td>
<td>H19</td>
</tr>
<tr>
<td>A *Q0Abc</td>
<td>K</td>
<td>-</td>
<td>S</td>
<td>D</td>
<td>PCPVLD</td>
<td>N</td>
<td>a</td>
<td>VL</td>
<td>S</td>
<td>7.05</td>
<td>H19, ECY509</td>
</tr>
</tbody>
</table>

Amino acid residue #1186 is alanine 'A' for both C4A and C4B alleles, however the nucleotide bases encoding this amino acid are different, 'gcg' for C4A alleles, and 'gcc' for C4B alleles. 'gcg' is represented by 'a' and 'gcc' by 'b'. C4A *3* is from this study and published by Bell et al. (1984), and Yu. (1991); C4B *2* is from Belt et al. (1984 and 1985). C4B *1* was from Chapter 4 (Table 4.11). Other examples were from donors in this study. 'PI' indicates the predicted isoelectric point by computer protein analysis. Residue #309 polymorphism was identified from Chapter 4 sequencing analysis (Figure 4.3).
C4A *CAN1 allele encoded 'Tyr' for residue #328, whereas nucleotides for residues #1157, #1186, #1188, and #1191, from the C4A *CAN1 allele encoded the same residues as detected in other C4B alleles (Table 5.6). Similar to the C4A *3, B *1, and B *2 alleles, nucleotide bases of the C4A *CAN1 allele encoded 'Ala' for residue #1267. This combination of amino acid residues together with the known conserved amino acid residues of the C4 gene, produced a PI value of 7.05, which is the same as that predicted for the C4A *3 allele. Therefore, C4A *3 and C4A *CAN1 alleles would be indistinguishable by protein gel electrophoresis.

PCR-RFLP analysis (Table 5.6), and Taq I RFLP analysis (Table 5.3), of the second type of C4B *Q0 allele found in Aborigines ECY743, ECY519, and the long C4B *Q0 allele of donor H14, found nucleotides which were identical to those of the C4A *3 and A *4 alleles at locus I of these donors. Therefore, C4 alleles of these donors would be indistinguishable by protein gel electrophoresis.

5.4.9 Aboriginal C4A *Q0 alleles identified by PCR-RFLP analysis, and their predicted isoelectric points (PI)

There appeared to be three types of C4A *Q0 non-deleted alleles in the Aboriginal population (Table 5.6), as heterozygosity of residues #1267 and #1054 in donor ECY509 was found (Table 5.4). Nucleotide bases of one type of C4A *Q0 allele, which was denoted 'C4A *Q0Abi', encoded amino acid residue 'Gly' for residue #1054 which is the same as the A *2~ allele (Table 4.10). Nucleotide bases of another type of C4A *Q0 allele, which was denoted 'C4A *Q0Abb', were similar to those of the A *3a~ allele (Table 4.10), encoding 'Ala' for residue #1267. Nucleotide bases of the the third type of C4A *Q0 allele, which was denoted 'C4A *Q0Abc', encoded 'Ser' for residue #1267 (Table 5.6), as was found for alleles C4A *3b~ and C4A *4, (Table 4.10).

In summary, more than one type of non-deleted C4A *Q0 allele is present in the Aboriginal population. Donor ECY509 is heterozygous for C4A *Q0Abi, and C4A *Q0Abc alleles, and donor H19 is
heterozygous for the C4A *Q0Abb, and C4A *Q0Abc alleles (Table 5.6).

5.4.10 PCR-RFLP analysis of Aboriginal donors having C4A Q0 alleles to screen for a 2 bp insertion into exon 29

It was necessary to determine if non-deleted C4A *Q0 alleles of donors H19 and ECY509 resulted from the same mutation as a Caucasian non-deleted C4A *Q0 allele described by Barba et al. (1993). Nucleotide bases encoding amino acid residues #1207 to #1284 were amplified from each C4 allele of Caucasian donors RCH555 (A Q0 B 1,2), ASD22 (A Q0 B 1) and ASD37 (A Q0 B 1), and Aboriginal donors H19 (A Q0 B 1), ECY509 (A Q0 B 1) and WCY1204 (A 3 B 1,12). PCR-RFLP analysis of this PCR product (254 bp), with the restriction enzyme Fok I, was undertaken to identify the presence of a 2 bp insertion into exon 29. Restriction analysis of C4 alleles with Fok I produces two fragments of 76 bp and 178 bp in length. If the 2 bp insertion is present, it will be found in the 76 bp fragment. During electrophoresis, this results in two additional bands migrating slower than the 76 bp band. One of these bands contains the 2 bp insertion (mutant allele), whilst the second, slower migrating band results from a combination of the wild and mutant type alleles. This type of pattern was detected for Caucasian donor RCH555 (Figure 5.8). Other C4A Q0 allele donors did not result in this pattern, and therefore did not have the 2 bp insertion in this gene region, nor did donor WCY1204, used as a control, who had functional C4A and C4B alleles. Donor ECY509 appears to have an additional PCR-RFLP banding pattern apart from the expected pattern. The origin of this additional pattern is unknown and requires further investigation by sequence analysis.

Barba et al. (1993), found that the C4A *Q0 alleles which resulted from a 2 bp insertion were in linkage disequilibrium with the HLA-DR 6 allele. Donor RCH555 was chosen to investigate the presence of the 2 bp insertion, as this donor had a C4A *Q0 allele and a DR 6 allele. Aboriginal donor ECY509 was heterozygous at the DR locus, having DR *0404, and DR *1402 alleles, which are DR 4 and DR 6
Figure 5.8 PCR-RFLP analysis of C4A Q0 allele donors, in search of a 2 bp insertion in exon 29. PCR fragments include nucleotide bases encoding residues #1207 to #1284 for each C4 allele. Restriction analysis resulted in PCR fragment sizes of 178 bp and 76 bp for each sample. Lane #3 represented donor WCY1204 (A 3 B 1,12), #4 donor ASD37 (A Q0 B I), #5 donor ECY509 (A Q0 B I), #6 donor H19 (A Q0 B I), #7 donor ASD22 (A Q0 B I), and #8 donor RCH555 (A Q0 B 1,2). Lane #1 is unrestricted PCR fragments, and lane #2 represents molecular weight marker pBR322 restricted with Msp I.
allotypes, respectively (Table 5.3). However, this donor did not have a 2 bp insertion in exon 29 of its C4 alleles. Donor ASD22 was heterozygous at the DR locus having DR 3,4 alleles, and ASD37 was homozygous for DR 3 alleles. The DR haplotype of donor H19 is unknown.

5.4.11 Transcription analysis of C4 *Q0 alleles in Aboriginal donors H11, H14, and H19

Now that it was possible to identify the short, non-deleted C4B *Q0 allele of donors H11 and H14, and the C4A *Q0 alleles of donor H19, RNA studies were performed to determine if these alleles were transcribed. Extraction of total RNA from cell lines of donors H11, H14 and H19, was used for both C4-specific cDNA and total cDNA analysis. This revealed that these donors could produce full length mRNA for the C4 Q0 allele which they possessed. PCR and PCR-RFLP analysis of cDNA from donors H11, H14 and H19, are shown in Figures 5.9, 5.10 and 5.11. A C4B-specific primer which detected nucleotide bases encoding 'Arg' for residue #1191, was used to amplify the C4A *CAA7 allele from cDNA of donors H11 and H14, (Figure 5.9). The same primer amplified the C4B alleles from cDNA of donor H19, (Figure 5.9). There was some genomic DNA contamination in each cDNA sample, but was easily distinguished from the cDNA-PCR product due to the different size of each product (542 bp for genomic DNA, and 342 bp for cDNA). These same results were obtained from both total, and C4-specific, cDNA preparations.

PCR-RFLP analysis, of cDNA containing all C4 alleles of donors H11, H14 and H19, using restriction enzyme Aci I, was carried out to identify nucleotides encoding residues #1188 and #1191. This analysis detected both C4A and C4B types of nucleotides thereby indicating that both C4A *3 and C4A *CAN1 alleles were fully transcribed (Figure 5.10A and B). Restriction analysis produced 544 bp, 41 bp, and 16 bp fragment sizes for the A *3 allele, indicating that amino acid residues 'Val' and 'Leu' were encoded by nucleotides for residues #1188 and #1191, respectively. Fragment sizes 306 bp, 230 bp, 41 bp, 16 bp, and 8 bp were detected for the C4A *CAN1 allele, indicating that amino acid residues 'Ala' and 'Arg' were
Figure 5.9 C4B-specific PCR amplification of nucleotides encoding amino acid residues #1091 to #1191 from cDNA samples of Aboriginal donors H11, H14 and H19. The cDNA PCR fragment size was 342 bp, whereas the genomic PCR fragment size was 542 bp. Lane #1 H11, #2 H14, #3 H19, #4 WCY1204 genomic positive control (primer pair 2, Table 5.1), #5 ECY743 (A 3 B Q0) genomic negative control (primer pair 2, Table 5.1), #6 ECY743 genomic positive control using different primers which are not C4B-specific, and amplified bases encoding residue #1091 to #1206 (primer pair 9, Table 4.1, Chapter 4), #7 molecular weight marker pBR322 restricted with Msp I, #8 PCR negative control, without a DNA template. Each cDNA PCR result showed some genomic contamination which resulted from cDNA preparation, whereas contamination was absent in the other negative controls.
Figure 5.10  PCR-RFLP analysis of nucleotide bases encoding residues #1188 and #1191, in cDNA samples from H11, H14 and H19. PCR fragments represented all C4 alleles. Figures 'A' and 'B' represent polymorphism of nucleotides encoding residues #1188 and #1191. PCR-RFLP fragments shown in Figure A were separated on 2.5% agarose gel and EtBr stained, whereas those in Figure B were separated on 8% polyacrylamide and silver stained. Lanes #3, #4 and #5, represent donors H11, H14 and H19, respectively, in Figures 'A', whereas lanes #3, #4 and #5 represent donors H19, H14 and H11, respectively, in Figure 'B'. Lane #2 in Figure 'A', and #1 in Figure 'B' are unrestricted PCR cDNA samples. pBR322 restricted with Msp I is the molecular weight marker shown in each figure.
Figure 5.11 C4A-specific amplification of cDNA samples from C4A Q0 allele donor H19. Figure 'A' shows C4A amplification of nucleotide bases encoding C4A isotypic residues #1101-#1106 to nucleotides encoding residue #1284 (primer pair 5, Table 5.1), where lane #1 represents donor H11, #2 is H14, #3 is H19, #4 is a genomic negative control (using primer pair 5, Table 5.1) donor 'EW-Vavy' (C4A Q0 resulting from deletion of the C4A gene), #5 is a cDNA positive control of donor H19 (using C4B-isotypic primer pair 6, Table 5.1), #8 is a PCR negative control without a DNA template. Figure 'B' shows C4A amplification including nucleotide bases for C4A isotypic residues #1101-#1106 to #1033 (primer pair #3, Table 5.1), where lane #1 represents cDNA PCR fragments of donor H11, #2 is H14, #3 is H19, #4 is a positive genomic control donor 'WCY1204' (primer pair #3, Table 5.1), #6 is a PCR negative control without a DNA template. Lane #1, #2 and #3 in 'B' have some genomic contamination of the cDNA preparation, whereas contamination was absent in the negative control. pBR322 restricted with Msp I is the molecular weight marker shown in each figure.
encoded by nucleotides for residues #1188 and #1191 (Figure 5.10A and B). The larger 544 bp fragment was clearly visible on a 2.5 % agarose gel stained with EtBr, whereas the 306 bp and 230 bp fragments were more readily detected by separation on an 8 % polyacrylamide gel which was silver stained (Figures 5.10A and B). Fragment sizes less than 60 bp could not be seen as they were lost during gel electrophoresis.

The molecular basis for the C4A Q0 allotype of Aboriginal donors has not been determined. However, PCR and PCR-RFLP analysis of cDNA from donor H19, identified C4A-specific nucleotides, thereby indicating that the C4A alleles are completely transcribed, (Figure 5.10A and B, and 5.11A and B). PCR-RFLP analysis of cDNA encoding residues #1188 and #1191 in donor H19, found nucleotides common to both C4A and C4B alleles, indicating that both C4A and C4B genes were completely transcribed (Figure 5.10A and B). In this analysis the C4A alleles were represented by 544 bp, 41 bp, and 16 bp fragments, whereas the C4B alleles were represented by 306 bp, 230 bp, 41 bp, 16 bp and 8 bp fragments, (Figure 5.10A and B).

C4A alleles were successfully amplified by two different pairs of C4A-specific primers (Figure 5.11A and B). The first pair of primers amplified a sequence region having nucleotide bases encoding C4A isotypic residues #1101-#1106 to nucleotide bases encoding residue #1284. This PCR amplification resulted in a fragment size of 574 bp (Figure 5.11A). The second pair of primers amplified the nucleotide base sequence region of C4A alleles encoding amino acid residue #1033 to nucleotides encoding the C4A isotypic residues #1101-#1106. The resulting PCR fragment size from this amplification was 262 bp (Figure 5.11B). Even though C4A-specific PCR-RFLP patterns were obtained by analysis of cDNA from donor H19 (A *Q0 B *I), it was not certain whether each C4A Q0 allele of the homozygous allotype was transcribed. Further PCR-RFLP analysis of nucleotide bases encoding residue #1267 is required to resolve this question as the C4A alleles of donor H19 were heterogeneous for nucleotides encoding this residue.
5.4.12 Association of the C4A *CAN1 allele with the C4B Q0 allotype in ten related Groote Island Aboriginal donors, and 50 non-related SLE Caucasian patients by PCR-RFLP analysis

A summary of PCR-RFLP analysis of each Groote Island Aboriginal donor is given in Table 5.7. PCR-RFLP banding patterns are the same as those shown in Figures 5.5 and 5.6. PCR-RFLP analysis of nucleotide bases encoding amino acid residues #1101 - #1106, #1157 and #1191 of ten C4B Q0 Groote Island Aboriginal donors, found that six of these donors carried the C4A *CAN1 allele. The remaining four Groote Island Aboriginals had similar PCR-RFLP patterns (Table 5.7) as C4B *Q0 allele donors ECY743 and ECY519 (Table 5.6). Each C4A *CAN1 allele donor had a DR 8 allele, while C4B *Q0 allele donors, with similar PCR-RFLP patterns to donors ECY743 and ECY519, had both DR 4 and DR 8 alleles. Unfortunately the inheritance of both types of C4B *Q0 alleles could not be shown as some family members were suspected to be unrelated.

A Caucasian SLE patient group of 50 donors (28 having a C4B Q0 allotype), were screened for the C4A *CAN1 allele by PCR-RFLP analysis of nucleotide bases which encoded amino acid residue #1188. Twenty three patients carried a C4B Q0 allele in a heterozygous state with either a B1 or B3 allele, four patients were homozygous C4B Q0, and the remaining patients had B1 and/or B2 alleles. Only one donor, 'SLE612', was found to have amino acid residue 'Ala' encoded by nucleotides for residue #1188, which was also detected in the C4A *CAN1 allele. This patient's ethnic background was English and New Zealand, and was the only donor with a DR 8 allele which was found also in Aboriginal donors from Groote Island having the C4A *CAN1 allele. These results suggest that the C4B Q0 allotype in Caucasians is not associated with the C4A *CAN1 allele, with the exception of donors who have DR alleles common to those found in Aboriginals.
Table 5.7 PCR-RFLP results for nucleotides encoding C4 polymorphic amino acids of Groote Island Aborigines

<table>
<thead>
<tr>
<th>Donor</th>
<th>C4 allotype#</th>
<th>#1101-#1106</th>
<th>#1157</th>
<th>#1191</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C4A</td>
<td>C4B</td>
<td>C4A</td>
<td>C4A</td>
</tr>
<tr>
<td>30017</td>
<td>A 3 B Q0</td>
<td>PCPVLD</td>
<td>LSPVIH</td>
<td>N&amp;S</td>
</tr>
<tr>
<td>30019</td>
<td>A 3 B Q0</td>
<td>PCPVLD</td>
<td>-</td>
<td>N&amp;S</td>
</tr>
<tr>
<td>30618</td>
<td>A 3 B 1,Q0</td>
<td>PCPVLD</td>
<td>LSPVIH</td>
<td>N</td>
</tr>
<tr>
<td>30620</td>
<td>A 4 B 1,Q0</td>
<td>PCPVLD</td>
<td>LSPVIH</td>
<td>N</td>
</tr>
<tr>
<td>40033</td>
<td>A 4 B Q0</td>
<td>PCPVLD</td>
<td>-</td>
<td>N</td>
</tr>
<tr>
<td>40034</td>
<td>A 3 B Q0</td>
<td>PCPVLD</td>
<td>-</td>
<td>N&amp;S</td>
</tr>
<tr>
<td>40036</td>
<td>A 4 B Q0</td>
<td>PCPVLD</td>
<td>-</td>
<td>N&amp;S</td>
</tr>
<tr>
<td>40041</td>
<td>A 4 B Q0</td>
<td>PCPVLD</td>
<td>-</td>
<td>N</td>
</tr>
<tr>
<td>40042</td>
<td>A 3 B 1,Q0</td>
<td>PCPVLD</td>
<td>LSPVIH</td>
<td>N&amp;S</td>
</tr>
<tr>
<td>40045</td>
<td>A 3 B 1,Q0</td>
<td>PCPVLD</td>
<td>LSPVIH</td>
<td>N&amp;S</td>
</tr>
</tbody>
</table>

# allotype determined by protein gel electrophoresis. '-' indicates absence of nucleotides which code for the respective C4 isotypic amino acid residues. PCR product encoding each C4 allele of a donor was used for analysis of nucleotides encoding residues #1101 to #1106. C4A specifically amplified PCR product was used for analysis of nucleotide bases encoding amino acid residues #1157 and #1191.
5.4.13 Frequency of the \(C4A\) *CAN1 allele in Aboriginal Australians from Cape York

To establish the frequency of the \(C4A\) *CAN1 allele in Aboriginal Australians, 50 donors from Cape York were screened by PCR-RFLP analysis using the restriction enzyme \(Aci I\). This enzyme distinguished between nucleotide bases encoding either amino acid residue 'Ala' or 'Val' for residue #1188 in the \(C4A\) alleles of each donor. Examples of the resulting PCR-RFLP banding pattern can be seen in Figure 5.6. Only two Aboriginal donors had a \(C4B\) Q0 allele in a heterozygous form with either a \(B\) 1 or a \(B\) 2 allele. Both of these donors had \(DR\) 4 and \(DR\) 5 alleles, which are common HLA class II alleles in Caucasian populations (personnel communication with Prof. Serjeantson, JCSMR). The remaining donors carried other functional \(C4B\) alleles which were detected by protein gel electrophoresis. From PCR-RFLP analysis of these donors, not one carried the \(C4A\) *CAN1 allele, indicating that very few \(A\) 3 or \(A\) 4 alleles are masking the \(C4A\) *CAN1 allele. The \(C4A\) *CAN1 allele is therefore more commonly associated with the \(B\) Q0 allotype.

5.5 Discussion

Association of the \(C4\) null allotype with disease emphasises the need to elucidate its molecular basis. Studies here and elsewhere, show that \(C4\) null alleles can result from large deletions or a small 2 bp insertion, as well as from \(C4\) isotypic gene duplication. It is possible that other undefined \(C4\) null alleles may be generated at the post-transcriptional level as has been demonstrated for \(C4\)-low protein producing strains of mice (Nakayama et al., 1990b; Pattanakitsakul et al., 1992a and b; Zheng et al., 1993), but such a phenomenon has not been demonstrated in man.

One type of \(C4B\) Q0 allele in this study was shown to comprise of a mixture of both \(C4A\)- and \(C4B\)-type polymorphic amino acid residues. The newly identified polymorphic residue, #306, together with PCR-RFLP analysis of other known polymorphic amino acid residues, enabled the molecular basis for the association of this \(C4A/C4B\) hybrid type of allele with the \(C4B\) Q0 allotype to be
determined. These results indicate that this unique combination of residues may enable the resultant protein to migrate into the anodal region by \( C4 \) protein gel electrophoresis, thus appearing as a \( C4B Q0 \) allotype. This novel \( C4A/C4B \) hybrid allele was thus denoted as \( C4A *CAN1 \).

Functional assays of the \( C4A *CAN1 \) and other \( C4 \) alleles with mixed \( C4A \) and \( C4B \) polymorphic amino acid residues are required. This analysis would determine if proteins from \( C4 \) alleles with mixed residue properties are functionally different from other more typical \( C4A \) and \( C4B \) alleles. Also, functional assays would determine the association of hybrid \( C4A \) and \( C4B \) alleles with different disease states.

As previously mentioned, \( C4 \) alleles with mixed \( C4A/C4B \) amino acid residues have been reported which are associated with the \( C4B \) null allotype. However, until now, the molecular basis for this type of association was unknown. Previous efforts were made to establish if these alleles were a part of a particular HLA-extended haplotype. However, the evidence was inconclusive. Partanen and Campbell (1989), detected a \( C4A *3, B *Q0 \) allotype having long \( C4 \) genes at both loci I and II, which is the same \( C4 \) gene organisation as detected for Aboriginal donors H11 and H14 from this study. They found that, as with donors H11 and H14, the \( C4 \) locus II gene had an allele with \( C4A \) isotypic nucleotide bases, and Chido and Rodgers antigenic determinants which were common to \( C4B \) alleles. Also, as with donors H11 and H14, this individual's \( C4B *Q0 \) allele was associated with the HLA class II \( DR 4 \) allele. However, as other family members were not informative, it could not be determined which haplotype this unusual allele was associated with, (either HLA-B 44, \( Bf S, C4A 3, C4B Q0, DR 4 \), or HLA-B 62, \( Bf S, C4A 3, C4B Q0, DR 4 \)). RFLP analysis together with PFGE, in a study by Dunham et al. (1989), found that a \( C4B Q0 \) allele in the extended haplotype HLA-B 44, \( C4A 3, C4B Q0, DR 4 \), was associated with the presence of a duplicated \( C4A \) gene at locus II. Studies here, and by Hammond et al. (1992), and Braun et al. (1990), found duplicated \( C4 \) genes which were functional by producing \( C4 \) protein. These studies indicate that the haplotype HLA-B 44, \( Bf S, C4A 3, C4B Q0, DR 4 \)
{detected by Partanen and Campbell (1989)}, may have a duplicated C4A allele as detected by Hammond et al. (1992). The other haplotype, HLA-B 62, Bf S, C4A 3, C4B Q0, DR 4, however, could be the one which is associated with the hybrid type C4A allele.

In addition to these studies described above, the extended haplotype HLA- B 44, DR 4, Bf S, C2 C, C4A 3, C4B Q0 was detected in 12 of 17 patients with non-deleted C4B Q0 (Schneider et al., 1986). However, whether or not this C4B Q0 allele was a duplicated C4A gene had not been investigated.

Of significance, is that a DR4 allele which is part of the extended haplotype HLA-B 44, C4A 3, C4B Q0, DR 4, associated with the C4B Q0 allotype, has been found with increased prevalence amongst patients with Felty's syndrome (Thomson et al., 1988). The molecular basis for the C4B Q0 allotype however, has not been investigated. Using advances made in the molecular basis of C4 null alleles, here, and by Hammond et al. (1992), and Braun et al. (1990), it would be possible to establish if this haplotype has duplicated C4A genes. This may represent the first case of disease association with this type of C4 gene organisation. The mechanism underlying this association was thought to be reduced efficiency of the classical-pathway-mediated immune complex clearance mechanisms (Morgan and Walport, 1991a). However, as Hammond et al. (1992) and Braun et al. (1990), have shown, the haplotype HLA-B 44, C4A 3, C4B Q0, DR 4 has a duplicated C4A gene, where both genes were functional, producing C4 protein. If the haplotype associated with Felty's patients contains the same C4 gene organisation as described by Hammond et al. (1992) and Braun et al. (1990), there may be an overproduction of C4A protein in patients with Felty's syndrome. In this case a breakdown of the classical-pathway-mediated immune complex clearance mechanism is highly unlikely, as the C4A protein has been shown to be more efficient than C4B in mediating binding of immune complexes in vitro to erythrocyte CR1 (Schifferli et al., 1987).
Results here show that both $C4B\ Q0$ allotypes, resulting from $C4A$ gene duplication or having $C4A\ *CAN1$ alleles, have functional alleles at each $C4$ gene locus. Therefore, these $C4$ allotypes are not strictly $C4$ deficient. However, homozygosity for $C4A$ alleles still results in deficiency of $C4B$ protein. $C4A$ and $C4B$ alleles have isotypic amino acid residues which are important for the differential binding of $C4$ protein to immune complexes or to hydroxyl-containing carbohydrate surfaces, such as those which occur on the surface of encapsulated microorganisms. However, there are no known functions which are exclusive for either $C4A$ or $C4B$ isotypes. In the case of those individuals with $C4A$ alleles at both loci I and II, this would result in higher concentrations of circulating $C4A$, and would be expected to partially compensate for $C4B$ deficiency. It is likely that deficiency of $C4$ protein in patients caused by deletions or mutation of alleles encoding these proteins, are the types of $C4$ null alleles which are associated with different disease states. Further functional studies of different $C4$ alleles are necessary to confirm these hypotheses.

Just as other previously identified $C4$ null alleles are associated with particular HLA-extended haplotypes, so too are the $C4$ null alleles described here. The new $C4A\ *CAN1$ allele is associated with two HLA alleles, including either $DR\ 4$ and $DR\ 8$ alleles. The $C4B\ Q0$ allotype resulting from duplication of $C4A$ alleles is also associated with both $DR\ 4$ and $DR\ 8$ HLA alleles in the Groote Island Aborigines and with $DR\ 8$ alleles in donor ECY519. Donor ECY743, in addition to the $DR\ 4$ allele, had a $DR\ 6$ allele. Of the SLE patients, only one had the $C4A\ *CAN1$ allele, and this allele was associated with the $DR\ 8$ allele, which is comparatively rare in the Caucasian population.

Studies here suggest that the molecular basis of nondeleted $C4B$ null alleles in Caucasian SLE patients is different from that of Aboriginal $C4B\ *Q0$ alleles, and these $C4B$ null alleles are associated with different $DR$ haplotypes more typical of those found in Caucasian populations. The non-deleted type of $C4B\ Q0$ alleles in Caucasians require further investigation.
Of the 50 Aboriginal donors from Cape York which had functional alleles at both C4 gene loci as determined by protein gel electrophoresis, not one had the C4A *CAN1 allele. This supports the hypothesis that the C4A *CAN1 allele is only associated with the C4B Q0 allotype. It is possible however, with gene duplication in the C4 gene region, that the C4A *CAN1 allele would be undetected by protein gel electrophoresis. For example, an individual with C4A *3, *CAN1, B *1,*2 alleles could be interpreted as C4A 3, B 1,2 by protein gel electrophoresis. This stresses the need for molecular analysis with PCR-RFLP to accurately determine the C4 allotype of some individuals.

In Caucasian population studies, only non-deleted C4B *Q0 alleles of the long type have been detected (Jenhani et al., 1992). Further analysis of these alleles by using the PCR-RFLP C4 allotyping protocol is necessary to establish if C4A alleles are present at both loci I and II as in the cases of Aboriginal donors ECY519 and ECY743 in this study.

It was not possible to establish if the C4 genes representing the B Q0 allotype in donors ECY743 and ECY519 were capable of producing RNA, as there was no way of distinguishing between the alleles at both C4 gene loci. It may be possible in future studies to test if both C4A genes at loci I and II are functional by a two-site immunoradiometric assay (Hammond et al., 1992). This assay measures C4A, C4B and total C4 protein concentration in sera. If each donor had a total of four C4A alleles (two on each haplotype), this assay would show a higher concentration of circulating C4A than an individual with two C4A alleles. However, there has been some criticism of this approach, as it is difficult to establish the base level for circulating C4A and C4B protein due to the variation of this level among individuals (Milis et al., 1992).

Molecular analysis of non-deleted C4A null alleles in this study, did not detect any unusual molecular properties which may result in a dysfunctional allele, nor did it provide an explanation for this allele's association with the C4A null allotype. However, this study has indicated that the defect may lie at the post-transcriptional level as
RNA studies of donor H19 found the C4A *Q0 allele to be fully transcribed. Further studies are required to establish if these alleles can produce the correct sized RNA transcript where all introns are successfully removed by splicing. Also, PCR-RFLP analysis of cDNA is necessary to determine if both C4A alleles of donor H19 are transcribed. The same RNA and cDNA studies are required for C4A Q0 allele donor ECY509.

It is uncertain as to where to begin searching for defects of these undefined C4 null alleles in the latter genetic processing stages. Knowledge of these types of defects in other human alleles or animal models may be useful in directing this research. Nakayama et al. (1990b), compared the transcriptional activity of C4 genes between high and low C4-protein-producing strains of mice and found no difference in transcriptional activity between high and low C4-protein-producing genes. However, they found that the steady state levels of C4 mRNA were ten times lower in low C4 protein strains than in high C4 protein strains. This suggested that the major regulation of C4 plasma levels occurred at the post-transcriptional level. Later studies by Pattanakitsakul et al. (1992a and b), discovered that the C4 mRNA of low C4 protein-producing strains was abnormally processed, resulting from a 200 bp insertion derived from a B2 sequence (short repetitive genetic element in rodents). Zheng et al. (1993), found a difference in the regulation of transcription of this C4 gene between two main sources of C4 biosynthesis including the liver and peritoneal macrophages. These animal model studies suggest it is possible that C4 Q0 alleles in humans which are able to transcribe mRNA, may be affected by other regulatory mechanisms at the post-transcriptional level, like those described in low C4 protein-producing strains of mice.

Determination of the molecular basis of C4B *Q0 alleles here, and for other C4 *Q0 alleles by other workers has been important for the development of the PCR-RFLP allotyping protocol. Each of these C4 *Q0 alleles can now be identified by either Taq I RFLP and/or by PCR-RFLP analysis. The potential use of PCR-RFLP as a molecular
C4 allotyping protocol is thus enhanced, and heterozygous carriers of these C4 'null' alleles can now be identified.

The genetic mechanisms involved with evolution of C4 *Q0 alleles are complicated, involving gene duplication and genetic recombination or gene conversion. The unusual C4A/C4B hybrid type of C4 allele has shown that C4 alleles can potentially be incorrectly ascribed by protein gel electrophoresis. Detection of this allele, associated with the C4B null allotype in the Aboriginal Australian donors from Darwin, suggests that the high frequency of C4B null alleles in the Darwin population could be accounted for, in part, by this hybrid type of allele.

In conclusion, studies reported in this chapter have increased the understanding of the molecular basis of C4 null alleles in Aboriginal populations, and have provided a means of identifying these alleles at the molecular level. Areas of further research have been identified and include the function of C4 proteins which are associated with the C4 Q0 allotypes described in this thesis, especially in relation to their association with different disease states.
Studies reported here have investigated the distribution of $C4$ alleles between two northern located Aboriginal Australian populations and three PNG populations from coastal, inland and highland regions. The distribution of $C4$ alleles was found to be distinctly different between the Aboriginal and PNG populations, and between each PNG population. A molecular-based $C4$ typing protocol was developed and its application has identified a number of novel and rare alleles. The molecular basis of the $C4B$ null allele in Aboriginal Australians has been established and found to result from either $C4A$ gene duplication or possession of a novel $C4A/C4B$ hybrid type of allele. The finding that the $C4A$ null allele can produce cDNA indicates the $C4A$ null allotype may result from a defect in the post-transcriptional processing of the allele.

Prior to this research, limited information was available concerning the distribution of $C4$ allotypes within Aboriginal Australian and PNG populations. The Aboriginal Australian populations were found here to be quite different to many of the other western Pacific indigenous populations having two frequently occurring $C4A$ alleles and two frequently occurring $C4B$ alleles. By comparison with the Aboriginal Australian populations each PNG population, especially from inland and highland regions, was almost monomorphic at the $C4$ gene locus by having one frequently occurring $C4A$ and one frequently occurring $C4B$ allele. Results here, and those by Ranford (1989), indicate that the distribution of the most frequently occurring $C4$ alleles in a population can contribute to understanding the origin of Aboriginal populations and their movement within Australia.

$C4A$ null alleles were found to be rare in PNG populations, they were absent in highland and inland populations and only a low frequency was detected in the coastal population. By contrast $C4A$ null alleles were found to be common in Aboriginal Australians. The frequency of $C4B$ null alleles in PNG populations was higher than that detected for the $C4A$ null alleles except for the inland population where $C4B$
null alleles were absent. The frequencies of $C4B$ null alleles for the highland and coastal PNG populations were slightly lower than those detected for the Aboriginal Australians.

The high frequencies of $C4A$ null alleles detected in Aboriginal Australian populations both here and by Ranford et al. (1987), suggest that this allele is common to these populations. Frequencies of $C4B$ null alleles in the Aboriginal Australian populations were similar to those found in the Caucasian population. However, molecular analysis indicated that these $C4B$ null allotypes identified in Aboriginal Australians were not true null alleles, but were instead, either duplicated $C4A$ alleles or a $C4A/C4B$ hybrid type of allele which has a similar PI value to a $C4A$ allele. Molecular analysis thus provides strong evidence for the conclusion that $C4$ protein allotyping by gel electrophoresis does not always give an accurate description of the $C4$ allotype.

A number of rarely occurring $C4$ alleles were identified in the Aboriginal Australians and PNG populations, some of which, were previously thought to be unique to other indigenous populations, for example, 'B 5' in Asians and 'B 96' in Chinese. The presence of these rare alleles could indicate a genetic link between different ethnic populations. This study has shown however, at the genetic level that it is not unusual for $C4$ alleles with different combinations of polymorphic amino acid residues to have the same or similar migration patterns by protein gel electrophoresis, thereby appearing as the same allotype. It is quite possible therefore that rare $C4$ alleles recorded in different ethnic populations may have different amino acid residue compositions, but the same or similar PI value. Consequently, it is important to establish the molecular basis of these rarely occurring $C4$ alleles, in order to define whether they have evolved independently, or if they have a genetic link. PCR-RFLP analysis would help to establish if the rare alleles identified in different ethnic groups have the same composition of polymorphic nucleotide bases and therefore verify any such genetic links between these ethnic groups.
A lack of Hardy-Weinberg equilibrium was detected for the distribution of $C4$ alleles in the East Cape York Aboriginal population, which is a population sample previously found to be in Hardy-Weinberg equilibrium at HLA-DR and HLA-DQ loci by Gao et al. (1992a). This demonstrates that the $C4$ protein gel electrophoresis technique does not always provide an accurate description of the distribution of $C4$ alleles in a population. In this particular study it's most likely that the inability to resolve $C4$ A3,4 heterozygous alleles has led to an overestimate of the number of $C4$ A3 alleles in this population. This provided further evidence of the need for a molecular-based $C4$ allotyping protocol.

Existing $C4$ sequence information used in conjunction with additional DNA sequencing results obtained here, enabled the development of a molecular-based protocol for $C4$ typing involving PCR-RFLP. This new protocol involved analysis of $C4A$ and $C4B$ alleles independently by amplification using primers which anneal to the $C4A$ and $C4B$ specific isotypic nucleotide bases. This strategy allowed ready identification of heterozygote and homozygote allotypes at the $C4A$ and $C4B$ gene loci that otherwise could be misinterpreted by protein gel electrophoresis.

The discovery of the polymorphic amino acid residue #306 was crucial to understanding the unique migration patterns of some of the rarely occurring $C4$ allotypes as polymorphism of this residue results in a charge change of the overall protein. Detection of polymorphic intron 9 nucleotide bases in this study provided a means for the identification of long and short $C4$ genes at the molecular level by PCR-RFLP. Sequencing of $C4$ alleles from intron 9 nucleotide base 280 to nucleotide bases encoding polymorphic residues #306 and #328 will determine which polymorphic residues are associated with short or long $C4$ alleles. To establish if the polymorphic residues #306 and #328 are associated with either $C4A$ or $C4B$ specific residues it is necessary to sequence from the $C4$ isotypic nucleotide bases to nucleotide bases encoding these residues. These additional studies would confirm the association of
polymorphic residues #306 and #328 with specific C4 alleles as reported in this thesis.

These studies have shown that each C4 allele consists of an array of polymorphic amino acid residues which appear to have come from ancestral C4 alleles by genetic recombination and thereby creating C4 allelic diversity. It is possible that the presence of the large 7 kb intron 9 in some C4 alleles and absence in others has contributed to instability in the C4 gene region, as this difference in length may cause misalignment and unequal cross-over of different, yet highly homologous C4 genes. The high 'G' and 'C' nucleotide base content in the C4d gene region may also contribute to the process of unequal cross-over between C4 alleles as these bases bind more strongly than 'A' and 'T' nucleotide bases (Yu et al., 1986).

PCR-RFLP typing of a number of C4 alleles (A 58~, B 12~, B 22~, B 4~, B 5~, and B 6~), for which the molecular basis was previously unknown, characterised the combination of polymorphic amino acid residues encoded by nucleotide bases of these alleles. These findings suggest that PCR-RFLP typing may provide a rapid method for identification of the genetic composition of other novel C4 alleles.

C4 allotyping results obtained by protein gel electrophoresis compared with those obtained later by PCR-RFLP analysis, were in some cases found to be conflicting. For example, alleles typed as A 6 were found to be A *58 by PCR-RFLP analysis. This finding highlights problems in the interpretation of C4 allotypes determined by protein gel electrophoresis. A typing protocol based on the unique sequence of each C4 allele cannot be easily misinterpreted and thus enables accurate C4 allele typing.

False interpretation of C4A or C4B allotype homozygosity will not result in deviation from Hardy-Weinberg equilibrium so long as the number of heterozygotes is double that of homozygous alleles. Deviation from Hardy-Weinberg equilibrium could, however, arise when heterozygous alleles, for example C4A 3,4, are interpreted as homozygous C4A 3 alleles. This would result in an increased number of homozygous C4A 3 alleles.
Molecular analysis has found heterogeneity for the \( C4A \ \text{3} \) allele, including the 'A *3a' allele (Yu, 1991), and the 'A *3b' allele (Anderson et al., 1992). In addition the molecular \( C4 \) typing protocol developed here, identified other polymorphisms of the \( C4A \ \text{3} \) allele, these were the 'A *3c~' allele, and the \( C4A \ *\text{CAN1} \) allele which co-migrates with \( C4A \ \text{3} \) alleles during \( C4 \) protein gel electrophoresis. Clearly, \( C4 \) protein gel electrophoresis is unable to resolve some alleles with different amino acid sequences but with similar PI values. Further complications can arise in interpretation of \( C4 \) protein gel electrophoresis patterns due to known duplication and deletion of \( C4A \) and \( C4B \) alleles.

It is highly likely that other polymorphic nucleotide bases remain undetected within the \( C4 \) gene. Only a single large (23 kb) \( C4A \ *\text{3} \) allele has been completely sequenced (Yu, 1991), and as such, no other sequence is available for comparison which might allow detection of other polymorphic nucleotide bases. All other \( C4 \) allelic sequence information has been provided from analysis of the polymorphic \( C4d \) region, and the region immediately flanking this site. It is clear that the \( C4d \) region is most important for the function of antibody/antigen binding of the \( C4 \) protein and therefore is the region of the gene most likely to have the majority of the \( C4 \) genes polymorphic residues encoded. As a result of these observations it can be hypothesised that by comparison with the polymorphic \( C4d \) region, fewer polymorphic residues will be identified outside of this region and are more likely to be allelic polymorphisms.

Detection of novel polymorphic nucleotide bases by sequencing presents a difficult task due to the length of the coding \( C4 \) gene (5.2 kb). However, molecular techniques for identifying nucleotide mismatches between a known and unknown sequence are available (Anderson et al., 1992), and may be applied to search the \( C4 \) gene for novel polymorphic nucleotides.

In future studies, if polymorphic nucleotide bases are found to be located too far from the isotypic region of the \( C4 \) gene for amplification with \( C4 \) isotypic primers, it may be necessary to use cDNA for the PCR-RFLP typing protocol. Analysis of cDNA may enable
amplification of distant polymorphic nucleotides as the introns are absent in this form of DNA. However, genomic DNA is preferred for PCR-RFLP typing as it is easily extracted from peripheral blood, and requires less preparation time in comparison to cDNA. An alternative method, compared to the use of cDNA for the analysis of distant polymorphic nucleotide bases, could be the use of allele-specific oligonucleotides which would specifically hybridise to the polymorphic nucleotide base of interest.

A molecular-based C4 allotyping protocol will prove to be very useful in future C4 population studies, as a more accurate account of the distribution of C4 alleles will be obtained, compared to the traditional method of C4 protein gel electrophoresis. A molecular-based protocol can identify alleles which appear to be the same by protein gel electrophoresis, and is able to determine which particular type of C4 null allele is present. In addition, the molecular-based protocol can identify hybrid types of C4 alleles which protein gel electrophoresis cannot.

Until these studies little was known of the molecular basis of C4 null alleles in ethnic groups apart from Caucasians. Only three types of C4 null alleles have previously been characterised: (1) Those representing approximately 60% of Caucasian C4 null alleles which result from large deletions of the entire C4 gene and usually an accompanying 21-hydroxylase gene; (2) Duplicated C4 genes associated with a C4 null allele at the alternative C4 locus and; (3) C4A null alleles with a 2 bp insertion resulting in a truncated message during transcription. Other studies have suggested another type of non-deleted hybrid C4 null allele consisting of amino acid residues common to both C4A and C4B alleles (Partanen and Campbell, 1989). The molecular basis of the hybrid type of C4 null allele and its association with the C4 null allotype according to protein gel electrophoresis, was determined in studies for this thesis.

Of the Aboriginal Australian C4 null alleles characterised in these studies, only those with extended HLA haplotypes having DR and DQ alleles commonly found in Europeans were found to have null alleles resulting from large deletions. Those individuals with
haplotypes more commonly found amongst Aborigines have non-deleted types of C4 null alleles. Some C4B null alleles are associated with duplicated C4A genes, as determined by RFLP and PCR-RFLP analysis. Other C4B null alleles were found to be associated with a hybrid type of C4A allele (C4A *CAN1) which co-migrates with the C4A 3 protein. Both types of non-deleted C4B null alleles were found to be associated with HLA-DR 4 and DR 8 alleles.

It is possible that the C4A *CAN1 hybrid type of C4 allele identified here, is the same or similar to the hybrid C4 allele reported by Hammond et al. (1992) and Braun et al. (1990), which was also associated with the DR 4 allele. Studies for this thesis provide a valid explanation for the C4A *CAN1 alleles' association with the C4B null allele. The evidence was provided by identification of the unique combination of polymorphic amino acid residues of C4A *CAN1. This included identification of the polymorphic amino acid residue #306 (not previously known to be polymorphic), which was important in establishing that C4A *CAN1 has a PI value the same as an C4A *3 allele. The C4A *CAN1 allele can now be identified by the C4 molecular allotyping protocol without the requirement for family studies which were necessary for Hammond et al. (1992), and Braun et al. (1990) to detect this type of hybrid C4A/C4B allele. Detection of hybrid allele-specific transcripts in this study proves that this novel type of allele is capable of producing an RNA transcript. However, further Northern analysis is required to determine whether the RNA transcript is of the predicted size. Rodgers and Chido antigenic determinant serological analysis would determine whether this protein is present in serum.

A number of studies in vitro have indicated selective roles for the protein products of C4A and C4B alleles under pathological conditions. Law, Dodds and Porter (1984) have demonstrated variations in binding specificity of C4A and C4B products, despite their high degree of amino acid homology (discussed in Chapter 1). C4A 3 is also known to be more efficient than C4B 1 in the inhibition of immune complex precipitation (Schifferli et al., 1986). However, only one functional study investigating the metabolic differences of C4A and C4B alleles in vivo has been performed
(Peake et al., 1989). Peake concluded that $C4B$ 1 was catabolised more rapidly than $C4A$ 3 in patients with pathological complement activation but not in control subjects. This difference, reported by Peake, was said to reflect the relatively greater extravascular distribution of $C4B$ at sites of immune complex deposition or, alternatively, different rates of catabolism of inactive $C4$ isotypes.

There are no reported cases which have compared the metabolic differences of different $C4A$ alleles and different $C4B$ alleles. These types of studies are necessary as they will also determine whether an individual with hybrid alleles, or duplicated $C4A$ or $C4B$ alleles, is more or less at an advantage with regard to elimination of immune complexes or microorganisms during an immune response. These types of studies may also resolve the uncertainty as to the role of $C4$ null alleles in the clinical situation. That is, whether $C4$ null alleles are directly associated with disease pathogenesis or whether they are associated with another gene causing the susceptibility. Analysis of $C4$ alleles in different indigenous populations have shown their are two common $C4A$ alleles ($A 3$ and $A 4$) and two common $C4B$ alleles ($B 1$ and $B 2$) in each population. It is possible that evolutionary forces have sort to maintain these alleles in each population as they may have a selective advantage in processing infectious agents during an immune response. Only functional studies comparing the metabolic differences of $C4$ alleles will resolve these questions described above. As few studies have investigated $C4$ allele products in vivo there is little experience with the complications which arise from working with this protein. In vitro studies of $C4$ protein can be difficult due to the instability of this protein. Possibly the instability of $C4$ protein reflects the lack of functional studies performed in vivo on this protein. However, Peake's study is encouraging as results indicated no variable loss of functional activity during purification between different $C4$ protein preparations.

A worthwhile study would be to determine whether the hybrid type of $C4A$ *CAN1* allele or duplicated $C4A$ alleles are associated with the $C4B$ Q0 allotype in Felty's Syndrome patients. This is because
previous studies have shown that the $C4B\ Q0$ allotype which is associated with this disorder, like that of the $C4$ hybrid allele and $C4A$ duplicated gene haplotype described here, is associated with the $DR4$ allele. Such a study would also seek to establish whether particular types of $C4$ null alleles are associated with different disease states.

Messenger RNA was identified for the $C4A$ type of non-deleted null allele by PCR analysis of cDNA which suggests that the $C4A$ null allele results from a post-transcriptional defect. This analysis indicated that the complete RNA transcript was produced. However, as the complete size of the transcript could not be detected by Northern analysis due to insufficient quantities of RNA in peripheral blood, it could not be shown that all introns were successfully removed by splicing. Therefore, the ability of this RNA transcript to translate protein is uncertain. Northern analysis of RNA isolated from liver, (the major producer of $C4$ protein), could be used to overcome this problem, however, liver samples from human donors with this type of $C4$ defect are not readily available. Therefore, it may be necessary to amplify, by allele-specific PCR, sections of the entire gene to ensure all introns have been removed.

Guinea-pig and murine models of $C4$ protein defects at the post-transcriptional level, as described by (Whitehead et al., 1983; Nakayama et al., 1990b, respectively), could be used as a guide for designing studies on genetic mechanisms responsible for $C4$ defective alleles in humans. Chido and Rodgers determinants have been detected on red blood cells of $C4$ deficient patients by Ballow et al. (1979), and O'Neill et al. (1979), however, these determinants were absent in plasma. These authors postulated there was an additional regulator locus for $C4$ deficiency. It is possible that this type of $C4$ deficiency may result from rapid consumption of $C4$ protein by a later acting component of the classical pathway.

The work described in this thesis has involved design and assessment of a molecular-based protocol for $C4$ allotyping which in conjunction with RFLP or PFGE analysis, can provide more information than any of the existing serological typing protocols. This
molecular protocol can now be expanded to cover other rarely-occurring $C4$ allotypes. The molecular basis of the Aboriginal Australian $C4B$ null allele has been determined, and results from two types of $C4A$ alleles including the novel $C4A\,*CANI$ allele. This thesis has contributed to knowledge of the complex genetics of $C4$, and of some evolutionary mechanisms underlying that complexity, and has indicated direction for future research in this field.
References


Bitter-Suermann, D., Kroenke, M., Brade, V., and Hadding, U. 1977. Inherited polymorphism of guinea-pig factor B and C4:
Evidence for genetic linkage between the $Bf$ and $C4$ loci. 
*J. Immunol.* **118**: 1822-1826.


Giles, C.M. 1985. 'Partial Inhibition' of anti-Rg and anti-Ch reagents. II. Demonstration of separable antibodies for different determinants. *Vox Sang.* **48**: 167-173.


Genetic variation in the HL-A system between Ainus, Japanese 
and Caucasians. In: Histocompatibility testing. Eds: Dausset, J. 

Restriction fragment length polymorphism caused by a deletion 
involving Alu sequences within the human a2-plasmin 

Moon, K.E., Gorski, J., and Hugli, T. 1981. Complete primary 
structure of human C4a anaphylatoxin. *J. Biol. Chem.* **256**: 
8685-8692.

Morel, Y., Bristow, J., Gitelman, S.E., and Miller, W.L. 1989. Transcript 
encoded on the opposite strand of the human steroid 21-
Natl. Acad. Sci. (U.S.A.)* **86**: 6582-6586.

Morgan, B.P. 1990. In: Complement. Clinical aspects and relevance to 

Morgan, B.P., and Walport, M.J. 1991. Complement deficiency and 

immunoassay for the quantitation of human C4 gene products. 
*Complement Inflamm.* **7**: 95-101.

Müller-Eberhard, H.J. 1986. The membrane attack complex of 

during meiotic recombination in the yeast *Saccharomyces 

Nakayama, K., Pattanakitsakul, S., Yokoyama, S., Kimura, H., Nonaka, 
M., and Takahashi, M. 1990a. Post-transcriptional regulation of 
complement C4 in low C4- producing strain of mouse. 


Olerup, O. 1990. HLA class II typing by digestion of PCR-amplified DNA with allele-specific restriction endonucleases will fail to unequivocally identify the genotypes of many homozygous and heterozygous individuals. Comment in: *Tissue Antigens* **36(2)**: 83-87.


Roos, M.H., Giles, C.M., Demant, P., Mollenhauer, E., and Rittner, C. 1984b. Rodgers (Rg) and Chido (Ch) determinants on human *C4*: characterisation of two *C4* B 5 subtypes, one of which contains Rg and Ch determinants. *J. Immunol.* **133**: 2634-2640.


Slightom, J.L., Blechl, A.E., and 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.


Winkler, H. 1930. Die Konversion der Gene. Fischer, Jena


