Methods for Investigating the Molecular Bases of Complement Terminal Component Deficiency

Andrew David Thomas

Thesis presented for the degree of MPhil

2012

Institute of Infection and Immunity

School of Medicine

Cardiff University

Heath Park

Cardiff, CF14 4XN

Acknowledgements

Summary

The purpose of this thesis is to identify and examine the molecular mechanisms which result in deficiency of the complement terminal pathway, a part of the innate immune system which can destroy pathogens by lysis through formation of the membrane attack complex (MAC). Patient samples with mutations in components C5, C7 and C8 were available and methods to study these mutations in either genomic, RNA or protein were developed.

Work on C5 Deficiency focused on a young African Xhosa woman with C5 deficiency, and identified a known nonsense mutation in exon 1 [c.84C>T; p.Q28X] and a novel missense mutation in exon 7 [c.754G>A] that leads to a conservative change [p.A252T]. Three intronic sequence variations were also identified. An in silico analysis of these mutations was made and a method for amplifying full length cDNA was developed.

C7 Deficiency studies focused on four Irish families for which a large genomic deletion was thought to be present. Mapping of the breakpoints and PCR across the deletion identified a 6.4kb deletion together with the insertion of a novel 8bp sequence [c.739+1262_1270-2387delinsGCAGGCCA]. A method for carrier detection was developed. Bioinformatics analysis indicated that the initial DNA breakage was Alu-mediated.

C8 Deficiency was investigated in three patients with C8β deficiency. Genomic studies revealed that two patients were shown to be homozygous for the common nonsense mutation in exon 9 [1282C > T; p.A428X. The other was a compound heterozygote for a novel duplication in exon 7 [c.1047_1053 dupGGCTGTG], and a previously reported nonsense mutation [c.298C > T; p.G91X] in exon 3.

Abbreviations

A   Adenine

aHUS   atypical haemolytic ureamic syndrome

APS   Ammonium persulfate

BLAST   Basic Local Alignment Search Tool

bp   Base pair

BS   Branch site

C   Cytosine

CCP   Complement Control Protein Repeat

cDNA   Complementary DNA

CHO   Carbohydrate

CRD   Carbohydrate recognition domains

CVF   Cobra venom factor

dH_2O Distilled water

DNA   Deoxyribonucleic acid

dNTP   Deoxynucleotide triphosphates

DTT   Dithiothreitol

ECL   Enhanced chemiluminescence

EDTA   Ethylenediaminetetraacetic acid

EGF   Epidermal Growth Factor like repeat

ELISA   Enzyme-linked immunosorbent assay

ESE   Exon splicing enhancer

FIMAC   Factor I/ Membrane Attack Complex C6/7 Module

G	Guanine
HCl	Hydrochloric acid
HRP	Horseradish peroxidase
Ig	Immunoglobulin
kb	Kilobase
KCl	Potassium chloride
kDa	KiloDaltons
LCCD	Late complement component deficiency
LDLRA	Low Density Lipoprotein Receptor class A
LPS	Lipopolysaccharide
MAC	Membrane attack complex
MACPF	Membrane attack complex proteins /Perforin like segment
MASP	MBL associated serine protease
MBL	Mannose binding lectin
MD	Meningococcal Disease
$MgCl_2$	Magnesium chloride
mRNA	Messenger ribonucleic acid
NaCl	Sodium chloride
NMR	Nuclear magnetic resonance
OPD	I ,2-*o*-phenylenediamine dihydrochioride
PCR	Polymerase chain reaction
PPT	Polypyrimidine tract
R	Purine
RESCUE	Relative enhancer and silencer classification by unanimous enrichment

rpm	Revolutions per minute
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SELEX	Systematic evolution of ligands by exponential enrichment
SR	Serine and arginine rich
SS	Splice site
TE	Tris-EDTA
TEMED	Tetramethylethylenediamine
T	Thymine
TAE	Tris-acetate-EDTA
TSP1	Thrombospondin type 1 repeat
UTR	Untranslated region
UV	Ultraviolet
V	Volts
v/v	By volume
w/v	Weight by volume
Y	Pyrimidine

Table of Contents

Chapter 1

Introduction

1.1 A History of the Complement System

The complement system is an essential part of the innate immune system and plays a significant role in host defence (Lambris JD., 1998).

In the late 19th century experiments with serum demonstrated the presence of an unknown component which could lyse bacteria.

Then, in 1896 a Belgian scientist, Jules Bordet, noted that whilst working with immunised animals that this unknown component could be further subdivided into two further categories. Using guinea pigs immunised with *Vibrio cholerae*, Bordet demonstrated that heat inactivated serum from immunised animals could have lytic activity restored following reconstitution with serum from guinea pigs that had not been immunised.

Therefore he surmised that two components are present, with one component losing it's lytic activity after heat treatment and showing no capability to adapt.

It was Paul Ehrlich who first used the term 'complement' after proposing that the component of serum which can be heat inactivated 'complements' the activity of what Ehrlich described as amboreceptors, which we now know of as antibodies.

The nature of complement was still a source of disagreement. Paul Ehrlich believed that each amboreceptor had a unique complement specific to the amboreceptor, whilst Jules Bordet believed that there was only one type of complement which could react with all amboreceptors.

The debate continued until it was noted that complement was capable of acting in combination with specific antibodies, or in an antibody independent manner, which was considered a clear indication that there was one form of complement.

1.2 The Complement System

The complement system, illustrated in figure 1, consists of approximately 30 proteins and is traditionally thought of as a system that forms a cascade to recognise and destroy pathogens. The ability to destroy pathogens is only one of several functional roles of complement
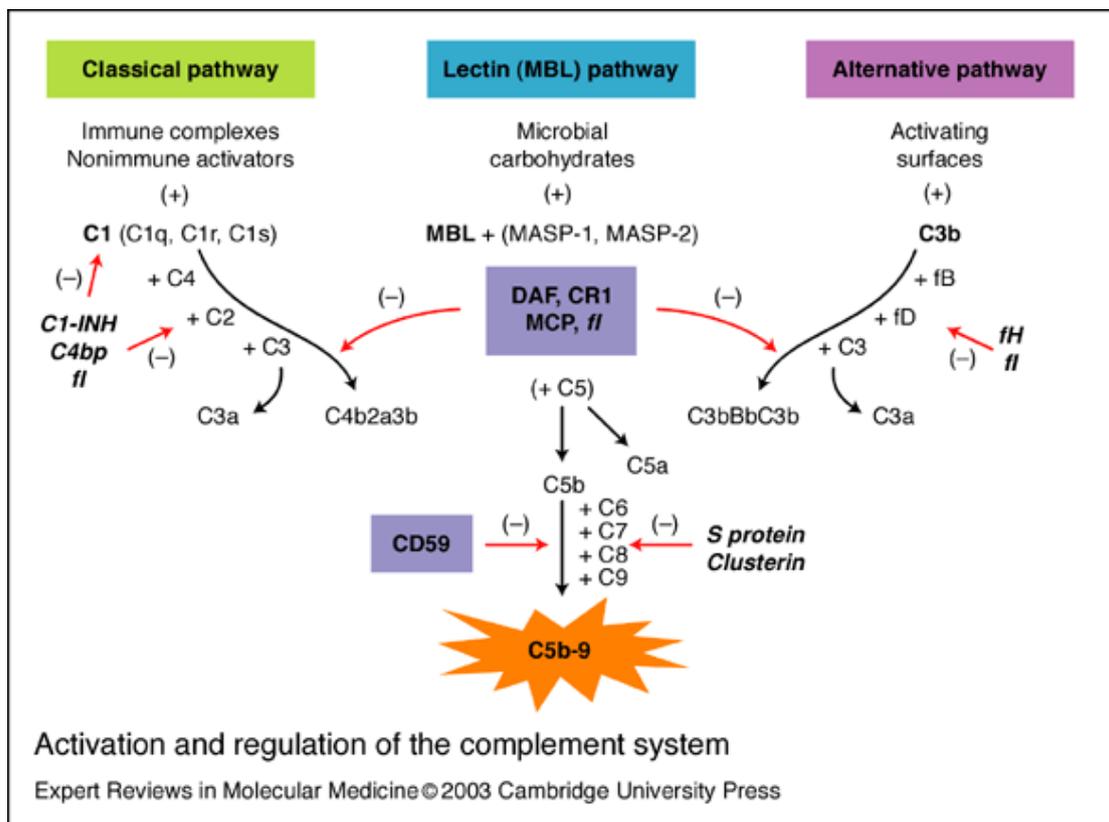
Figure 1. The complement system, taken from Gasque et al., expert reviews in molecular medicine 2003

These complement proteins are abundant and can be found freely circulating in the blood, and they can comprise around 4% of the total blood protein (Morgan and Gasque., 1997). The complement system is activated by several mechanisms, including the binding of C1q to Fc regions of immunoglobulins.

Following activation enzymes quickly cleave specific components to initiate the complement cascade (Lambris et al., 1998).

This then results in a comparatively large amplification of the cascade which in turn results in the formation of the membrane attack complex and then lysis of the target cell (Podack et al., 1984b).

Although other factors such as thrombin may lead to complement activation (Amara et al., 2010), the three main biochemical pathways that activate the complement system are the classical, alternative and mannose-binding lectin pathway (Morgan and Gasque., 1997).

The complement system performs several functions which can generally be described as;

Lysis: The formation of the membrane attack complex can result in the lysis of bacteria, damaged or infected cells and viruses (Cole and Morgan., 2003).

Opsonisation: This is a process where cells are biochemically marked by complement for removal by cellular immune components (Morgan BP., 2005).

Inflammation: Certain complement fragments such as C5a are powerful anaphylatoxins and can induce a powerful pro inflammatory response (Basta et al., 2003, Gerard et al., 1994, Hugli et al., 1978)

Innate-Adaptive bridging: The ability of complement to bind to antibody creates a significant link between the innate and adaptive immune systems (Morgan and Marchbank et al, 2005).

Immune clearance: The complement system plays a significant role in assisting the removal of immune complexes for eventual deposition in the spleen and liver (Bockow et al., 1981, Law et al., 1984).

1.3 The biosynthesis of complement Components

Although the majority of proteins that constitute the complement system are synthesised by the liver, there are significant amounts produced by other cells such as macrophages, monocytes, Neutrophils and some epithelial cells. The main sites for the synthesis of each component are shown in table 1 below

Component	Primary Synthesis Site	Secondary Synthesis Site
C1q	Macrophages	Follicular dendritic cells, cells of monocyte macrophage lineage
C1r	Liver	Epithelial, Endothelial and CNS Cells
C1s	Liver	Monocytes, Epithelial, Endothelial and CNS Cells
C2	Liver	Monocyte Macrophage cell lineages, Fibroblasts
C3	Liver	Monocytes, Macrophages, Fibroblasts, T Cells, Astrocytes, Adipocytes.
C4	Liver	Monocytes, Macrophages, Mammary Glands, Lung, Spleen, Kidney, Brain and Testis
C5	Liver	Lung, Spleen, Monocytes, Macrophages and Foetal Intestine.
C6	Liver	Macrophage
C7	Granulocytes, Liver	No major secondaries
C8	Liver	Monocytes, Macrophages, Fibroblasts and Astrocytes
C9	Liver	Monocytes, Fibroblasts and Glial cells
FACTOR B	Liver	Mononuclear Phagocytes and Fibroblasts
FACTOR D	Adipose Tissue	Monocyte Macrophage cell lineages
MBL	Liver	Kidney
MASP1	Liver	No major secondaries
MASP2	Liver	No major secondaries

Table 1. The primary and secondary biosynthesis sites for complement components. Adapted from Morley and Walport, The Complement Facts book, Academic Press 2000.

Biosynthesis for all complement components is not always a steady, continuous process. Many components, the most notable being C3, are acute phase reactants and the synthesis of these components is strongly dependant on the local cellular environment such as the presence and concentration of cytokines such as IL-1, IL-6, IFNα or Lipopolysaccharide(LPS) (Purwar et al., 2006 and Migita et al., 1990). Even at this point a further layer of complexity is added by individual cell type response where, for example C3 production by Neutrophils is actually down regulated by IL-1 (Botto et al., 1992). Furthermore, not all complement proteins are secreted rapidly, as demonstrated by Neutrophils which have been observed to store significant amounts of C7 (Hogasen et al., 1995). In the complement terminal pathway both C5 and C6 are relatively strong acute phase reactants compared to C7, and this can often lead to a molar deficit of C7 during an acute phase response (Lachmann et al., 1970).

1.4. The Complement Activation Pathways

There are three main pathways by which complement is activated, and these are the classical, alternative and mannose binding lectin pathways (Agostoni et al., 1992, Matsushita., 1996, Pangburn., 1983).

The classical pathway typically requires an antibody-antigeninteraction for activation which then binds to C1q (Sim and Reid., 1991). The antibody required to activate the classical pathway is either IgG or IgM, as other immunoglobulin classes do not activate complement via C1q. The ability of IgG isotypes to activate complement via C1q in descending order of efficiency

from IgG1>IgG3>IgG2>>IgG4 which is a poor C1q activator of complement (Bindon et al., 1988).

The C1 complex is formed when C1q binds to either IgM or IgG complexed with an antigen, or occasionally when C1q binds to a target cell surface (Sim and Reid., 1991). This leads to the activation of C1r and C1s, and two molecules each of these C1 subcomponents form another complex with the C1q complex creating macromolecular C1. It is the activation of C1 that leads to the cleaving of C4 into C4a and C4b. This then exposes a highly labile thioester bond within C4b, which under normal circumstances would be quickly hydrolysed (Law et al., 1997). In the presence of a suitable activating surface such as a microorganism or an immune complex C4b can form clusters near the activating C1 complex (Sim and Reid., 1991).

This bound C4b can then act as an acceptor site for C2, which is then cleaved by the C1 complex or MASP. This leads to the formation of C4b2a, a C3 convertase.

The alternative pathway, unlike the classical pathway, is an antibody independent activation pathway that can be triggered by various substances such as zymosan, teichoic acid or lipopolysaccharide (Pangburn 1983). In aqueous solution, C3 continually undergoes a low level of hydrolysis to form $C3(H_2O)$ (Lachmann, P. J. and R. A. Thompson., 1970). This is capable of binding factor B, which acts as a substrate for factor D. This in turn forms $C3(H_2O)Bb$, an unstable fluid phase C3 convertase capable of cleaving C3 to C3a and C3b. In the presence of a suitable activating surface, bound C3b will bind factor B. The C3bB can then undergo further cleavage by factor D to form the convertase C3bBb (Krishnan et al., 2009).

Given that the C3 convertase is formed by C3, this forms the basis of a continual and increasing amplification loop. The C3Bb complex then binds an additional C3b molecule which results in the formation of the C5 convertase C3bBbC3b (Gros et al., 2008).

Although the mannose binding lectin (MBL) pathway can be activated by the hydrolysis of C3, the presence of activating surfaces is the mechanism most readily associated with the activation of this pathway (Epstein et al., 1996). These surfaces normally present molecules such as bacterial carbohydrates or lipopolysaccharides (LPS). These are capable of binding to MBL and initiating the pathway.

The lectin pathway has many structural similarities to the classical pathway. The first classical pathway component C1q and mannose-binding lectin (MBL), share a similar 'bunch of tulips' structure as shown in figure 2 below.
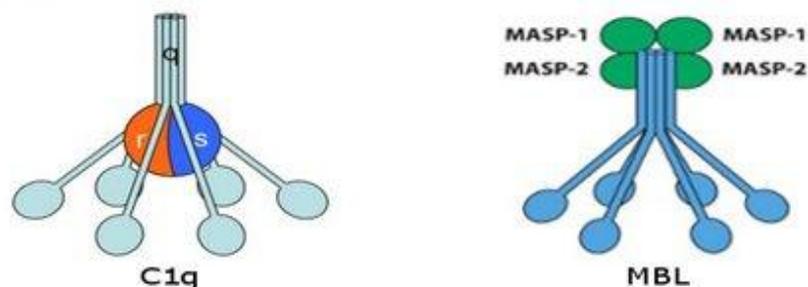
Figure 2: a comparison of the structures of c1q and mbl

In addition to this the MBL pathway shares C2, C3 and C4 with the classical pathway as the MBL-MASP complex can cleave C2 and C4.

This pathway is activated by binding mannose-binding lectin to surface containing mannose, LPS or carbohydrates on the pathogen surface, which activates the MBL-associated serine proteases, MASP-1, and MASP-2. This then cleaves C4 into C4a and C4b and cleaves C2 into C2a and C2b (Garred

et al., 2003). Following this C4b and C2a then form the C3-convertase, as in the classical pathway.

The activation of the classical, alternative or mannose binding lectin pathway involves a C3 convertase cleaving and activating C3, resulting in the formation of C3a and C3b which in turn results in an amplification of activation via further cleavage and activation. C3b has very significant opsonising capacity, which when bound to a cell surface leads to increased phagocytic killing by immune cells (Blom et al., 2009).

The complement fragments C3a and C5a are powerful anaphylatoxins capable of inducing a strong pro-inflammatory response. The cleavage of C5 results in the formation of C5a and C5b (Lambris et al., 1998).

It is C5b that can initiate the complement terminal pathway and form the membrane attack complex, which consists of C5b, C6, C7, C8 and C9. The membrane attack complex is the final product of the terminal pathway cascade, and it forms a transmembrane channel capable of causing lysis of the target cell by membrane disruption (Podack., 1984a) as shown in figure 3.
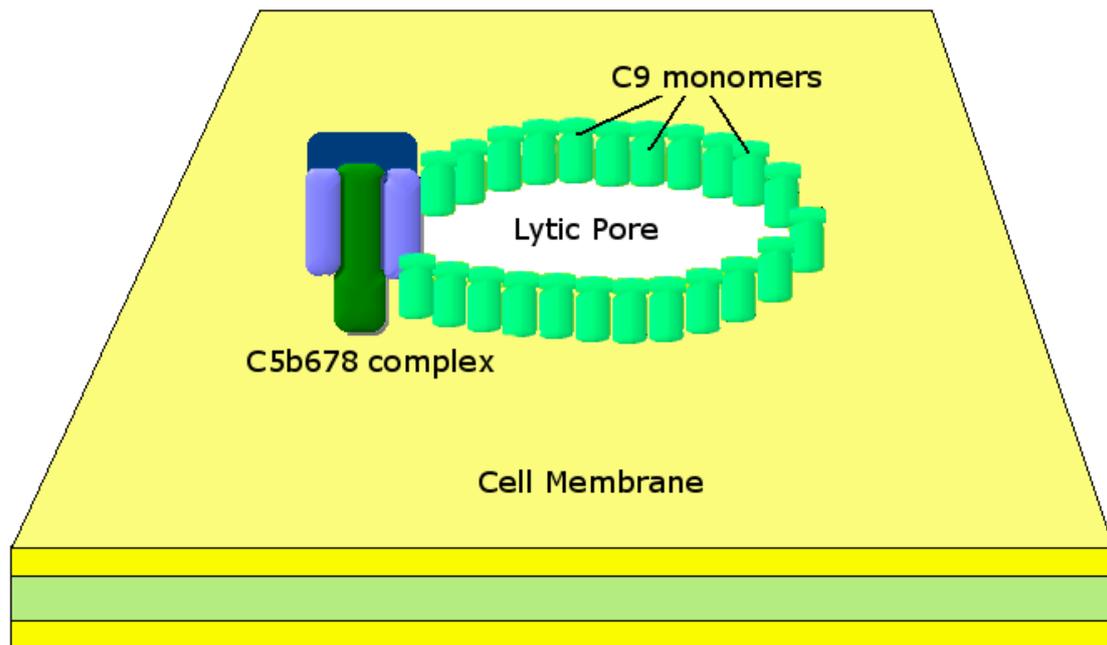
Figure 3. An illustration of the membrane attack complex. The lytic pore is formed by the polymerisation of C9.

The complement system can be extremely damaging to host tissues, and therefore its activation requires careful regulation. The complement system is regulated by complement control proteins, which are relatively abundant in serum compared to the complement proteins themselves (Morgan and Walport., 1991). Some complement control proteins such as MCP or DAF are present on the membranes of cells to prevent them from being subjected to autologous complement attack (Fosbrink et al., 2005). Specific terminal pathway regulators are clusterin and CD59, which inhibits C9 polymerisation during the formation of the membrane attack complex to avoid cell lysis (Morgan and Walport, 1991).

The three activation pathways converge at the formation of a C3 convertase. Given that C3 forms a critical part of the complement cascade, it's structure and relative position of the thioester bond is shown in figure 4 below. It is this thioester bond which governs the functional activity of the C3 molecule, and therefore differs from the terminal pathway which is largely dependant on

conformational changes. The classical pathway results in the formation of C4b2a which can cleave C3 to C3a and C3b (Morgan and Gasque., 1997).

Figure 4. A diagram representing the structure of C3. In addition to the position of the thioester bond, the N terminus (N), C terminus (c) and disulphide bonds (S) are also shown.

In the presence of an activating surface, membrane bound C3b can bind to factor B. The factor B is cleaved by factor D for form Ba and Bb. The complex C3bBb remains intact whilst the Ba fragment is released. This forms a positive amplification loop as C3bBb cleaves more C3 to begin the loop again. It is the addition of a second C3b molecule that creates a C5 convertase (Pangburn et al., 1983).

1.5 The Terminal Pathway

The activation of either the classical, alternative or MBL pathway would normally lead to the formation of a C5 convertase and the activation of the terminal pathway with the release of C5b and the strong anaphylatoxin C5a. (Gerard and Gerard., 1994) The chemistry of the terminal pathway differs from that of the C3 family of proteins. Whereas, for example, C3 and C4 have activity mediated by an internal thioester bond, C5 does not have such a bond and activity is dependant on molecular conformational changes (Gros et al., 2008). The terminal pathway as a whole requires conformational changes

11

resulting from binding interactions to work. Upon interaction of C3b in the convertase an acceptor site for C6 is exposed.

The binding of C5b to C6 forms C5b6 which in turn forms a complex capable of binding C7 (DiScipio et al., 1988). A further conformation change on the C7 part of the C5b67 complex allows binding to the target cell membrane (Mold C et al., 1998). The terminal pathway cascade continues when C8 binds to the C5b67 complex using the C8β part of the C8 molecule. A further conformational change allows the C8α part of C8 to bind to the cell membrane thus increasing the stability of the complex at the cell surface. The membrane attack complex (MAC) is then formed when C9 binds to the C5b678 complex and then C9 polymerises (Gros et al., 2008).

The binding of C7 to C5b6 allows C5b6 to be released from C3b. The hydrophobic portion of C7 that binds to the cell membrane allows the complex to localise once again to the target cell surface (Mold et al., 1998). The time between the release of C5b6 from C3b and binding of C7 to the target cell membrane is small but between these two events the complex is not physically bound to a cell surface. Therefore, between the release from the C3b complex and C7 binding event it is possible for the MAC to localise to a nearby cell surface and cause damage by a process known as bystander lysis (Koch-Brandt et al., 1996).

Upon the formation of the MAC on the cell surface, it is entirely possible that lysis will occur as a result of membrane disruption. It is also possible that a number of signalling pathways within the cell will be activated and a pro inflammatory response induced.

Several other non classical effects of the membrane attack complex have been reported. In some cases the membrane attack complex can cause cell proliferation or activation. It have also been noted that MAC deposition can induce resistance to further complement attacks, and that it can also induce or protect against apoptosis (Cole and Morgan, 2003).

The complement terminal pathway genes are located on chromosomes 1, 5 and 9 (Campbell et al., 1988) as shown in figure 5
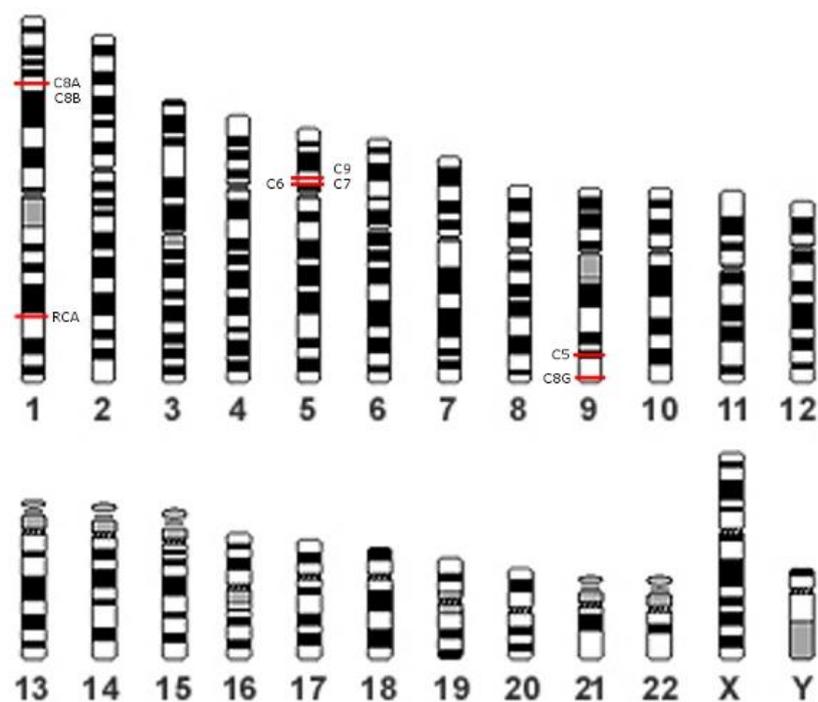
Figure 5. The location of terminal pathway genes and the Regulators of Complement Activation. Figure adapted from www.pubmed.com.

Complement C5

C5 biosynthesis occurs as a single chain precursor of 1676 amino acids which includes an 18 amino acid leader sequence. The precursor contains both the C5 α and β chains linked by an arginine rich linker sequence (Ooi and Colten, 1979). The leader and linker sequences are processed during secretion, at which point the mature C5 molecule adopts a two chain structure linked

13

together by disulphide bonds. The molecular weight of C5 is approximately

190kDa, and it is expressed primarily in the liver, although other sties such as

the lung, spleen and monocytes are also important. (Kohler and Müller-

Eberhard., 1967) The genomic location for C5 is 9q33 as shown in figure 5.
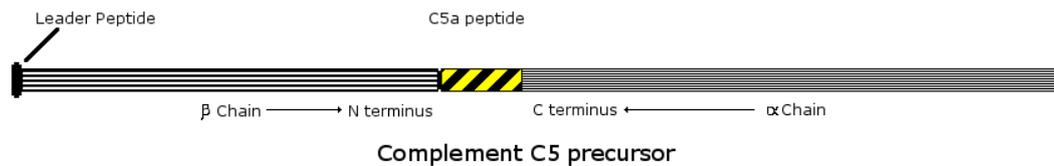
Figure 6. Complement C5 protein precursor. The arginine rich linker sequence is represented at the centre of the molecule

Although deficiency of C5 is rare compared to other terminal pathway

components, a number of defects have been reported as shown on Table 2

below

Gene	Mutation	Consequence	Location
C5	84C>T	Q19X	Exon 1
C5	892C>T	Q298X	Exon 9
C5	1115A>G	Splicing Defect	Exon 10
C5	1183_1184CAG>CTCT	Splicing Defect	Exon 15
C5	2536T>C	Missense	Exon 20
C5	4521C>T	PTC	Exon 36
C5	4884CCC>GC	Missense	Exon 40

Table 2. Mutation currently reported in the C5 gene

Complement C6

C6 is also synthesised as a single chain polypeptide of 934 amino acids. It

has a leader peptide, which is 21 amino acids long (Hobart et al., 1995). The

mature C6 molecule has a molecular weight of around 100kDa, and is

synthesised primarily in the liver. C6 is an acute phase reactant. The C6 gene

is located at 5p12, in close proximity to C7 and C9 (Campbell et al., 1988).
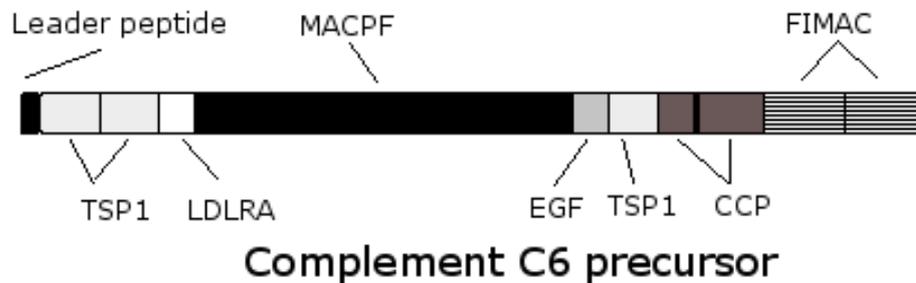
Figure 7. Complement C6 protein precursor. Figure abbreviations are listed in the main abbreviations section above.

The C6 gene has a number of defects associated with it, as shown in table 3

below

Gene	Mutation	Consequence	Location
C6	234delC	Termination Codon	Exon 2
C6	446+3A>C	Splicing Defect	Intron 4
C6	821delA	Termination Codon	Exon 7
C6	822delG	Termination Codon	Exon 7
C6	1138delC	Termination Codon	Exon 8
C6	1542T>A	Termination Codon	Exon 10
C6	1879delG	Termination Codon	Exon 13

Table 3. Mutations currently reported in the C6 gene

Complement C7

Complement component C7 is synthesised as a single polypeptide of 843

amino acids which includes a 22 amino acid leader sequence (Hobart et al.,

1995). It is a relatively weak acute phase reactant which can lead to a molar

imbalance with C5 and C6 during an acute phase response (Lachmann and

Thompson., 1970). C7 is synthesised primarily by granulocytes, with no

significant secondary synthesis site as yet being identified. The C7 gene

approximately 80kB and is located at genomic position 5p12-14 close to C6 in

a 3' to 3' orientation (DiScipio et al., 1988).
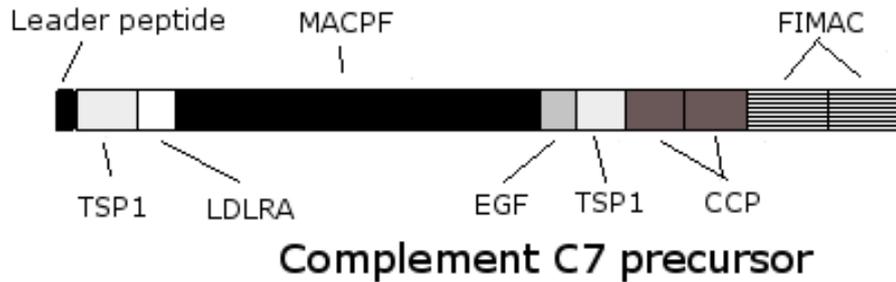
Figure 8. Complement C7 protein precursor. Figure abbreviations are listed in the main abbreviations section above.

The defects associated with complement C7 are shown in table 4 below.

Gene	Mutation	Consequence	Location
C7	189T>G	C41W	Exon 3
C7	1135G>C	G357R	Exon 9
C7	1561C>A	R499S	Exon 11
C7	1741-3delT	F569fs14X	Exon 12
C7	1924delAG	S620fs10X	Exon 14
C7	2107C>T	Q681X	Exon 15
C7	4009+2T>C	Splicing Defect	Intron 16

Table 4. Mutations currently reported in the C7 gene

Complement C8

C8 is unique in the terminal pathway insofar as it is synthesised from three

separate polypeptides, C8α, C8β and C8γ (Rosa et al., 2004).

C8α and C8 γ form a disulphide linked heterodimer, with C8β being linked non

covalently to form the mature C8 molecule. In a solution at high ionic strength

the C8α-γ complex can be isolated and purified. Each sub unit of C8 is

16

(Michelotti et al., 1995) encoded by a different gene, with the lipocalin C8γ

gene being located on a different chromosome (Kaufman et al., 1994). Of all

the complement proteins, C8γ is the only lipocalin. These molecules have a

roughly hollow hemispherical shape (from the Latin calyx meaning cup) and in

the case of C8γ the ligand for the central hollow space is yet to be identified.

The defects associated with each of the α, β and γ subunits of complement

C8 are shown in figure 9 below
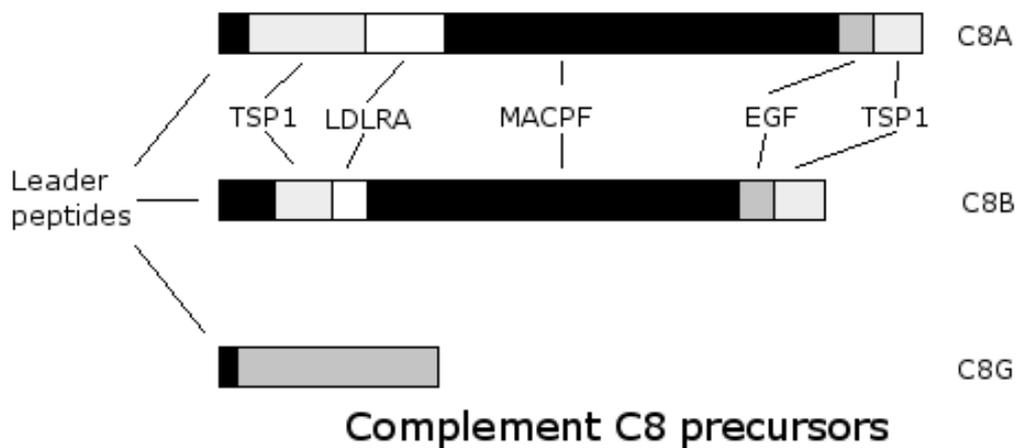
Figure 9. Complement C8 protein precursors

Gene	Mutation	Consequence	Location
C8A	952-12G>A	Termination Codon	Intron 6
C8A	952-15G>A	Termination Codon	Intron 6
C8A	267+1T>C	Splicing Defect	Intron 2
C8A	1407C>T	Termination Codon	Exon 10
C8B	271C>T	G91X	Exon 3
C8B	336delC	T112Tfs23X	Exon 3
C8B	361C>T	R121X	Exon 3
C8B	605delC	P202Rfs5X	Exon 5
C8B	820C>T	R274X	Exon 6
C8B	1047_1053dupGGCTGTG	L350Gfs8X	Exon 7
C8B	1282C>T	R428X	Exon 9
C8B	1309C>T	Termination Codon	Exon 9
C8G	None	None	None

Table 5 Mutations currently reported in the C8 genes

Complement C9

It is the ability of C9 to polymerise and form tube like structures that governs the lytic function of the membrane attack complex as shown in figure 2. C9 is synthesised as a single chain polypeptide that is located within 3Mb of the C6 and C7 genes. At 90Kb long, C9 shares considerable homology with a protein expressed on NK cells often referred to as C9 related protein (Müller-Eberhard HJ et al., 1987).
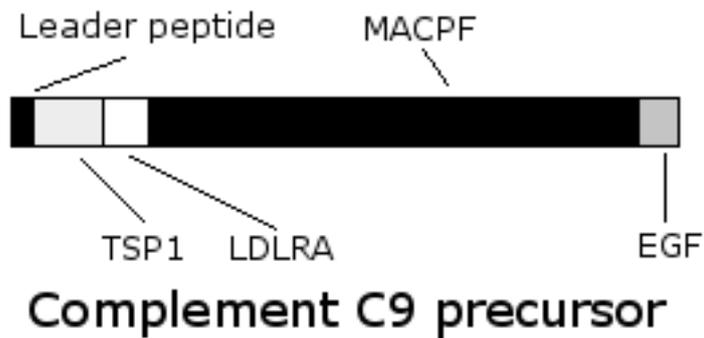

Figure 10. Complement C9 protein precursor. Figure abbreviations are listed in the main abbreviations section above.

The C9 protein has fewer reported defect than other terminal pathway proteins, which are listed in table 6 below

Gene	Mutation	Consequence	Location
C9	166C>A	C53X	Exon 2
C9	350C>T	R115X	Exon 4
C9	464C>T	R153X	Exon 4

Table 6. Mutations currently reported in the C9 gene

1.6 Complement Deficiency and Disease

Historically, complement deficiency has been considered when a patient has had a number of recurrent infections or an infection has taken abnormally long to clear (Colten and Rosen., 1992).

The complexity of the complement system is such that it plays a significant role in the majority of diseases, and therefore specific symptoms are sometimes difficult to attribute to complement deficiency (Lambris et al., 1998).

Deficiencies of components in the classical pathway can lead to an increased susceptibility to bacterial infections (Haeney et al., 1998), in particular polysaccharide encapsulated bacteria (Colten and Rosen., 1992) A deficiency of C1 inhibitor has been shown to cause hereditary angioaedema (HAE) (Agostoni and Cicardi., 1992). In addition to this, deficiencies of the classical pathway have been strongly linked to a disease which is similar to systemic lupus erythematosus (Manoharan and Madaio., 2010).

Mannose-binding lectin deficiencies are thought to be amongst the most common complement deficiencies, in some populations a prevalence as high as 30% of the population are reported as being deficient (Garred et al., 2003) It must be pointed out that in addition to complete deficiency the range of MBL present in human serum can vary between 5-10µg/L and 10,000µg/L and so the function point at which MBL deficiency can be defined may not be clear. In addition to this four variant allotypes of MBL exist and are relatively abundant

(Garred et al., 2003). A reduced MBL concentration has been linked to increased susceptibility to bacterial infections.

Alternative pathway deficiencies are also associated with an increased susceptibility to bacterial infections (Figueroa et al., 1991).

Specific complement component deficiencies are associated with particular diseases. Factor H deficiency is associated with a disease called membranoproliferative glomerulonephritis and atypical haemolytic ureamic syndrome(aHUS), and CD59 deficiency is associated with paroxysmal nocturnal hemoglobinuria (Morgan and Walport., 1991).

Complement terminal pathway deficiency leads to an increased susceptibility to Neisserial infections and recurrent meningococcal disease (MD). Although terminal pathway deficient individuals do experience an increased risk of disseminated gonorrhoea, it is the risk of recurrent *Neisseria meningitidis* infections that poses the most severe health risk (Morgan and Orren., 1998).

Complement sufficient individuals do not normally experience multiple *Neisseria meningitidis* infections, and although infection poses a risk to all age groups, a comparison of the age of first infection in both complement sufficient and deficient patients suggests that there is no difference between the age distributions of complement sufficient and deficient groups, implying that complement terminal pathway deficient individuals are a suitable research cohort for the study of Neisseria infections in humans (www.neisseria.org).

1.7 The molecular heterogeneity of Complement Components

1.7.1 Protein investigations of Terminal Complement Pathway Deficiency

A protein based study of each terminal complement pathway component should determine the nature of the deficiency. It is advisable to use a polyclonal antibody to the complement component of interest as it may be possible for an abnormal form of the protein to be missed by a monoclonal antibody (Lipman et al., 2005). The presence of an abnormal form of protein may also be seen if the patient serum sample is separated electrophoretically prior to immunoblotting. Normally a western blot can be completed within a day, producing a result strongly indicating terminal pathway abnormality

1.7.2 Genomic investigations of Complement Terminal Pathway deficiency

Although determining the molecular basis of the deficiency can often be achieved by the genomic PCR amplification of the sequence around known defects, a wider study may be necessary. It would comprise all exons and their surrounding intronic regions, the promoter region and both the 5' and 3' untranslated regions. The development of a method for generating these genomic PCR fragments was recommended. Occasionally, a compound defect may be present, such as a large deletion or insertion, which can be

obscured by the amplification of DNA from the unaffected allele (Ball et al., 2005). In addition to this there is the possibility that epigenetic modifications, such as changes in methylation patterns, may affect the production of complement components (Chen et al., 2007).

1.7.3 RNA investigations of Complement Terminal Pathway deficiency

Performing a study of mRNA may not always be a straightforward matter. In addition to the standard problems of producing cDNA and avoiding mRNA degradation, most terminal pathway components are synthesised in hepatocytes and these are not readily available from patients. Although biosynthesis does occur in many blood cell types, identifying the correct cell type for a given assay can be essential when the abundance of a particular transcript in a deficient individual may be of critical.

Even at this stage, the identification of a defect may not be clear. Some mutations further within an intronic region can have pathological consequences (Blyth et al., 2010), and some epigenetic modification can affect gene expression. On occasion, an observed mutation may not be strongly associated with splicing or result in a significantly different amino acid change (Rameix-Welti., 2007).

1.8 The aims of this study

The purpose of this study is to develop methods of investigating complement terminal pathway deficiency and to apply these method to patients suspected of having complement terminal pathway deficiency.

1.8.1 The investigation of C5 defects in a case of C5 deficiency in South Africa

Cases of C5 deficiency are rare in Western countries, with less than 100 being reported to date. In this section studies focus on a young Xhosan woman who was diagnosed with C5 deficiency after two attacks of Meningococcal disease (MD). I have collaborated in studies conducted on proteins, genomic DNA and cDNA samples.
In addition to this work was undertaken to ascertain the best source of patient sample material. This focused on the isolation or enrichment of specific cell types in order to maximise the potential to observe low levels of C5.

1.8.2 The characterisation of a large C7 genomic defect in the Irish Population

The identity of the C7 molecular defect in four Irish families was investigated (Thomas et al., 2012). Previous studies had showed a large genomic deletion,

and in this study the molecular identity of the deletion breakpoints has been determined and a novel 8bp insert identified.

1.10.3 The identification of C8β genomic defects in UK patients

A number of patients resident in the UK have presented with recurrent Meningococcal disease, and diagnosed with C8β deficiency (Arnold et al., 2009). In this study the molecular bases of these defects have been determined and the results suggest the frequency with which specific defects occur may be higher than expected.

Chapter 2

Experimental Methods

2.1 Isolation of nucleated cells from a sample of whole blood

The removal of erythrocytes and subsequent isolation of nucleated cells is a useful technique in the production of genomic DNA, RNA or the EBV immortalisation of B cells. Although serum does contain a little genomic DNA from cells that have undergone lysis, the highest concentration of genomic DNA within the blood exists within viable nucleated cells (Miller et al., 1988). Producing RNA suitable for reverse transcription requires a rigorous process; primarily due to the ease with which RNA can be degraded.

The method detailed here was designed to allow a lapse of about 12 hours between sample collection, and the isolation of nucleated cells. This enabled the protocol to be used when a blood sample needed to be transported back to the laboratory.

Method

Samples of about 15 ml of blood were collected into tubes containing heparin, and when necessary they were placed in a suitable container for transportation.

DNA isolation for a single sample was performed by taking 5 ml of the blood sample and gently mixing it with 5 ml of sterile PBS. The blood mixture was then overlaid very carefully onto 10 ml of Lymphoprep. It was this step that determined the quality and yield of nucleated cells.

The sample was centrifuged for 30 minutes at 900g, using low braking conditions and at room temperature. The serum lymphoprep interface

contains the buffy coat with nucleated cells as show below. This layer should be visibly cloudy if nucleated cells are present.
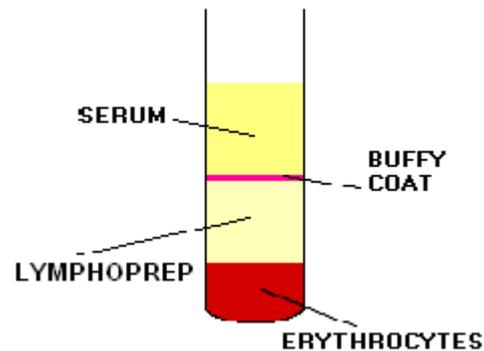
Figure 11. The visible layers produced after a lymphoprep gradient separation are represented. In practice the buffy coat is observed as a thin white opaque band.

The buffy coat was harvested into a sterile 15ml falcon tube using a sterile 1ml tip. An excess of sterile PBS was added to the tube to wash the nucleated cell sample, which was centrifuged for 10 minutes at 1300g and at 4°C. This wash step was then repeated another two times. The cells were resuspended in 1 ml of sterile PBS.

In order to assess cells isolated from the buffy coat, a haemocytometer was used.

The haemocytometer was loaded with a mixture of 10µL of cells and 10µL of trypan blue. The total cells were counted as well as the viable (non blue) cells. The cell density per ml is given as count x (2×10^4) and can be calculated for both total and viable cells. The two measurements were taken to assess the possibility of RNA degradation due to cell lysis.

The optimum number of cells for RNA isolation is 10^6 cells. Using the cell density counts above, the volume of sample necessary to provide 10^6 cells was calculated.

2.2 The isolation of Neutrophils from whole blood

Neutrophils are the most common cell type in blood circulation. They are short lived and play a significant role in complement biology such as responding to the presence of C5a (Ember et al., 1998), or in the synthesis of complement component C7 (Hogasen et al., 1995). Neutrophils can synthesis and store large amounts of C7, and it is for this reason that neutrophil activation should be avoided.

Method

When isolating neutrophils all steps were completed at 4°C and excessive vibration was avoided in order to limit activation. In addition to this, glass tubes were not used as the cells become adherent to the glass surface and cannot be recovered.

The method separated neutrophils in three steps:

The first step: Dextran sedimentation. This step removed most of the erythrocytes, which sediment at the bottom of the tube. The leukocytes and lymphocytes remained suspended in solution.

The second step: Hypotonic lysis. This step removed any remaining erythrocytes and platelets. The leukocytes and lymphocytes did not undergo lysis.

The third step: Lymphoprep sedimentation. It was during this step that mononuclear cells were separated from neutrophils. The neutrophils sank to

the bottom of the Lymphoprep, and mononuclear cells remained at the interface. It is worth noting that the mononuclear cells can also be harvested during this procedure.

Each sample was processed by firstly taking 20ml of whole blood which was then carefully poured down the inside of a 50ml conical tube containing 4 ml ACD buffer. The tube was gently inverted several times to mix the contents. Then, 12 ml of 6% Dextran solution was pipetted into the ACD/blood mixture and was then gently inverted 20 times to ensure the mixture is homogenised. The mixture was then pipetted into 4 x 15ml tubes, which equated to around 10ml per tube.

The tubes then stood at room temperature until separation occurred, which normally takes around 1 hour. The separation time can be reduced by centrifuging the tubes at 20g for 20 minutes at room temperature.

The supernatant was then pipetted into a 50 ml falcon tube, and centrifuged at 1150 rpm for 12 minutes at 4°C using low braking conditions.

Next, the pellet was resuspended in 12 ml of the ice-cold dH2O. The mixture was homogenised by repeated pipetting. Hypotonic lysis was carried out by adding 4 ml of cold 0.6 M KCl and mixing several times. The solution was diluted to 50 ml with cold sterile PBS. The tubes were then centrifuged at 1300 rpm for 6 minutes at 4°C using a high braking conditions.

The hypotonic lysis stage was performed up to three times to ensure complete removal of erythrocytes. Each stage can result is some cell loss, so the balance between cell yield and erythrocyte contamination was judged according to experimental requirements.

The supernatant was then discarded, and the pellet resuspended in 2.5 ml of cold sterile PBS.

Very carefully, the cell suspension was layered over 3 ml of Lymphoprep in a 15 ml tube. The tubes were centrifuged at 1500 rpm for 30 minutes at 4°C using high braking conditions. The supernatant was aspirated and the pellet resuspended. Cells remained viable suspended in HBSS for about 4 hours at 4C, or two hours at room temperature. Nevertheless, the cells were always used as quickly as possible.

2.3 The isolation of nucleic acids from blood and serum using Trizol phase separation reagent

The isolation of nucleic acids does not necessarily rely on the availability of whole blood. Although RNA stability is effectively negligible in dried blood and stored serum samples, it is possible to isolate genomic DNA from such sources. Although they do not contain nucleated cells and therefore are not preferred as high yield sources, they do contain DNA from cells that have undergone lysis.

The process of isolating nucleic acid depends upon the initial starting material, although in each case there was a method convergence at the phase separation level.

Whole Blood

The cells were first harvested using Lymphoprep as described above, and 10^6 cells were subjected to lysis using 1ml of Trizol reagent.

The lysate was then incubated at room temperature for 5 minutes to allow complete dissociation of nucleoprotein complexes. Then 200µL of chloroform was added to the tube which is then shaken vigorously for 15 seconds. The mixture was then incubated at room temperature for 5 minutes. The tube was then centrifuged at 12,000g for 15 minutes at 4°C. Upon removal from the centrifuge the contents had separated into an aqueous and organic phase. The upper phase was the aqueous phase and contained the RNA. The lower pink organic phase contained DNA and protein.

To isolate RNA, the colourless aqueous phase was then transferred to a clean tube and 500µL of isopropanol was added. The sample was centrifuged at 12000g for 30 minutes at 4°C. The supernatant was discarded and pellet was washed twice with 75% ethanol, after being centrifuged at 12000g for 30 minutes at 4°C between washes. The pellet was air dried and resuspended in 10mM EDTA 10mM Tris pH 8.0. Care was taken to avoid over drying the pellet as this significantly impeded resuspension.

When necessary the pink organic phase was used to isolate DNA, with the highest DNA concentration being near the interphase.

The organic phase was transferred to a clean tube and after complete removal of the aqueous phase the DNA was precipitated by the addition of 300µL of 100% ethanol. The sample was mixed by inversion and then

incubated at 15 to 30°C for 5 minutes. The DNA was precipitated by centrifuging at 2,000g for 5 minutes at 4°C.

The DNA pellet was then washed twice in a solution containing 0.1 M sodium citrate in 10% Ethanol solution. At each wash, the DNA pellet was incubated in the washing solution for 30 minutes at room temperature, and centrifuges at 2,000g for 5 minutes at 4°C. After these two wash steps, the DNA pellet was suspended in 2ml in 75% ethanol, and then incubated for 20 minutes at room temperature. The pellet was air dried and dissolved in 10mM EDTA 10mM Tris pH 8.0.

Dried Blood

The presentation of a dried blood sample, such as a blood spot on filter paper, can be used as a source of genomic DNA. The material containing the dried blood was excised with a clean sharp scalpel and incubated in 500µL of Trizol overnight with constant mechanical agitation on laboratory rollers.

The volume of Trizol reagent used for extraction was half of that used for extraction from whole blood, and therefore all downstream reagent volumes were similarly reduced by a half.

Similarly to whole blood, chloroform is added and DNA extracted from the organic phase as described above.

Dried blood was not a suitable source material for RNA extraction (unpublished result).

Serum

On occasion a serum sample was used in place of a white blood cell pellet as described above provided the serum volume is below 50µL. In cases where the sample volume used was more than 50µL, it was possible to centrifuge the sample under vacuum until the sample volume is below 50µL. All reagents and volumes used are then as per the whole blood method.

In common with dried blood samples, serum is not a suitable source material for RNA isolation.

2.4 The assessment of RNA quality

The risk of RNA degrading during the RNA isolation protocol or during storage means that RNA quality should be assessed prior to use. This was done either by measuring the RNA sample's UV absorbance or by RNA gel electrophoresis.

Method

UV Absorbance: The equipment routinely used to measure UV absorbance was a nanodrop ND1000 UV spectrophotometer. A 2µL sample of RNA elution buffer was first used to generate a blank reference for the nanodrop spectrophotometer. A 2µL sample of RNA was placed on the nanodrop measuring pedestal. The reader arm is then lowered and a UV absorbance measurement was taken. A UV absorbance measurement was taken for all

RNA isolations. The nanodrop generates a wavelength absorbance graph which was used to assess degradation, contamination or storage degradation by comparing graphs before and after storage.

RNA gel electrophoresis: This method was optimised for RNA fragments of between 2kB and 6kB. The percentage of agarose can be varied for RNA outside of this range.

It was possible to resolve the 18S rRNA and 28S rRNA bands in the RNA sample in order to assess potential degradation of the sample. In a good sample the 28S band has an intensity of about twice that of the 18S band.

A suitable volume of around 2L of l0x FA gel buffer was prepared to pour the gel. In addition to this around 2L of FA gel running buffer was prepared to fill the RNA electrophoresis tank.

Firstly, 200ml of agarose gel was made by mixing 2.2g agarose, 20ml, l0x FA gel buffer and then RNase-Free water was added to 200 ml.

The mixture was heated in a microwave, with the vessel swirled occasionally until the agarose was completely dissolved.

The agarose was cooled to 65°C in a water bath, and then 3.6 ml of 37% (12.3 M) formaldehyde and 2μL of ethidium bromide solution were added.

The gel was cast in a fume cabinet. It was ensured that the gel was completely set before removing the gel comb. The gel was placed in the electrophoresis tank which was then filled with FA gel running buffer. The gel was equilibrated in FA gel running buffer for at least 30 minutes.

In an RNase free sterile micro centrifuge tube, 5μL of 5x RNA loading buffer was added to 20μL of RNA sample, and denatured by incubating at 65°C for 4

minutes. Immediately after this incubation the sample was placed on ice for 10 minutes.

The RNA samples were loaded onto the gel and electrophoresed at a low voltage until the dye had migrated about 2/3 of the way across the gel. The gel was then visualized and photographed under UV light.

2.5 The reverse transcription of mRNA to produce cDNA

Although a number of RNA dependant DNA polymerase enzymes were commercially available, only Thermoscript and Superscript III were used. These products were supplied by Invitrogen (Paisley, UK). Thermoscript is a high temperature avian derived enzyme and Superscript III is a proprietary mutant of reverse transcriptase that is active at 60°C and has a half-life of 220 minutes.

Although it is possible to use gene specific primers to initiate the cDNA first strand synthesis reaction, the possibility of having causative mutations under gene specific primer annealing sites precludes this option. Therefore, two options were considered for priming the first strand synthesis reaction

Oligo(dT): a more specific priming method, was used to hybridize to 3' poly(A) tails, which are found in the vast majority of eukaryotic mRNA. Given that mRNA constitutes between 1 and 3% of the total RNA, the amount and complexity of cDNA produced is considerably less than with random hexamers. The cDNA synthesis performed in these experiments were performed with Oligo $(dT)_{12\text{-}18}$. Although Oligo $(dT)_{20}$ has been reported to give

more consistent results at higher cDNA synthesis temperatures (Veres et al., 1987), it has been reported that such reaction can be less robust (Noonan and Roninson., 1988).

Random Hexamers: Random hexamers are the most non-specific priming method (Breslauer et al., 1986), and are typically used when it is difficult to copy the full length mRNA. Using this method, all RNAs and not just mRNA's in a population are templates for first-strand cDNA synthesis. It is the PCR primers that confer specificity during the PCR amplification stage.

The length of cDNA produced was dependant on the concentration of random hexamers used in the first stage synthesis reaction. In generally, the lower the hexamer concentration the greater the maximum length of the cDNA produced. In order to maximize the size of cDNA produced, one should determine the ratio of random hexamers to RNA experimentally for each cDNA synthesis reaction.

Two guidelines for the generation of cDNA that may be beneficial are firstly, if generating cDNA at temperatures above 50°C using oligo(dT)$_{12-18}$, the amount of Superscript III may be raised to 2µL to increase yield (Houts et al., 1979). Secondly, the addition of an RNase inhibitor for low levels of template RNA is strongly recommended (Scheele et al., 1979). Using a recombinant RNase inhibitor with higher starting concentrations of RNA template did not produce any deleterious effects. It may therefore be used routinely during first strand synthesis.

Method

All reactions were set up for a 20µL reaction volume, utilising 100 to 500ng of total RNA. An initial mixture of 1µL of either of oligo(dT)$_{12-18}$ or 50ng of random hexamers, RNA template, 1µL 10mM dNTP Mix and sterile dH$_2$O to a volume of 13µL was produced. This mixture was heated to 65°C for 5 minutes and then incubated on ice for 1 minute. After being centrifuged briefly to collect the tube contents 4µL 5X First-Strand Buffer, 1µL 0.1M DTT, 1µL RNase inhibitor and 1µL of Superscript III were added. If random hexamers were used, then an initial incubation at 25°C for 5 minutes was performed. The first incubation was conducted at 50°C for 45 minutes to facilitate the reverse transcription, unless dealing with a difficult template where the incubation can be increased to 55°C. The reaction was heat inactivated by heating the cDNA to 70°C for 15 minutes.

The cDNA was then used as a template for amplification in PCR. However, amplification of the PCR target required the removal of RNA complementary to the cDNA, which was achieved by adding 1µL of RNase H and incubating the mixture at 37°C for 20 minutes. Wherever possible, primers were designed to overlap exon boundaries.

2.6 Polymerase Chain Reaction

This technique allows the amplification of large amounts of specific DNA from small amounts of initial template. The reaction undergoes a phase during which the amount of template doubles with each PCR cycle, assuming 100%

efficiency. As primers and dNTPs are consumed in the reaction, the increase in PCR product will slow down. Given that PCR product serves as a template for the reaction in the next PCR cycle and that typically 25 to 35 cycles of PCR can be performed, the amount of product capable of being produced is therefore large.

Method

A PCR mixture was produced in accordance with the components and amounts listed in Table 7 for a 50µL reaction. The PCR mix must be made in a clean DNase, RNase and pyrogen free tube certified for PCR. The components were thawed on ice and kept on ice throughout pipetting. The enzyme used for standard PCR amplification was Promega GoTaq (Southampton, UK) unless otherwise stated.

Component	Volume (µL)	Final Concentration
DNA template	0.5	Less than 250ng[*]
5x Buffer	10	1x
$MgCl_2$	In buffer	1.5mM
Primer (sense) 50pM µL^{-1}	0.25	12.5pM
Primer (sense) 50pM µL^{-1}	0.25	12.5pM
10mM dNTPs	1	0.2mM
GoTaq DNA Polymerase	0.25µL	1.25 units
dH_2O	To 50µL	N/A

Table 7. The individual components for a PCR reaction mixture. (*) The amount of DNA can vary significantly, although excessive quantities such as 250ng should be avoided.

Also additional magnesium chloride was used where specified. It was also possible to use a buffer without magnesium and add back magnesium in order to achieve concentrations below 1.5mM.

Thermal cycling normally consists of four steps. The first step is a single denaturation step of between 1 and 5 minutes. Then the following three steps are cycled, a denaturation step of 30-60 seconds, an annealing step of around 30 seconds and an extension step of around 60 seconds per kilobase. Once produced, the PCR reaction mixture was subjected immediately to thermal cycling. The number of thermal cycles used can vary but these are normally between 25 and 35 cycles. All thermal cycling was performed on an MJ Research DYAD PCR machine. MJ Research is now part of Bio-Rad Laboratories (Hemel Hempstead, UK). This machine was equipped with a heated lid, and therefore mineral oil was not required. The following thermal cycling profile is typical.

Initial Denaturing:	95°C	120 seconds
Denaturing:	95°C	30 seconds
Annealing:	55-65°C	30 seconds
Extension:	72°C	60 seconds per kilobase
Number of DAE cycles:	35°C	35
Final Extension:	72°C	600 seconds

On completion of thermal cycling the product was kept at 4°C, which is normally included in the PCR program. The annealing temperature used was dependant on the characteristics of the primer pair used.

2.7 Optimising Polymerase Chain Reactions

The use of polymerase chain reaction (PCR) is routine in both research and clinical settings. Extensive guidelines for the design of primers, purification of DNA template and choice of enzyme exist (Dieffenbach and Dveksler., 1995). It was noted, however, that the empirical refinement of specific PCR is often required. Although some elements are routinely optimised, it was found beneficial to consider optimising all regent components and all steps in the thermal cycling conditions for difficult reactions.

Method

The annealing step: This was optimised by running a PCR reaction with the annealing step set to a thermal gradient. This took the form of twelve identical reaction mixtures being subjected to nearly identical thermal cycling conditions in which the only variation was that each successive tube experiences a higher annealing temperature. An example of this would be where the first tube was subjected to an annealing temperature of 50°C and the twelfth tube was subjected to an annealing temperature of 60°C with the intervening tubes being subjected to incremental temperature increases defined by the manufacturer of the PCR thermal gradient cycling machine. Post PCR samples were electrophoresed on an agarose gel and imaged under ultra violet light to ascertain the optimum annealing temperature for the specified PCR reaction. Generally, the higher the PCR annealing

temperature, then the higher the reaction specificity (Dieffenbach and Dveksler., 1995).

Magnesium concentration: Magnesium ions are a critical cofactor for the function of thermostable DNA dependant DNA polymerases (Dieffenbach and Dveksler., 1995). Although other factors influence successful PCR reactions, the concentration of magnesium ions and annealing temperature form the basis of most optimisations. A range of magnesium concentrations between 0.5 and 4mM in increments of 0.5mM were explored. As the magnesium concentration increased, the efficiency of magnesium as a PCR cofactor increases to a maximum. Beyond this point the primers begin to anneal in a non specific manner, thus generating multiple non-specific products and eventually no product at all. The magnesium concentration and annealing temperature may be optimised in a single round of experiments using a 96 well PCR plate as shown in figure 12 below
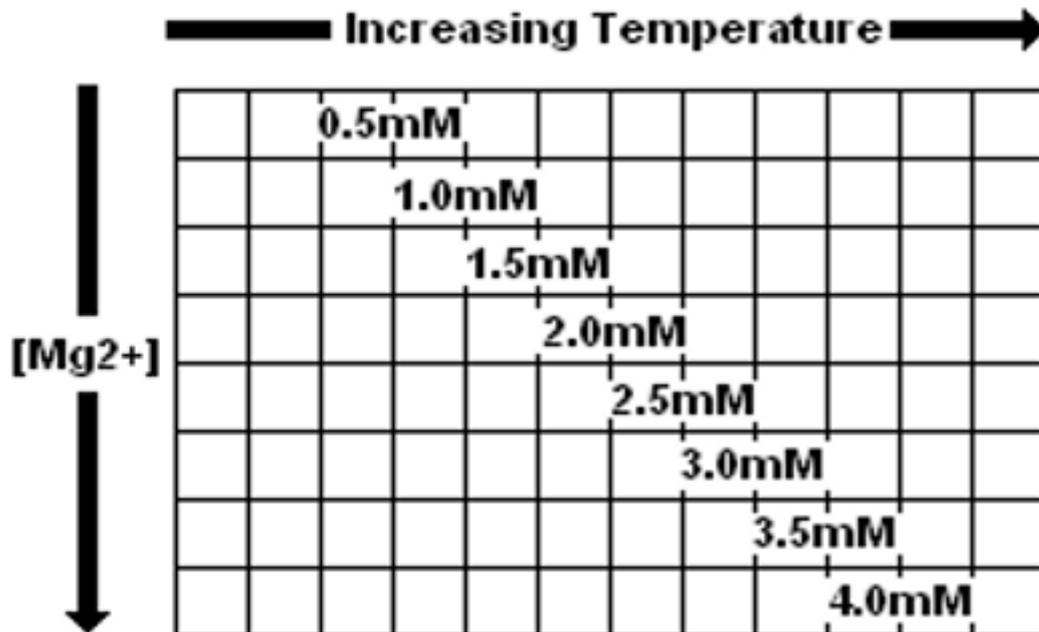
Figure 12: A schematic diagram of an initial PCR optimisation plate. This enables the optimal annealing temperature and magnesium chloride concentration to be quickly identified.

Buffer Optimisation: Commercially available polymerases are supplied with both a specific buffer and instructions relating to buffer concentration. It is possible to change the buffer concentration from between 0.5 and 2 times that recommended by the manufacturer. As the concentration of buffer increased the specificity of the PCR reaction also increased (Dieffenbach and Dveksler., 1995).

The type of buffer may be changed depending on the nature of the PCR to be attempted. An NH_4 based buffer tends to result in a more robust reaction, whilst a KCl based buffer will result in a reaction with higher specificity.

Deoxynucleosides (dNTPs): Most manufacturer protocols will recommend, for example, a dNTP concentration of 0.2mM for each component when performing a PCR with 1.5mM $MgCl_2$. Although this can be optimised there is one critical aspect that must be observed. In order for the reaction to succeed free magnesium ions are required, and at the point where the dNTP to magnesium concentration breaks parity the reaction becomes rapidly inhibited. Although the magnesium concentration can be increased to counter this, the concentrations listed above should yield sufficient product for most downstream applications.

Also dNTPs are prone to degrade with excessive freeze thawing, and should therefore be stored in small aliquots prior to use.

Primer Design and Concentration: Although many PCR primer design programs are available free of charge via the internet such as primer3 (http://primer3.sourceforge.net/), there are some aspects of the design process that benefit from careful selection (Innis et al., 1999). The primers should not contain sequences that readily self anneal or dimerise, and they

should have a GC content of between 40-60% and similar melting (T_m)

temperatures (Innis et al., 1999) . They should also be of similar length, with

around 18 to 25 bases being typical. Modern sequence analysis has allowed

two additions to the design process. Firstly, the target template can be

scanned for regions of repetitive DNA and this should then be avoided

(Cordaux and Batzer., 2009). Secondly, the primers themselves can be

subjected to a BLAST (Basic Local Alignment Search Tool) to ensure that the

primer sequence is complimentary to the target annealing site only (Altschul

et al., 1990).

The required primer concentration should be determined empirically. The

presence of strong primer dimer bands when imaged under UV light can

sometimes be indicative of excessive primer concentration.

Polymerase: When choosing the polymerase for a reaction, the end

requirements of the reaction dictate which polymerase to use. A simple PCR

based screen requires a robust reaction where fidelity is less important than

robustness, so an enzyme derived from Thermus Aquaticus (Taq) may be

most appropriate. In a reaction where the PCR product is to be cloned a

higher fidelity enzyme derived from an organism such as Pyrococcus

Kodakaraensis (Pfx) or Pyrococcus Furiosus (Pfu). In addition to a higher

fidelity, both enzymes possess a 3' to 5' exonuclease activity by which they

can exhibit a proofreading ability. In addition to this Taq will generate a

product with a TA overhang, whilst the other two enzymes will generate a

blunt ended product (Sambrook and Russell, 2001).

The majority of DNA dependant polymerases sold commercially are produced

using recombinant DNA technology. This has enabled a range of polymerases

to be generated which have specific mutations which enhance certain aspects of the behaviour of the polymerase (Merkens et al., 1995). Some modifications may enhance the ability of the polymerase to function in the presence of molecules normally regarded as PCR inhibitors, whilst others may have an influence on fidelity or the maximum possible size of the PCR product produced.

Finally, it has often been noted that heating the initial PCR mixture from ice temperature to an initial denaturing step of around 95°C takes a finite amount of time. During this time the reaction mixture passes a warm phase below the annealing temperature of the reaction where non specific primer annealing may occur. This can lead to the generation of non specific PCR products, and can be eliminated using a PCR hot start (Chou et al., 1992). Although this can involve simply placing reaction tubes onto a pre heated thermal cycler, many polymerases have an inhibitory anti-polymerase monoclonal antibody included which breaks down at the higher denaturing temperatures an thus provides an automatic hot start.

Adjuvants: Supplementing the PCR reaction mixture with additional components can often be useful. The following list of adjuvants is commonly used but not exhaustive.

Adjuvant	Concentration	Notes
Dimethyl Sulfoxide (DMSO)	2-10%	
Betain	1-1.5M	Do not use Betain HCl
Glycerol	1-10%	
Non ionic detergent	0.1-1%	Triton X100, NP40, Tween20
Formamide	1-10%	
BSA	<10%	

Table 8: a selection of some commonly used PCR Adjuvants

The use of these additives can produce unwanted effects. Although DMSO can assist with PCR amplification, particularly in GC rich templates, it can reduce the activity of Taq at higher concentrations (Appendix III). Betain and Betain monohydrate form the basis of many commercial PCR enhancing solutions, although it should be noted that Betain hydrochloride is a PCR inhibitor (Appendix III). Glycerol can help with the reducing of secondary structure, as can many non ionic detergents. Non ionic detergents can also help to neutralise any SDS carried over from DNA purification protocols. Occasionally at higher concentrations, formamide appeared to reduce the activity of Taq (Appendix III). The use of BSA to stabilise PCR enzymes has proved valuable, particularly where DNA has been extracted from ancient sources and so may be contaminated with inhibitors such as melanin (Appendix III).

In all cases, it is wise to optimise the concentration of adjuvants and to select the lowest possible concentration of any additives in order to reduce any negative effects the additive may have.

2.8 Automated Fluorescent Sequencing

The determination of genetic sequence is critical to the identification of genetic sequence variants and the identification of pathological mutations. Although chip based sequencing techniques are available, a common method of generating genetic sequence data relies of the Sanger di-deoxy sequencing. First published in December 1977 (Sanger et al., 1977), this

technique employs the use of both normal dNTPs and ddNTPs each labelled with a specific fluor. Thus, for any given nucleotide, the DNA synthesis can either continue following the inclusion of a dNTP or terminate with the inclusion of a fluor labelled ddNTP. Since each ddNTP is labelled with a specific fluor, as shown below, the terminating nucleotide for each synthesised strand bear a fluor label.

Figure 13: The ddNTP fluor labels used in automated fluorescent sequencing.

Given that each fluor labelled DNA strand is of a different length depending on the point at which synthesis termination occurred, the separation of these strands by capillary electrophoresis will then allow the determination of DNA sequence as shown below in Figure 14.

Figure 14: A schematic representation of the Sanger di-deoxy sequencing method.

The process of determining genetic sequence by this method has been simplified due to the availability of commercial kits and reagents such as Big Dye available from Applied Biosystems.

Method

The concentration of a purified PCR fragment was determined by UV spectrophotometry using a nanodrop ND1000. Then, using 500ng of DNA template, a sequencing reaction mixture was set up as shown in the table below.

Component	Volume (μL)
DNA temple	Volume equivalent to 500ng
Big Dye v3.1	2
ABI Buffer	4
Primer	0.5
dH2O	To 20
Total Volume	20

Table 9: A standard sequencing reaction mixture

The mixture is placed in a PCR tube and centrifuged briefly to collect the

contents. It is then subjected to;

95°C for 2 minutes and then,

35 cycles of

95°C for 30 seconds

50°C for 10 seconds

60°C for 4 minutes

This was followed by rapid cooling to 4 °C which is held until hold until ready

to purify.

The tube contents were centrifuged briefly in order to prepare for the

purification process. Firstly, 60μL of absolute ethanol is added to the tube. It

was important to keep the final ethanol concentration between 67-71%. The

tube was placed on ice for 30 minutes and centrifuged at 13000g for 30

minutes at 4°C. The supernatant was carefully removed and the pellet, which

was not always visible, was washed with 300μl of 70% ethanol. The

supernatant was once again removed and the wash step was repeated. The pellet was dried under vacuum. The electrophoresis step using an ABI 3130XL was performed by a Cardiff University core managed facility, central biotechnology services.

2.10 SDS-PAGE Gel Electrophoresis

The detection of proteins by SDS-PAGE gel electrophoresis is a technique widely used in biological research to separate proteins according to their electrophoretic mobility (Shapiro et al., 1967). The SDS gel electrophoresis of samples relies on those samples having identical charge per unit mass, and due to binding of SDS results in separation by size.

Method

An SDS polyacrylamide mixture was up as follows.

Upper Gel: 2.4mls upper buffer, 1.252mls of 40% acrylamide, 6.14mls of dH_2O, 0.1mls of ammonium persulfate and 0.01mls of TEMED.

Lower Gel: 7.5mls lower buffer, 7.5mls of 40% acrylamide, 14.5mls of dH_2O, 0.3mls of ammonium persulfate and 0.03mls of TEMED.

The ammonium persulfate solution was added to each mixture last as it initiates polymerisation. It was not until the lower gel had set that the upper gel had ammonium persulfate added.

The percentage of acrylamide can be reduced in order to observe lower molecular weight products.

The SDS Page gel equipment used was a Bio-Rad mini protean II vertical gel system (Bio-Rad Laboratories, Hemel Hempstead, UK)

The gel holder was assembled after all surfaces and glass plates were cleaned with 70% Methanol and dried thoroughly. Assembly occurred in accordance with the manufacturer's instructions.

The comb was placed the gel plates and a spot about 5-8mm below the bottom of the comb was marked. This was the level to which the lower gel was poured. The comb was removed and the lower gel poured.

After 45 minutes a check was made for polymerisation then the top of the gels were rinsed with distilled water.

The upper gel was then poured and the comb re inserted between the plates. After polymerisation the comb was removed and the gel clamped into the electrophoresis apparatus.

The sample was increased to 30μL with running buffer and loading dye and then heated to 95°C for 5 minutes before loading on the gel. The sample was then loaded onto the gel, taking care to avoid spillage. A suitable protein marker ladder was also loaded onto the gel.

The gel was then electrophoresed at between 100 and 180V until the loading dye front approached the bottom of the gel. Care should be taken to avoid the gel front bowing or distorting during electrophoresis, which is normally caused by a high current or voltage.

2.10 Western Blotting

Western blotting is an analytical technique used to detect specific proteins in tissue homogenate, tissue culture supernatant or other sample capable of being electrophoresed on an SDS PAGE gel (Renart et al., 1979). It uses gel electrophoresis to separate proteins by the length of the polypeptide which occurs under denaturing conditions or by the protein structure in non denaturing conditions. All proteins are then transferred to a nitrocellulose membrane, where they are then detected using antibodies specific to the protein of interest. Prior to transferring protein onto the nitrocellulose membrane the protein must first undergo SDS PAGE gel electrophoresis as described in Experimental Methods.

Method

Electrophoresis Transfer using Wet Blots

The gel and the nitrocellulose membrane were soaked in cold transfer buffer for 15 minutes to remove electrophoresis salts and detergents. In addition to this two filter papers of the same size as the gel were also soaked in transfer buffer. The transfer cassette is loaded as follows.

A fibre pad, a filter paper, the SDS PAGE gel, the nitrocellulose membrane, another filter paper and another fibre pad.

The cassette was loaded into the electrode module, ensuring it is loaded into the correct polarity orientation. The cassette was loaded into the buffer tank together with a magnetic stir bar and integral ice packs. The electrophoresis

tank was filled with transfer buffer to the appropriate level, and then placed on a magnetic stirrer at a low speed. The electrodes were attached to the tank and which was then run at 100 V for 1 hour.

The membranes were rinsed four times with dH_2O and blocked with 5% non-fat milk-TBST solution for at least 30 minutes, although overnight is acceptable, while agitating on a mechanical rocker. The membranes may be stored in TBST at 4°C for up to 1 week.

The blocking solution was removed and the membranes rinsed in TBST once. Then 500µL of 1/1000 dilution of primary antibody diluted in TBST was added. It was occasionally necessary to optimise the amount of antibody used for each protein of interest. The membranes were then left at room temperature for 30 minutes while shaking, after which they were then washed 4 times for five minutes per wash in TBST at room temperature.

Then 500 µl of 1/1000 diluted conjugated secondary antibody was added. Again this occasionally needed optimisation. An incubation for 30 minutes at room temperature with shaking was then required. Again the membranes were washed four times for five minutes per wash with TBST at room temperature.

The membranes were developed using an ECL kit (GE Healthcare, Amersham, UK) and processed using automatic developer. A variety of exposure times are used to determine the best signal to noise ratio for the final photograph. Typically 15 seconds, 1 minute and 30 minutes were used.

Chapter 3

The Investigation of C5

Deficiency in a South African

Patient

3.1 Introduction

Complement component C5 deficiency is an infrequently observed complement terminal pathway deficiency, with less than 100 cases of C5 deficiency having been reported from ethnically diverse families (Lopez-Lera et al., 2009, Boelaert et al., 1985, Pfarr et al., 2005). The cases of C5 deficiency have been reported amongst Afro Americans, South Americans, South Africans and Moroccans (Wang et al., 1995). This chapter describes the molecular investigation of a South African Xhosan woman in her early thirties, who has had two attacks of meningitis caused by *Neisseria meningitidis*. A diagnosis of C5 deficiency was made by immunodiffusion assay and by demonstrating that no lytic activity was present in a haemolysis assay.

The patient has been prescribed monthly bicillin injections and has been observed for over ten years. She has suffered no further attacks since diagnosis.

Previous work in this laboratory together with Ms Claire Edwards (University of Surrey, M.Sc. 2006) focused on sequencing all 41 exons present in genomic DNA. Five sequence variations were observed, most significantly a heterozygous nonsense mutation in exon 1 c.84C>T which lead to premature termination (p.Q19X). This mutation has been reported previously in two Afro-American families (Wang et al, 1995). This is shown below in figure 15.
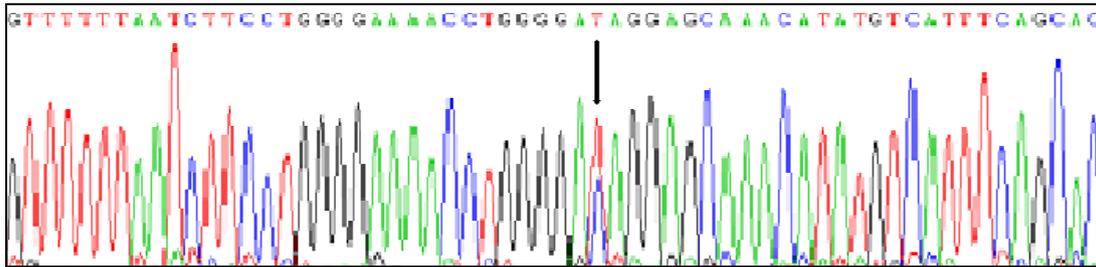
Figure 15. The published C5 exon 1 defect p.Q19X observed in the C5 deficient patient

The location of the defect on the first exon suggests that this is a null-allele and that no protein with any characteristics of C5 would be produced (Chang et al., 2007, Chan et al., 2007)

Another coding sequence variation was identified in exon 7 [c.754G>A; rs112959008], which causes a substitution of alanine at position 252 for a threonine [p.Ala252Thr]. This variation occurs within the alpha-2-macroglobin domain in the C5b section of the C5 molecule (DiScipio et al, 1998). The sequencing electropherogram is shown in figure 16 below.
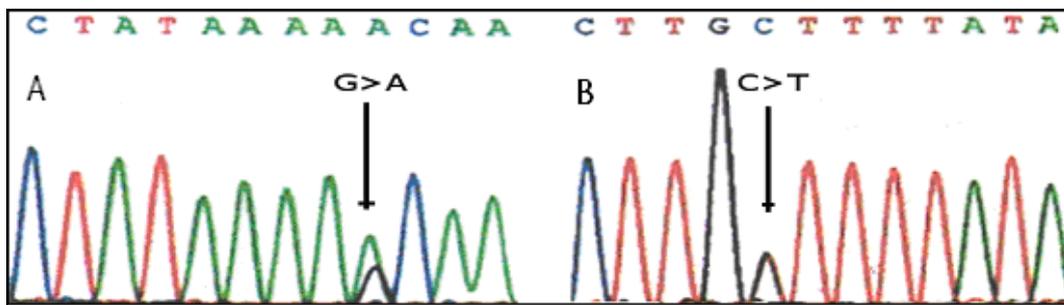
Figure 16. The c.754G>A sequence variation observed on C5 exon 7. The figure shows exon 7 sequenced (A) in the forward direction and (B) the reverse direction. The location of the variation is marked

In addition, two intronic sequence variations were observed in intron 25 and are shown in Figure 17. These were both close to exon 26 and are heterozygous T>G substitutions [IVS25-18T>G and IVS25-22T>G].

A further sequence variant was a homozygous substitution of G for a T at position – 12 within intron 32. [IVS32-12G>T] This is a known polymorphism

listed as rs10985112 and is also shown in figure 17 alongside the intron 25 variation.

TATTTATTT   TCT T GTT TTAT GT GCCA GCAT CCAC T
155   162   50   57   64   71

Intron 25   Intron 33
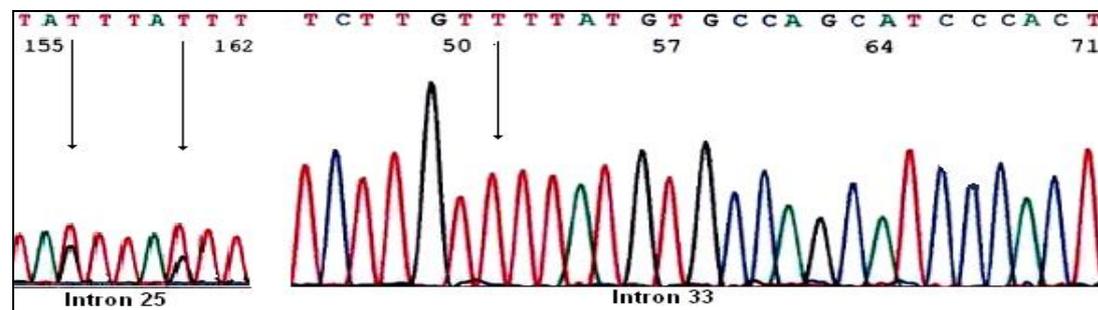
Figure 17. The C5 polymorphisms observed in introns 25 and 32 for the patient.

To summarise, five sequence variations were identified (table 10). Only 1 of these, the truncating mutation in exon 1 (p. Q19X) can be considered pathogenic (Wang et al., 1995). The other 4 are of unknown clinical significance. The missense mutation in exon 7 is a relatively conservative substitution that has been identified previously on a single allele from 1000 genomes originating from the Bantu population in Central and South Africa and has been deposited in the database (dbSNP XML build 135) as rs112959008. The two intronic sequence variants upstream of the canonical splice acceptor site of exons 26 have not been previously reported and are also of unknown clinical significance, as is the polymorphism rs10985112 in intron 32.The possibility that a pathogenic mutation had been missed and was located elsewhere in the C5 gene was therefore considered. This chapter includes the results of this sequencing analysis. The chapter also includes the description of bioinformatic analyses of the identified sequence variants. In addition efforts to optimise RT-PCR amplification of C5 from human cell types are presented.

Location	Position	Change	dbSNP Reference
Exon 1	c.84C>T	Q19X	rs121909587
Exon 7	c.754G>A	A252T	rs112959008
Intron 25	c.32311-18T>G	Unknown	None
Intron 25	c.32311-22T>G	Unknown	None
Intron 33	c.4163-12C>T	Unknown	rs10985112

Table 10. The sequence variations observed on the patient C5 gene

3.2 Materials and Methods

The primers used for PCR amplification and their respective annealing temperatures were designed with primer3 software and are shown below in table 11.

3.2.1 Bioinformatics

Four prediction software packages were used to analyse mutations, these programs can be classified either as splicing prediction programs or protein function programs.

RESCUE-ESE: This resource is web based and available at http://genes.mit.edu/burgelab/rescue-ese/ .

ESEFINDER: This is a web based resource at http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi (Cartegni et al., 2003)

PolyPhen-2: This resource is at http://genetics.bwh.harvard.edu/pph2/ and is used for predicting the possible effect of an amino acid change on the structure and function of a human protein.

SIFT: This is a web based resource available at http://sift.jcvi.org/ which allows the end user to check for changes in protein functionality using multiple alignment data to predict tolerated substitutions or potential defects. Putative promoter sequences were analysed analyses using MatInspector software at www.genomatix.de. CpG island motifs were identified using EMBOSS software available via the web server at http://www.ebi.ac.uk/Tools/emboss/cpgplot/index.html

3.2.2 Identifying secondary structure in C5 mRNA

The C5 mRNA sequence was examined using RNAfold software available at http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi for secondary structure by calculating lowest mean free energy. Data on centroid structure and partition functions was also obtained with this software.

3.2.3 The design of PCR primers for amplifying C5 mRNA

The mRNA sequence was examined for repetitive Alu, LINE and SINE elements using www.repeatmasker.org. GC content of C5 mRNA was analysed using Vector NTI software (Invitrogen, UK) for successive DNA segments of 10bp. Primers were initially designed with primer3 software available at http://primer3.sourceforge.net/, with GC content of 50% selected as optimum (Regions with a GC content of<35% or >65% were excluded). The proposed primers positions were adjusted manually to avoid SNPs and regions with high secondary structure complexity. Full length mRNA was

amplified in 7 overlapping fragments using the primers listed in table 11

below.

Primer	Sequence	%GC Content
C5 cDNA 1F	ATCCGTGGTTTCCTGCTACC	55
C5 cDNA 1R	TTTGCATTGCTGTTTGCATC	40
C5 cDNA 2F	AAAGTAGTCACTGAGGCTGACG	50
C5 cDNA 2R	AGACACAGTTTGGCCTGGAG	55
C5 cDNA 3F	TCACAGGAGAACAGACAGCAG	52
C5 cDNA 3R	CAGTGTTTGAAATGCCAACG	45
C5 cDNA 4F	TTTCCAGAAAAGCTGGTTGTG	43
C5 cDNA 4R	TAGCACTTCCACCCTTCCAC	55
C5 cDNA 5F	CTGGAAACAGGAAATCATTGG	43
C5 cDNA 5R	CCACCTCCATACCTCTGCTC	60
C5 cDNA 6F	ACACTGGTACGGCACGTATG	55
C5 cDNA 6R	CGGCTCCTTCACAGACTTTTC	52
C5 cDNA 7F	CAGTCCTGCCACTTTCACAG	55
C5 cDNA 7R	CAGCAAACATTCCTGAGTGG	50

Table 11. The primers designed for the amplification of C5 cDNA in 7 fragments

3.2.4 Optimising the amplification of C5 mRNA

The reverse transcription reactions were performed as described in materials

and methods section 2.5. The secondary structure reducing additives were

used at a volume of 5µL in a 20µLreaction and are as follows;

A.   dH2O – This was used as a control to maintain constant volume.

B.   DMSO (Fisher, UK) – A working concentration of 20% was used.

C.   Betain (Sigma Aldrich, UK) – Betain and Betain Monohydrate are

thought to reduce secondary structure,

D.   Formamide (Oxoid, UK)– A final concentration of 10% was used.

E.   Triton X100 (Fisher, UK) – A final concentration of 1% was used.

F.   BSA (Invitrogen, UK)–A final concentration of 20% was used.

G.   Invitrogen PCR Enhancer Solution (Invitrogen, UK) – This is a

commercially available solution containing a proprietary mix of components. A final concentration of 10% was used.

The fidelity of the reverse transcriptase in this modified reaction was assessed using the housekeeping gene Glyceraldehyde 3-phosphate dehydrogenase (GapDH). An additive was considered unacceptable if a PCR amplification of GapDH cDNA produced either non-specific PCR products or significantly reduced quantities of the GapDH amplicon.

3.2.5 The selection of appropriate cell types for C5 cDNA amplification

Buffy coat was prepared from a sample of 15mls of whole blood as described in materials and methods. The following cell types were then enriched and used for the preparation of mRNA and corresponding cDNA;

(1) Buffy Coat: This was used directly for mRNA production and served as a relative control. (2) EBV Immortalised B Cells: Buffy coat from a healthy volunteer was co-cultured with a sample of live EBV virus derived from B95-8 cells. Cyclisporin A was added to reduce T cell proliferation. The process of transformation took 4 to 6 weeks. Activated EBV transformed B cells were produced by co-culturing with L cells and an anti human CD154 antibody.

(3) Monocytes: These were prepared by placing buffy coat in 15mls of RPMI F10 media (Invitrogen, UK) in a 25cm^2 tissue culture flask overnight which resulted in monocytes becoming adherent. The remaining cells were aspirated and the monocytes then harvested.

(4) Neutrophils: These were isolated as described in materials and methods.

.

3.3 Results

3.3.1 Extended Genomic Sequencing

The 5' untranslated regions (UTRs) of genes may contain elements capable of regulating gene expression, particularly translation. However, the C5 gene 5'UTR is 31bp in length, which is a relatively short UTR and is not predicted to have a significant effect on gene expression. Sequencing the patient's genomic DNA across this region did not identify any variation.

3'UTRs of genes contains elements which may affect the mRNA stability. The C5 gene 3'UTR is 401bp (www.ensembl.org) and contains a number of known sequence variants (table 12) for which there is no information concerning their possible effect on C5 expression. Sequencing the patient's genomic DNA that included the 3'UTR did not identify any of these rare alleles nor find additional changes. These results confirm that no other mutations were present within the C5 transcript sequence.

Location (from stop codon)	SNP Identifier	Population Genetics
+78	rs147540583	No Data
+111	rs141560628	No Data
+301	rs148353209	No Data
+324	rs118181017	0.008 CHB / JPT
+327	rs144332635	No Data

Table 12: The location of sequence variations in the C5 3' flanking region, together with SNP reference number and available gene frequency data. (CHB) Chinese living in Beijing (JPT) Japanese living in Tokyo

Sequencing analysis was then extended to include the C5 promoter region. This region has only been partially characterised (Carney et al., 1991) Putative sequence elements including a TATA-like element and CCAAT box

were identified together with a number of transcription factor binding sites (Figure 18), However, the minimal promoter sequence was not defined through functional analysis and non of these elements have been shown to function in vivo.

Sequence upstream of C5 was re-analysed by *in silico* methods. Using MatInspector software at www.genomatix.de, a TATA-less core promoter was identified at position -379 to -429 (relative to the translation ATG start site). Although this seemed too far upstream from the reported transcriptional start site at -31 and was likely to be artifactual, it did provide the impetus to explore regions further upstream and in particular look for CpG islands. Vertebrate CpG islands (CGIs) are short interspersed DNA sequences that deviate significantly from the average genomic pattern by being GC-rich, CpG-rich, and predominantly nonmethylated. Most, perhaps all, CGIs are sites of transcription initiation, shared DNA sequence features adapt CGIs for promoter function by destabilizing nucleosomes and attracting proteins that create a transcriptionally permissive chromatin state. 2Kb of sequence upstream of C5 was therefore examined for CpG island motifs using EMBOSS software available via the web server at http://www.ebi.ac.uk/Tools/emboss/cpgplot/index.html. A CpG island motif was located at position -1725 to -1570, relative to the translation ATG start site. These features are shown in figure 18 below
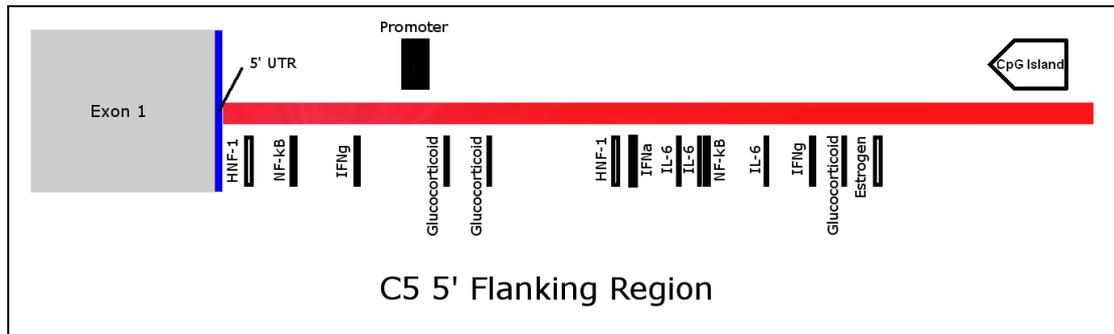
Figure 18. A diagram representing the relative positions of specified sequence elements involved in enhancing gene transcription.

In addition, a list of known sequence variations within this region and their respective positions were generated from www.ensembl.org data and shown below in Table 13.

Location (from start codon)	SNP Identifier	Population Genetics
-202	150172957	No Data
-227	138666345	No Data
-340	114156353	0.008 YRI
-419	41307964	0.008 YRI
-428	36223102	0.042 YRI
-705	148922946	No Data
-835	36223103	No Data
-930	36223100	Venter: USC
-942	5900469	No Data
-957	79187303	0.500 USC
-959	62578436	0.500 USC
-1126	57815396	No Data
-1416	138308342	No Data
-1553	36223437	0.058 CEU
-1705	78945889	No Data
-1884	36223436	0.067 CEU

Table 13: The location of sequence variations in the C5 5' flanking region, together with SNP reference number and available gene frequency data. (YRI) Yorubi living in Ibidan. (USC) US Caucasians and (CEU) Utah resident of central European ancestry

This extended region of genomic DNA was amplified by PCR and subjected to automated fluorescent sequencing. No additional sequence variants were detected and none of the published rare alleles were present. These results provided strong evidence that no additional mutations are present within the C5 gene. It was therefore decided to re-examine those mutations already identified by *in silico* methods.

3.3.2 Bioinformatics

The semi-conservative substitution in exon 7 (p.Ala252Thr) replaces a hydrophobic amino acid residue for a polar uncharged amino acid residue. A phylogenetic comparison between human C5, and other species indicated that Ala252 is highly conserved (figure 19). A phylogenetic comparison between human C5, and C5 from macaque, marmoset, giant panda, cow, pig, rat, dog, opossum, zebra finch, chicken and frog was made. The comparison between Human and Gorilla C5 does show a conservative change from Alanine to Leucine, although the gorilla data is derived from one individual. The predicted consequence of this change on protein function was investigated using different prediction software (PolyPhen2, SIFT). Using PolyPhen2, the p.Ala252Thr change was considered to be possibly damaging to protein function, with a score of 0.690 (on a scale of 0.000 -1.000). The phylogenetic comparison result below in figure 19 implies that position 252 (indicated by a black box) is highly conserved. An examination of the species compared revealed that the amino acid residue was conserved in the Japanese puffer fish (fugu rubripes) but not in the spotted green pufferfish

(Tetraodon nigroviridis). The spotted green pufferfish C5 protein has a valine

at the equivalent 252 position.

This amino acid change was also considered deleterious using SIFT, a

prediction software based on the degree of conservation of amino acid

residues in sequence alignments derived from closely related sequences.

Figure 19. A comparison of human C5 against those of other species listed above. The comparison is centred on position 252.

Each mutation was further evaluated for possible effects on C5 mRNA

splicing using RESCUE-ESE and ESE-FINDER software.

The mutated base c.754G is 5bp from the 3'end of exon 7 and so an effect on

splicing was considered a plausible explanation; perhaps through disrupting

an exonic splicing enhancer (ESE) sequence. ESEs are pre-mRNA *cis*-acting

elements required for splice-site recognition and have been shown to bind SR

(serine/arginine) proteins that are involved in regulating and selecting splice

sites in eukaryotic mRNA. Interestingly, ESEfinder predicted the loss of a

putative SR40 binding site in the mutated sequence, indicating that this

mutation may affect the correct splicing of exon 7, an illustration of splice site

for the normal and mutated sequence are shown below in figure 20.

Figure 20. A representation of splicing motifs and their relative strengths for normal exon 7
and mutated 754G>A sequence. The blue bar represents the position of the exon. The height
of each bar is directly proportional to the relative splicing strength. The position of each bar
indicates the nucleotide sequence of each splicing motif. The varying colour of each bar is for
illustration purposes only.

The three remaining intronic sequence variants were also analysed. The

c.3231-18T>G and c.3231-22T>G variants in intron 25 are within the

polypyrimidine region of the splice acceptor site and could potentially interfere

with correct splicing of exon 26 (Aguilar-Ramirez., 2009). Neither sequence

variant was predicted affect the strength of the cognate splice acceptor site of

exon 26, or create or abolish putative ESE motifs. However, both T>G

transversions were predicted to result in the creation of possible cryptic splice

acceptor sites (ApG), as shown in figure 21 below, upstream of exon 26. Recognition of these sites by the cells splicing machinery would, in theory, result in an in-frame insertion of six and 7 amino acids respectively.
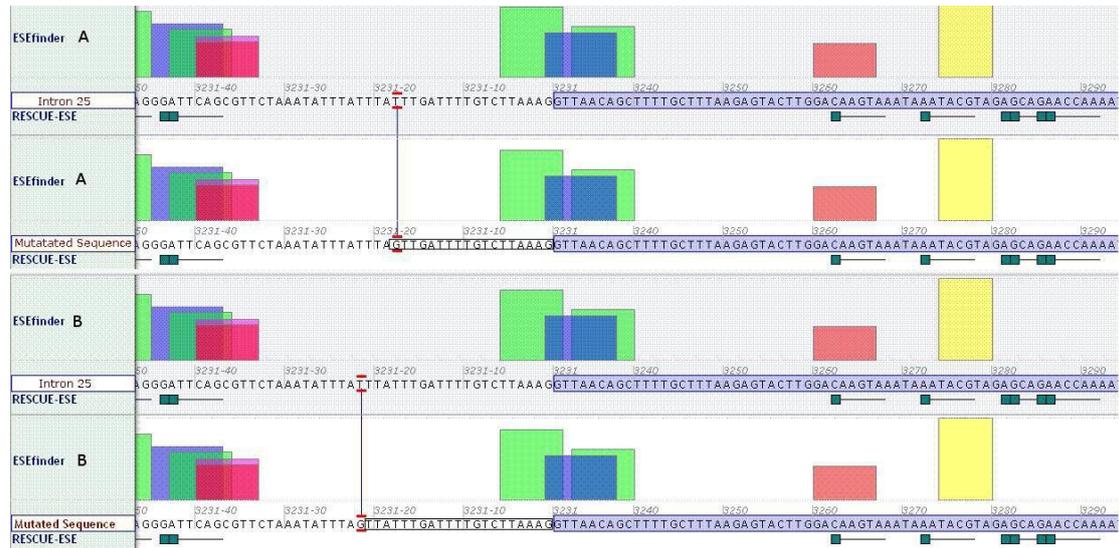
Figure 21. A representation of splicing motifs and their relative strengths for normal intron 25 and the two mutated intron 25 sequences. The -18 sequence variation is shown in the upper panels A and the -22 sequence variation is shown in the lower panels B. The blue bars represents the position of the exon. The blue bar represents the position of the exon, with the nucleotide sequence leading to the six and seven amino acid inclusions shown in the boxed area preceding the exon. The height of each bar is directly proportional to the relative splicing strength. The position of each bar indicates the nucleotide sequence of each splicing motif. The varying colour of each bar is for illustration purposes only.

The c.4163-12C>T polymorphism in intron 32 also lies within the polypyrimidine tract of the splice acceptor site. The replacement of one pyrimidine for another would not be expected to have an effect on splicing of exon 32 and *in silico* analyses also predicted a negligible reduction in the strength of a putative ESE, and no effect on the strength of the splice acceptor site as shown in figure 22 below.
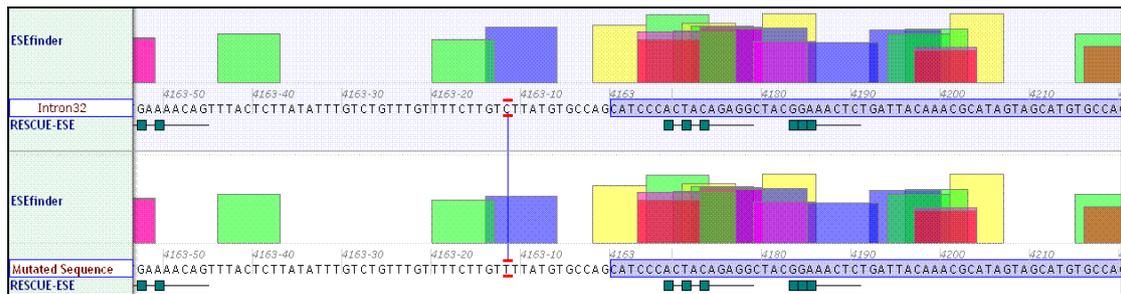
Figure 22. A representation of splicing motifs and their relative strengths for normal intron 32 sequence and mutated c.4163-12C>T sequence. The blue bar represents the position of the exon. The height of each bar is directly proportional to the relative splicing strength. The position of each bar indicates the nucleotide sequence of each splicing motif. The varying colour of each bar is for illustration purposes only.

3.3.3 C5 mRNA secondary structure

The theoretical effect of the c.754G>A change on the secondary structure of the mature C5 mRNA was also investigated. C5 mRNA secondary structure was analysed using RNAfold. A comparison of the mean free energy algorithm for both C5 Normal mRNA and the C5 A252T mRNA is depicted in figure 23 and shows a high degree of secondary structure within the C5 mRNA molecule and subtle alterations due to the c.754G>A change. However, when the algorithm was expanded to include a centroid function, which calculates the minimum possible distance between branched structures in accordance with the Boltzmann equation, the change to predicted mRNA structure was pronounced, as shown in figure 24 below. Whether these alterations in the secondary structure due to a single base substitution reflect the true nature of mRNA in vivo is open to question. The possibility that an increase in secondary structure may interfere with gene expression, for instance, by interfering with ribosome progression during translation remains a theoretical possibility
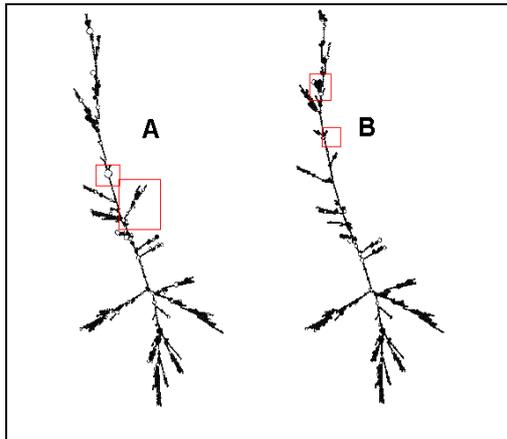
Figure 23. A comparison between (A) the reference C5 mRNA sequence and (B) the C5 mRNA exon 7 variation sequence. Selected changes to secondary structure are highlighted within the red boxes.
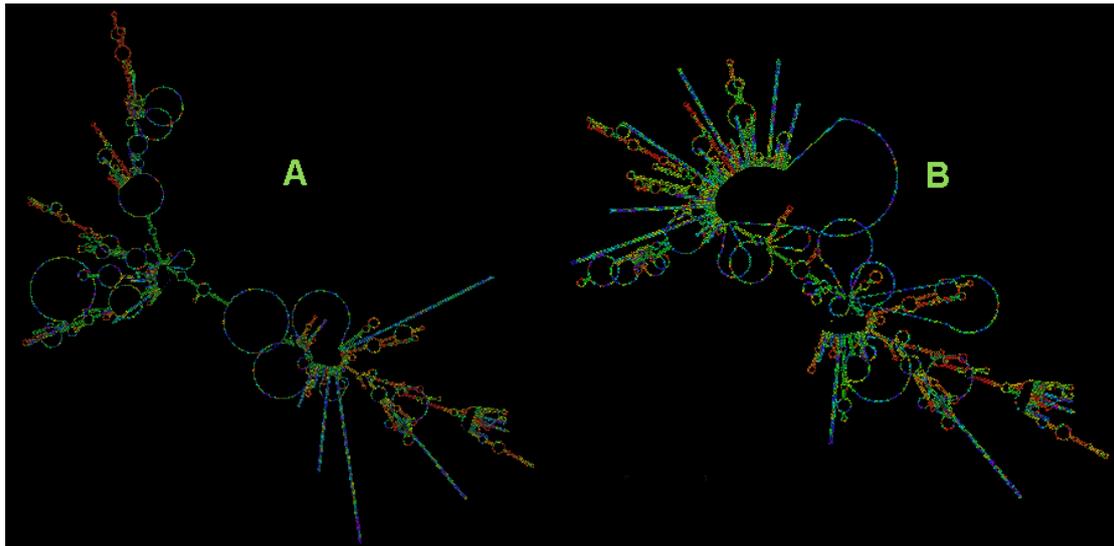
Figure 24. The secondary structure predicted with a centroid function for (A) the reference C5 mRNA sequence and (B) the C5 C84T variant sequence.

3.3.4 The optimised RT-PCR amplification of C5 mRNA

In silico analyses indicate possible effects of the sequencing variants in exon 7 and intron 25. A direct study of these sequence variants on C5 mRNA stability and/or splicing would require efficient amplification by RT-PCR

A method of amplifying C5 cDNA has been described by Delgado-Cervino (2005), and utilises a series of overlapping PCR reactions. This method was initially used to investigate the mutations in exons 1 and 7. We and others (personal communication) met with difficulties amplifying the 5' end of the C5 cDNA, encoded by the first 10 exons, and a second round of PCR using nested primers was required, however, this often led to the amplification of aberrantly spliced transcripts. This may be due in part to the length of C5 mRNA, which is approximately 5kb long. However, as described above, it is also possible that a highly complex secondary structure may impede the procession of reverse-transcriptase and or polymerase during RT-PCR. Therefore, a new protocol for C5 was developed. Initially the reduction of

70

secondary structure was attempted using a reverse transcriptase with a high reaction temperature. The enzyme Thermoscript RT (Invitrogen, UK) was reported to work up to 65°C. However, this proved unsuccessful.

A new set of primers were therefore designed. Firstly, using sequencing utility software RepeatMasker, human C5 cDNA was shown not to be harbour any repetitive sequence elements. Primers were selected based on their GC content and were located in regions of low secondary structure complexity and did not overlap with known SNPS. Secondly a series of different additives, including those known to relieve secondary structure were tested. Initially, these were tested on a control GapDH PCR amplification to check for enzyme fidelity changes. The GapDH PCR products are shown on an agarose gel in figure 25 below. Addition of DMSO, Betain, formamide or Triton X100 resulted in reduced yield and non-specific amplification, whereas BSA and the PCR_x enhancer produced a specific product with comparable yields to the water control. Since Invitrogen PCR_x additive is known to relieve difficult regions of secondary structure, this additive was used in subsequent amplifications of C5.
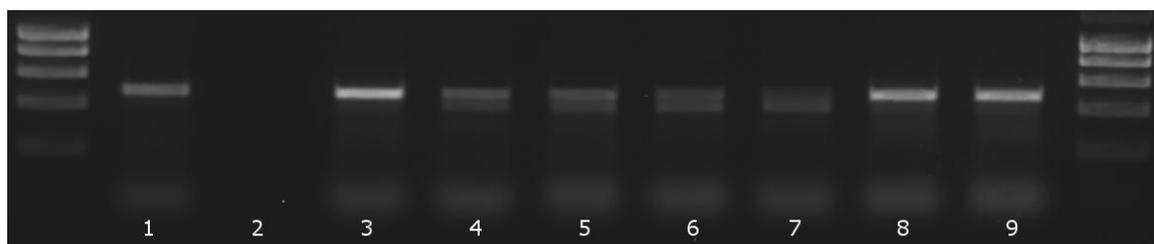
Figure 25. The amplification of the GapDH control sequence using cDNA templates with reverse transcription steps modified with (1) dH_2O (2) NTC (3) No Additives (4) DMSO (5) Betain (6) Formamide (7) Triton X100 (8) BSA (9) Invitrogen PCR_x enhancer

These amplification conditions were then used to amplify C5 using 35 PCR cycles from HepG2 cells, which are a known to express high levels of C5.

Full length C5 cDNA was successfully amplified by RT-PCR amplified in 7 overlapping reactions) as shown in the figure 26 below, confirming the utility of this method. This amplification method was then applied to cDNA obtained from different cell-types.
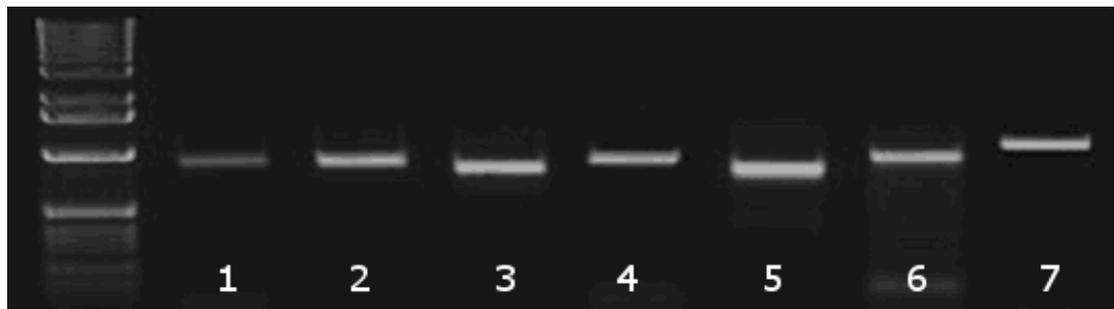
Figure 26. The amplification of C5 cDNA from a HepG2 cDNA template reverse transcribed in the presence of Invitrogen PCR$_x$ enhancer solution. Each PCR product represents a segment of C5 cDNA amplified by the primers in Table 11, and have a corresponding numbering shown in the figure.

3.3.5 Appropriate source material for the amplification of C5 mRNA

Two PCR reactions were used for a semi-quantitative comparison of C5 expression between different cell preparations. These reactions corresponded to fragment 1 and fragment 5 of C5 cDNA as described above. Fragment 5 was selected as it was robust and suitably distal from fragment 1 which was included to provide an assessment of efficient cDNA synthesis. Amplification of C5 cDNA from monocytes and EBV immortalised B cells was poor and discounted as monocytes and EBV immortalised B cells both appeared to produce C5 mRNA at comparatively low levels as shown in a representative figure 27 below
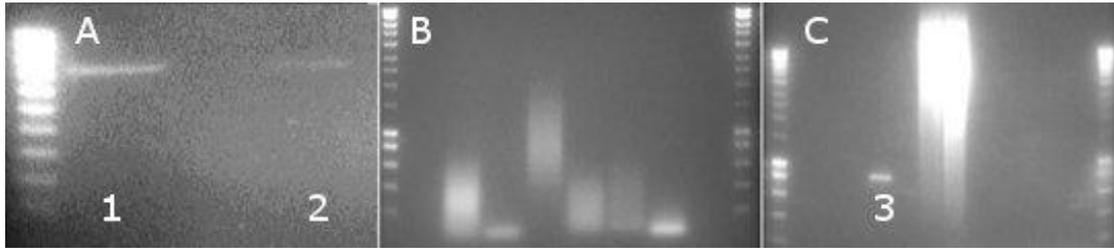
Figure 27. A representative amplification of C5 cDNA by (A) monocytes, (B) EBV immortalised B cells and (C) activated EBV immortalised B cells. Faint bands were occasionally observed for the monocyte template and are shown in Lanes 1 and 2. Normal EBV immortalised B cells did not give a distinct PCR product as shown in gel B. Activated B cells rarely gave a distinct PCR product, an example of which is seen in Lane 3.

The PCR amplification of C5 cDNA from buffy coat and Neutrophil derived cDNA templates gave a more abundant product as shown in figure 28 below. Comparison of the relative intensity of these fragments indicated that full length cDNA was efficiently synthesised in from both cell preparations suggesting that full length cDNA had been reverse transcribed efficiently for each template.
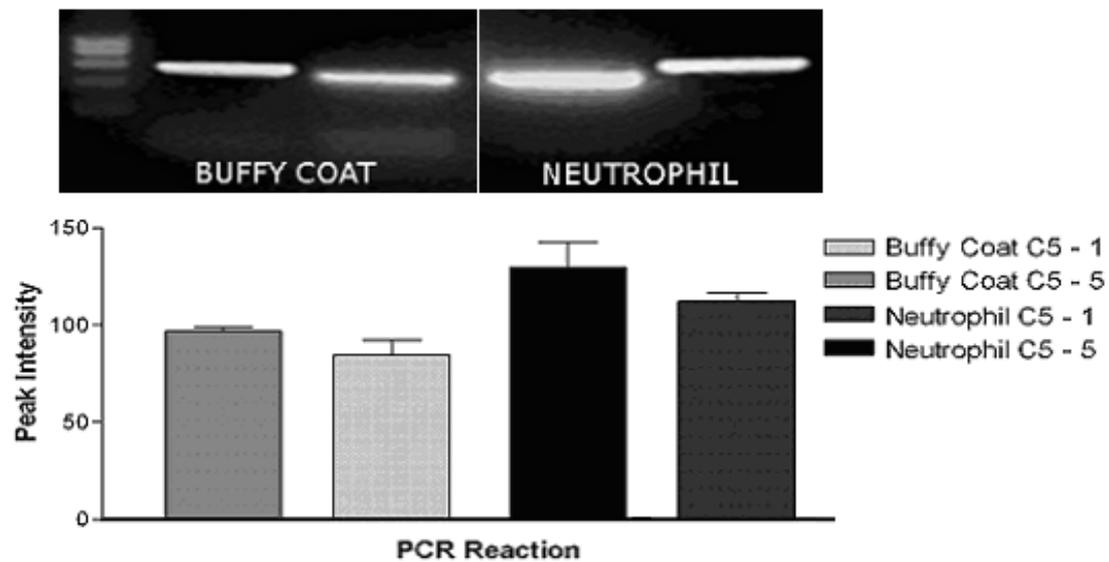
Figure 28. The upper panel shows representative amplification of selected C5 cDNA fragments for both buffy coat (left) and Neutrophil (right) derived cDNA templates
The lower panel shows a graphical representation of Densitometry data for each of the PCR amplifications for buffy coat and Neutrophil derived template.

3.4 Discussion

The genomic sequencing of all 41 exons and promoter region constitutes less than 10% of the total C5 gene, and therefore a causative mutation perhaps embedded deep within an intron may have been missed. However, five sequence variations were identified in this South African woman with C5 deficiency. One of these is a published nonsense mutation in exon 1. Unfortunately for the remaining variations, particularly the intronic mutations, no suitable tissue material was available from the patient in order to directly study potential effects on C5 mRNA processing and therefore the molecular mechanisms that result in C5 deficiency in this patient remain speculative. The strongest candidate is c.754G>A in exon 7. Genomic sequencing of the patient's offspring demonstrated that the exon 1 and exon 7 mutations segregate independently, thus showing they exist on separate alleles in patient genomic DNA. The sequence variation in exon 7 has been identified in one additional allele from 1000 genomes originating from the same population as the patient described here. Whilst it remains possible that this is a rare SNP of no clinical significance, it may also prove to be a relatively common cause of C5 deficiency in this population.

It may exert its effect on C5 expression via a number of possible mechanisms, it is unlikely that stable but inactive C5 with an Ala252Thr substitution is produced as this would still be detectable via a polyclonal antibody based immunoblot, and no such protein is observed (data not shown). It may inhibit protein secretion from the cell, but this too may be unlikely as C5 would be detectable from cells that have undergone lysis.

Undetectable levels of C5 could be caused by protein instability, or mRNA instability. The latter may result from mis-splicing involving exon 7 (Graveley et al., 1998, Graveley et al., 2008). Furthermore, the profound effect that c.754G>A may have on mRNA folding is intriguing and outcome of this mutation suggests another possibility whereby translation inhibition and therefore mRNA stability is a possible.

In order to gain an insight into the nature of the A252T C5 mutation and the mechanisms leading to deficiency it may be possible to clone functional C5 and ligate it into an expression vector to allow exogenous expression. Site directed mutagenesis may then allow the creation of a plasmid containing A252T mutation. Protein and mRNA studies on cells containing either the normal C5 or A252T C5 plasmid may yield important data regarding the molecular mechanism of deficiency for this mutation.

Chapter 4

The characterisation of a large

C7 genomic defect present in the

Irish population

4.1 Introduction

A study of known genetic defects responsible for C7 deficiency revealed that the genetic defects responsible are very heterogeneous (Thomas et al., 2012).

To date there are over 25 different C7 gene sequence variations that have been deposited in the Human Gene Mutation Database [HGMD] http://www.hgmd.cf.ac.uk/ac/index.php. Mostly, these are restricted to individual families, although occasionally the same C7 defect is found in several unrelated families living within a particular community such as the pathogenic G357R change found in Moroccan Jews living in Israel (Fernie et al., 1997b).

A gene deletion defect was suspected following the regional failure to PCR amplify exons 8 and 9 (shown here in figure 29, and supported by Southern Blot analyses).

Figure 29. The regional failure of PCR for exons 8 and 9 and using a control DNA sample and a heterozygous sample (shown in the first two lanes for each exon) and two deletion defect positive subjects.

Direct demonstration of the deletion in heterozygotes is prevented by the amplification of exons 8 and 9 from the normal allele (Figure 29 above).

The work described in this chapter includes investigations to fully characterise the breakpoints of the deletion. Using this information a method to screen for

it in homozygotes and heterozygotes was developed. A possible mechanism of deletion is also discussed.

4.2 Methods

4.2.1 Patients

The subjects reported in this study were members of four unrelated Southern Irish families; within each family the proband was diagnosed because of meningococcal infections. There is no evidence that any of the families are related. Families 1 and 2 have been described previously (Egan et al., 1994; Fernie et al., 1997b); Family 3 has been reported as Family A (O'Hara et al., 1998). Fresh blood samples were obtained from two members of Family 3 for this study. A fourth family that has not been reported previously and includes a single C7D patient, parents and two C7 sufficient siblings was also investigated in this study.

4.2.2 PCR walking and amplification across the deletion boundaries

The conditions for PCR amplifications are as described in Materials and Methods. The primers used are listed in Table 14

Primer Name	Sequence
Intron 7 Region 1F	CAATATCATGTCACTGCACTTCAGC
Intron 7 Region 1R	GCCTCTAGCATTTGCATTTAAGAAG

Intron 7 Region 2F	CAAAGCCCTTATTACCTTGAAATCTTC
Intron 7 Region 2R	TCAGATTTCTCAAGAGGAAGGTTTTAG
Intron 7 Region 3F	GAGGCTTTTTAGATGGTGCATGA
Intron 7 Region 3R	GCTCCATAGTCAGTTTCTCAGTCATAAA
Intron 9 Region 1F	GGAAGCATCAGGTGATTTGTTAAAA
Intron 9 Region 1R	TCTCTTAACCTCATGATCCACCCA
Intron 9 Region 2F	AGAGATTGCTTAATGGGTATAAACCTGT
Intron 9 Region 3F	ATCATGAGGTTAAGAGATCGAGACCA
Intron 9 Region 2+3R	CCTGACTGTGCTATCAAACACTAGAATT

Table 14. The primer sequences used to restrict the potential size of the C7 gene deletion by gene walking

The locations of the gene walking PCR fragments are shown in figure 30, and following the restriction of the size of the potential deletion defect using this method, and attempt was then made to amplify across the defect using the sense and antisense primers in table 9 that most closely bracket the deletion defect.
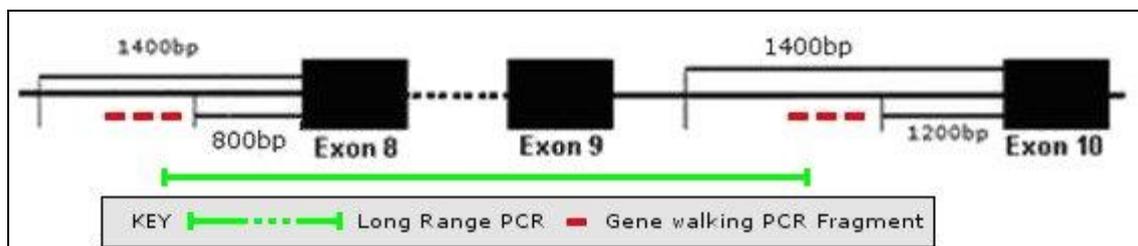
Figure 30. The locations of PCR gene walking fragments and the region across the deletion defect for which amplification is desired.

4.2.3 The routine identification of the large C7 genomic deletion

The deletion was screened using a duplex PCR method using the primers listed below. The cycling conditions were an initial denaturing step of 95°C for

2 minutes then 35 cycles of 95°C for 30 seconds, 55°C for 30 seconds and 72°C for 30 seconds.

Primer Name	Sequence
Deletion Defect F	TCTCCACATTATATTGTTCACAAAGTGG
Deletion Defect R	CGTCTTCATTGTTAGCTGTGTTTTCA
Normal Allele F	CTTCATTATAAAGTAGCTTCACCTGCAA
Normal Allele R	TCTGAGGGAACTGCAGATTTCTTAG

Table 15. The primers used for the routine detection of the C7 deletion defect

4.2.4 Gene Nomenclature

For this study, C7 exons are numbered 1-18. Historically, exons were numbered 0-17. Nucleotides are numbered from the initiating methionine of the Coding Reference sequence for human C7 (NM_000587). This is consistent with the guidelines of the HGVS (www.hgvs.org/rec.html).

4.3 Results

Our initial attempts at long range PCR to amplify across the deletion using primers in exons 7 and 10 were unsuccessful. We therefore decided to use PCR-walking to localize the breakpoints. DNA from the homozygous C7D patient of Family 1 was analysed using the primer pairs (listed in Table 10) that were progressively proximal to the deleted exons. Positive amplification indicated that the sequence was normal while failure to amplify indicated deleted sequence. The results are shown in Figure 31 below.

Figure 31. The successful amplification of each of the PCR walking amplification for primers listed in table 9. The first lane (left to right) is a DNA ladder. The land 2 to 4 are the 3 intron 7 fragments and the lanes 5 to 7 are the intron 9 fragments.

Using this strategy, it was determined that the deletion could be no longer than 7317bp, based on the consensus C7 gene sequence obtained from the Sanger Institute. These results were consistent with the previous estimations of O'Hara et al who concluded that the 5' boundary was 800-1400bp 5' of exon 8 and the 3' boundary 1200-1800bp 5' of exon 10 (O'Hara et al., 1998). Primers: intron 7 region 3F and intron 9 region 2+3R were then used to amplify across the boundary as shown in figure 32 below.
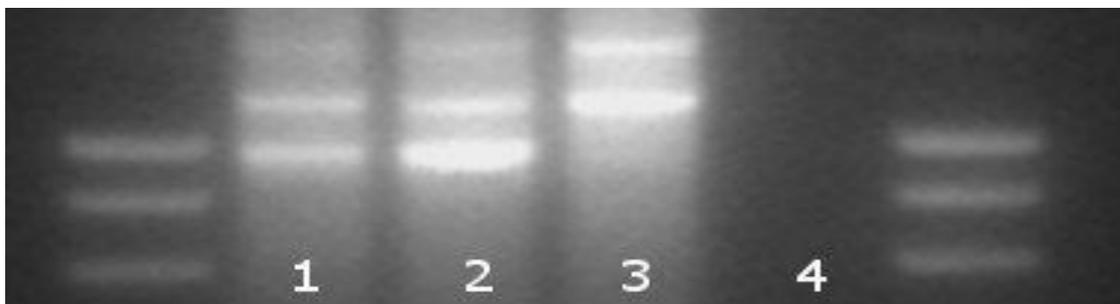
Figure 32. An amplification across the deletion boundary for (1) a heterozygote, (2) A deletion defect homozygote, (3) a normal control and (4) a no template control.

Although extensive non-specific amplification occurred, a 429bp product was visible in a homozygous C7D individual and a heterozygote but not control DNA. DNA sequencing is shown in figure 33 below.

81
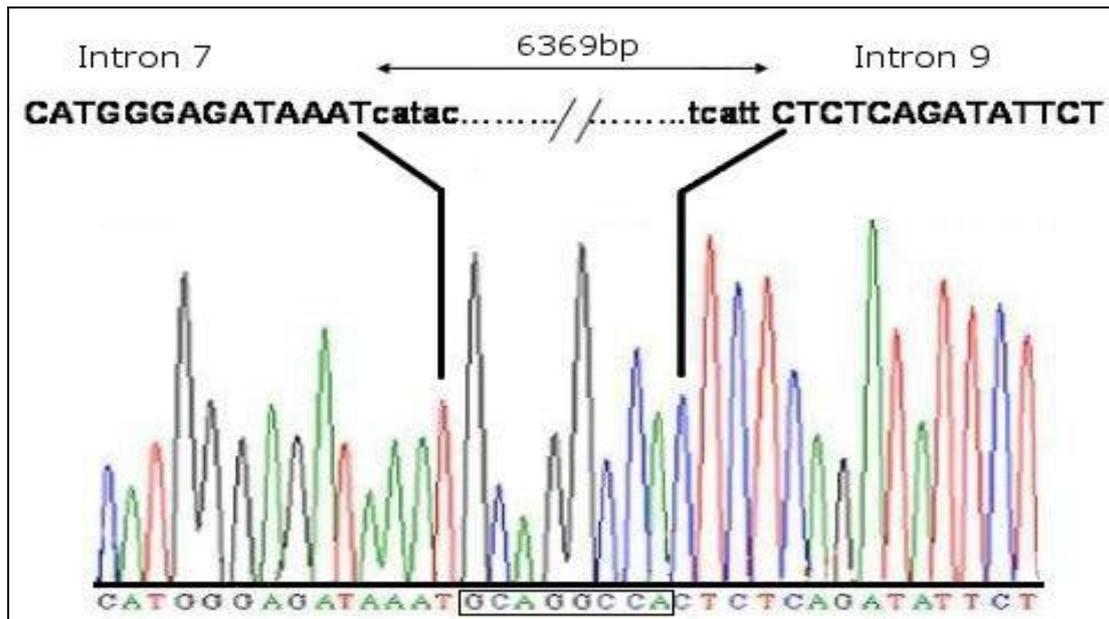
Figure 33. A sequencing figure describing the identity of the deletion and novel 8bp inclusion. This sequence was consistent between each of the four families examined.

Alignment with normal C7 demonstrated that the deletion was a complex rearrangement consisting of a 6369bp deletion and, interestingly, the insertion of a novel 8bp sequence. [c.738+1262_1270-2387delinsGCAGGCCA].

4.4 Screening for the deletion using a duplex PCR

A duplex PCR for the simultaneous amplification of normal sequence and across the deleted sequence was established using additional primer pairs listed in table 10. One set of primers [Deletion Defect F and Deletion Defect R] were used to amplify across the deletion and result in a 302bp product when amplifying across the deletion. Amplification of normal sequence using these primers cannot proceed using specified cycling conditions, because the product is too long. The other set of primers [Normal Allele F and Normal Allele R], located within exon 8 and therefore within the deleted sequence.

These primers amplify a 257bp product when a normal template of the C7 gene is present.

Using this method we analysed the proband and relatives in our families. The result of screening in Family 2 is shown in Figure 35. An investigation of ten haplotype markers on both the C6 and C7 genes on each of the family members implied that the deletion defect was consistent with a particular haplotype (Fernie et al., 1997a). This haplotype indicated that the deletion defect status for each family member, from left to right in figure 34 was two heterozygotes, three deletion homozygotes, three heterozygotes and then two normal individuals. This matches the result shown by routine screening duplex PCR.

Figure 34. A duplex PCR for the routine detection of the C7 deletion defect for family 2

A fourth unrelated family where the deletion defect was suspected was also investigated. The proband was diagnosed as C7D post mortem. The results are shown in Figure 35 below, and indicates that not only was the proband homozygous for the deletion defect, but several heterozygotes are present within family 4.

Figure 35. A duplex PCR for the routine detection of the C7 deletion defect for family 4. The figure (left to right) shows a DNA ladder, two heterozygotes, a deficient patient, a heterozygote and an unaffected individual.

4.5 Discussion

The sequencing of PCR fragments amplifying across the DNA deletions in the index cases in Families 1, 2 and 3 revealed identical breakpoints and 8bp insertion. In addition, investigations of the affected member of Family 4 also confirmed homozygous deletion. Complete characterisation of the DNA sequence of the deletion enabled us to develop a simple duplex PCR-based assay to test for the presence of both one non-affected allele and one affected allele (Thomas et al., 2012). We also analysed the index case from Family 1 who is heterozygous for the deletion defect and a splicing defect in intron 1 (c.7-1G>A). Both deletion-positive and deletion-negative DNA fragments were amplified, thus confirming the patient is compound heterozygote.

An investigation into the origin of the C7 deletion defect was conducted using *in silico* methods. This work was performed in collaboration with Dr Nadia Chuzhanova in Department of Medical Genetics Cardiff University. Dr Chuzhanova is credited with generating the final hypothesis for the origin of the deletion defect.

Initially, the sequences flanking both breakpoints were screened for the presence of direct repeats, inverted repeats and symmetric elements listed in Appendix IV by means of complexity analysis (Gusev et al., 1999).

The program RepeatMasker was used to determine whether any of the repetitive sequence elements identified by means of complexity analysis were either Alu or LINE elements. In addition, ±10bp flanking each breakpoint were screened for the presence of 37 DNA sequence motifs (or their complements) of length ≥5bp known to be associated with site-specific cleavage/recombination, high frequency mutation and gene rearrangement (Abeysinghe et al. 2003) as well as various 'super-hotspot motifs' found in the vicinity of micro-deletions/-insertions and indels (Ball et al., 2005). These DNA sequence datasets were also screened for the presence of 15 additional recombination-associated motifs (plus their complements) identified by Cullen et al. (2002) plus a CCTCCCT motif associated with a classical meiotic recombination hotspot (Myers et al. 2005; Myers et al. 2006). The motifs examined are listed in Appendix IV.

Inspection of the C7 genomic sequence identified five Alu sequences (AluJo, AluSx, AluSg, AluSc and AluSx) in the vicinity of the deletion/insertion breakpoints. The orientations of these highly homologous sequences (1-5) are denoted by arrows in Figure 36a. We postulated that two slipped structures, formed between homologous regions 1 and 2 and between regions 4 and 5 respectively, could have contributed to the initial DNA breakage events (Figure 36b). However we are unable to explain either the exact positions of the deletion breakpoints or the precise mechanism of the insertion by any one simple mutational model.

Deletion of the DNA fragment may have been mediated by two short direct repeats of the form CATW. This implies that non-homologous end joining, a pathway that repairs double-strand breaks in DNA, may have been the mechanism responsible for this mutation (Seluanov et al., 2004).

Interestingly, the complement of the deletion hotspot consensus sequence, TGRRKM, found in close proximity to the second breakpoint (Figure 36b) may have played a role in defining the exact location of this breakpoint. A 6bp copy of a GCAGGC fragment resembling the inserted bases was found within repeat 4 in the same orientation whilst a 7bp fragment, TGGCCTG (complementary to CAGGCCA), was identified at position 441. Although the underlying mutational mechanism is unclear, AATCATAC direct repeats, spanning the first breakpoint, were found immediately upstream of this complementary fragment.
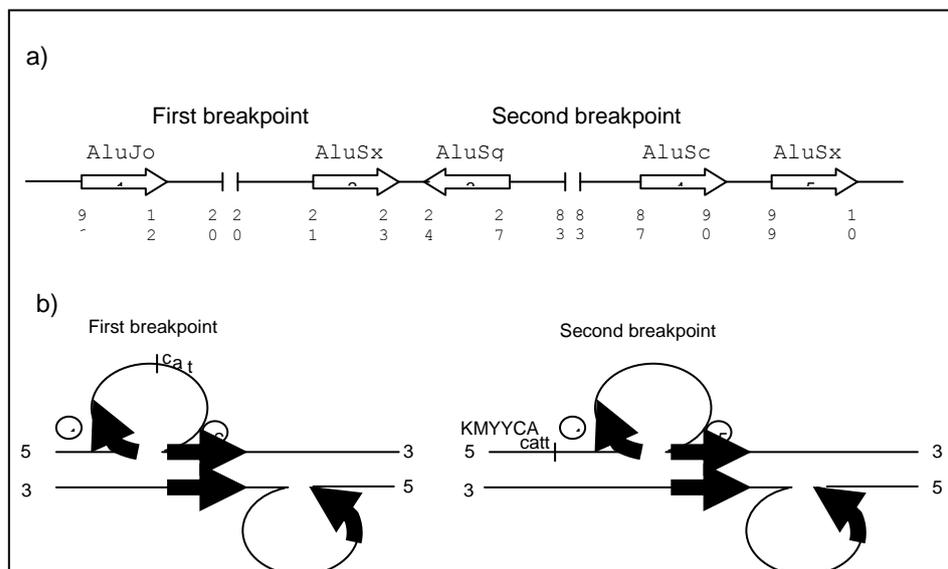
Figure 36. A potential mechanism for the origin of the C7 deletion defect. The relative positions and orientations of both the Alu and CATW sequence are shown.

Neisseria meningitis infections are very serious, and one of the C7D patients reported here died of meningitis. The disease also has important sequelae and the more infections individuals suffer, the more likely are they to suffer long-term complications such as deafness, limb-loss from disseminated intra-vascular coagulation, or mental impairment (Orren et al., 2006). The deletion defect appears to be a relatively common cause of C7 deficiency in Ireland and genetic screening is needed to determine the frequency in this population.

Chapter 5

The identification of the

molecular defects in three C8

deficient patients

5.1 Introduction

The molecular defects in three UK resident patients, each of whom had presented with recurring episodes of meningitis were determined. The patients were referred for molecular screening based on their clinical history and tests which demonstrated the complete absence of haemolytic activity. The identity of the missing component was tested in a modified CH50 assay supplemented to physiological levels with C5, C6, C7, C8, or C9, haemolytic activity in the patients' serum could be restored to levels comparable to normal serum upon addition of pure C8. (data not shown).

C8 deficiency is less common than deficiency of C6 or C7 and may result from lack of the α–γ chain or the β chain. Deficiency of C8α-γ has been predominantly reported amongst Afro-Caribbean people, Hispanic people and Japanese (Würzner et al., 1992). C8β deficiency is more common, with the majority of cases occurring amongst Caucasian people although C8B deficiency has been reported in other areas such as South America (Rosa et al., 2004)

The C8B gene is 38 kb long and contains 12 exons. A single common mutation, c.1309C > T, that introduces a premature stop codon (R428X) in exon 9, is the causative mutation in approximately 85% of null alleles examined in 41 cases of C8β deficiency [Saucedo et al., 1995 and Kaufmann et al., 1993]. The remaining 15% are other C>T transitions or single base deletions, all of which result in premature termination of translation. These mutations are restricted to 4 of the 12 exons (3, 5, 6, and 9).

The work described in this chapter includes investigations to identify which C8 protein was deficient and the molecular investigations to identify the sequence variation at the genomic level.

5.2 Methods

5.2.1 Patients

Three UK patients, all male, presented with recurring neisserial meningitis infections and during the course of this study consented for whole blood to be taken for molecular screening. None of the patients were related and all came from different ethnic backgrounds, with one being UK British and two being UK Asian. They were aged between 20 and 35 and all experienced their initial infection during early childhood.

Patient 1: A UK British Male

Patient 2: A UK Asian Male

Patient 3: A UK Asian Male

.

5.3 Confirmation of Missing Component by Western Blotting

Serum samples (30µl, following a 1:10 dilution) were analyzed by non-denaturing SDS-PAGE and Western Blotting as described in Experimental Methods using a pAb rabbit anti-human C8 (Dako, Ely, Cambridgeshire) at a 1:2000 dilution as a primary antibody. Protein was detected by incubation with an αRbIgG-HRPO conjugated secondary antibody (1/1,000 dilution; Jackson ImmunoResearch Laboratories, Philadelphia) for 30 min and the membranes developed using ECL reagent.

A selective exon amplification approach to mutation analysis of C8β deficiency was used. Due to the frequency and position of known mutations a strategy was implemented in which exons 3; 5; 6; and 9 were amplified and sequenced in an initial screen of the C8B gene in these patients.

DNA was isolated from patient EDTA-blood as described in Experimental Methods.

The genomic DNA isolated from the patient blood samples was assessed for quality and quantity using UV spectrometry. Each sample was diluted to 100ng/µl and used in a PCR to amplify the C8B gene. Each patient was screened initially by amplification and sequencing of only those exons in which mutations had previously been identified i.e. in exons 3, 5, 6, and 9. The PCR products were excised from the agarose gel and purified using a geneclean III kit (MP Biomedical, France) and the purified products used in a sequencing reaction in both the sense and antisense orientations as described in Experimental Methods. The sequencing electropherograms were analysed using BIOEDIT software (citation in program) to assess data quality.

Sequencing of the remaining exons was performed if no mutations were identified by this strategy.

5.4 Gene Nomenclature

The C8B gene is annotated according to current recommendations (Human Gene Nomenclature Committee). Nucleotides are numbered using the coding reference sequence NM_000066 in which the A of the ATG translation initiating codon is numbered +1

5.5 Results

5.5.1 Confirmation of Missing Component by Western Blotting

To determine which subunit of C8 was absent, C8 protein was analyzed by non-denaturing SDS-PAGE followed by western blotting using an anti-human C8 polyclonal antibody (Figure 37). In normal human serum, C8 is visualized as two bands (lane 1). These are the covalently linked α–γ chains (87 kDa) and the β chain (64 kDa). Figure 37 shows the result of analysing serum of patient 1. There was a clear and specific loss of the β chain (lane 2) in the patient's serum, confirming deficiency of the C8β subunit in this patient. Analysis of the patient's serum and comparison with normal human serum demonstrated that C8 α–γ chains were present but reduced in amount, as previously reported in C8β deficiency (Pallares et al., 1996). The analysis of

the remaining patients showed identical results. They were performed by

other members of laboratory staff and are not shown.
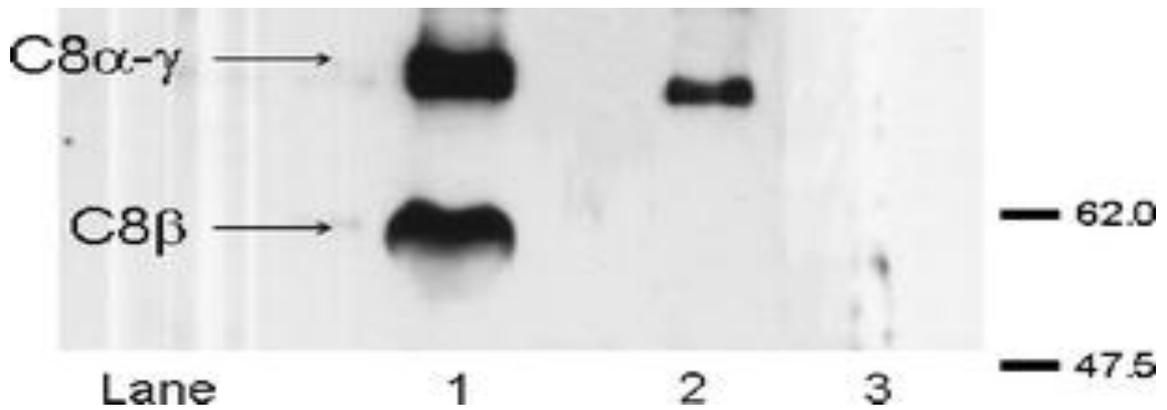
Figure 37. The western blotting of C8β using a polyclonal antibody for both a deficient patient
(Lane 2) and normal human serum (Lane1)

5.5.2 Genomic Investigations

Each of the exons (3, 5, 6 and 9) were successfully amplified and were the

correct size from each patient. A representative amplification is shown below

in figure 38.
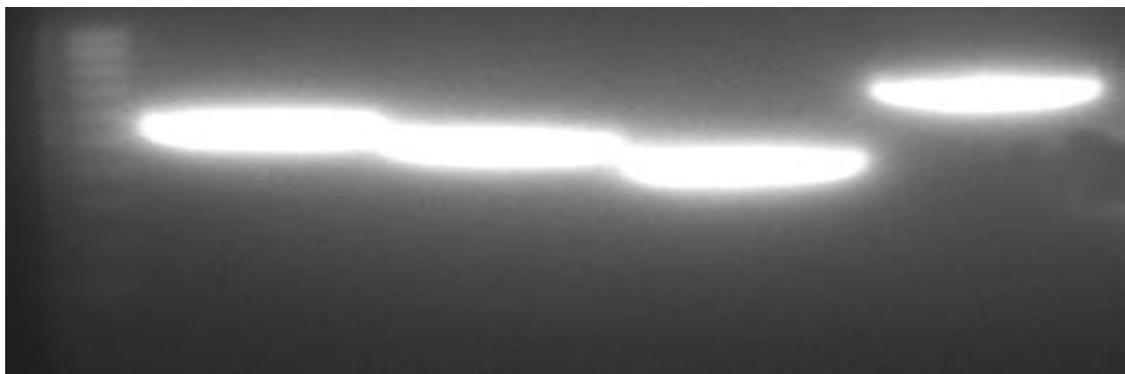
Figure 38. The amplification of C8B exons 3, 5, 6 and 9 and their flanking intronic sequences.
The PCR appeared robust during all amplifications.

5.5.3 Patient 1

Sequencing results and analysis are shown in figure 39. Sequencing of the patient's C8B gene identified a heterozygous mutation c.271C>T within exon 3. This transition causes a nonsense change and introduces a premature stop codon at glutamine 91 (p.G91X). No other mutation was identified in this initial screen. Sequencing of the remaining exons identified a novel mutation caused by a duplication c.1047_1053dup GGCTGTG, in exon 7.
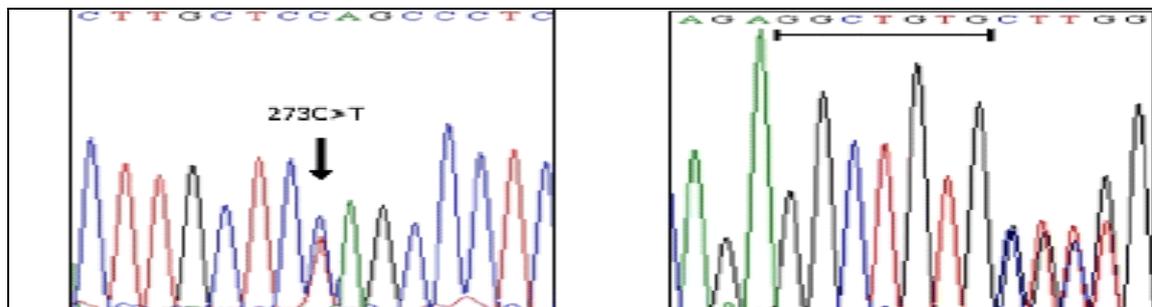
Figure 39. Sequencing results for patient 1 showing heterozygosity for both the exon 3 mutation and the novel c.1047_1053 dupGGCTGTG mutation

Sequencing of parental DNA showed that the patient's father carried the novel c.1047_1053 dupGGCTGTG mutation, whereas his mother carried the c.C271 > T (Q91X) mutation in exon 3, thus demonstrating that these mutations are found on separate chromosomes in the patient confirming compound heterozygosity (data not shown).

5.5.4 Patient 2 and 3

Sequencing results and analysis are shown in Figure 40. Sequencing of the patients' C8B gene demonstrated that they were both homozygous for the common mutation, c.1309C > T, that introduces a premature stop codon (R428X) in exon 9.
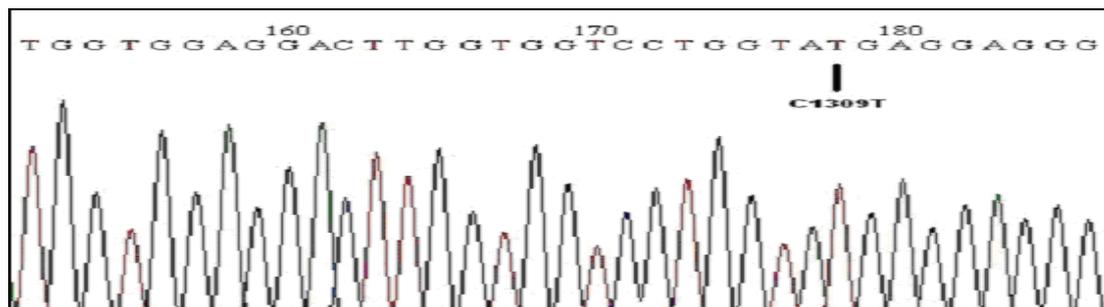
Figure 40. The sequencing results for patient 1 C8B exon 9 in which homozygosity for the c.1309C>T defect is shown. Identical results for patient 2 were also observed

5.5.5 Discussion

C8β deficiency is by far the more frequent cause of C8 deficiency in Caucasians (Arnold DF et al., 2009) and is commonly caused by a nonsense mutation in exon 9 (R428X) together with rarer mutations in exons 3, 5 and 6. A strategy of selected exonic amplification to screen the C8B gene for causative mutations in C8β deficiency was assessed in 3 patients with C8β deficiency. In Patients 2 and 3, sequencing demonstrated that both were homozygous for the common c.1309C>T transition supporting the finding that it is responsible for most reported cases of C8β deficiency. However Patient 1 presented differently. This patient was shown to be a compound heterozygote of an uncommon, but previously described nonsense mutation, c.271C > T in exon 3, (Q91X), (Saucedo et al., 1995) and a novel duplication, c.1047_1053

95

dupGGCTGTG in exon 7. It is likely that this tandem duplication, in common with the majority of short insertion mutations, is derived from a slippage-like process during DNA replication (Messer and Arndt 2007). It introduces a frame shift that results in seven novel amino acid codons (G-C-A-Y-G-H-K) immediately downstream of the amino acid L350, followed by a premature stop codon. Although either mutation results in the theoretical truncation of the C8β protein., no truncated C8 protein was detected by western blotting of this patient's serum, suggesting that the proteins are either unstable or the transcripts containing these mutations are unstable and degraded via the nonsense mediated decay pathway.

These findings confirm that the c.1309C>T mutation is the commonest genetic cause for C8β deficiency and that a selected screen for it as a primary analysis is appropriate. However, these results also suggest that C8β deficiency is more heterogeneous than previously thought. The c.271C>T, (p.G91X) mutation has been reported previously as 298C>T in two other families (Saucedo et al., 1995) and its presence in a British patient indicates that it may be more common than previously considered. Furthermore the absence of the c.1309C>T mutation in this patient, indicates that sequencing of the entire coding region of the C8B gene may be required in some instances.

6 Concluding Remarks

Within a clinical research context, an effective screening protocol of common, causative mutations should be considered for each gene. This approach would make the most efficient use of resources prior to screening the remaining exons, UTR's and promoter regions. However, as more human genomes are sequenced, the number and type of genetic defect are expected to increase. Likewise, this is true for complement terminal pathway genes. This is of particular concern when new sequence variants are identified where the effect of the mutation is not obvious, such as apparently conservative amino acid substitutions. Bioinformatics and segregation analyses within families can provide supporting evidence, however, experimental verification of a causative mutation is often necessary.

C5

In cases of C5 deficiency a selected exon approach is appropriate since it is a relatively large gene with 41 exons. Thus, exons 1, 7, 10, 15, 20, 35, 36 and 40 should be screened for known defects. The functional effect of the C5 exon 7 sequence variant remains unclear.

It may be possible this variation interferes with normal C5 exon splicing, and verifying this would require minigene experimental data.

It may be that the exon 7 defect interferes with the structure of the mRNA molecule, either causing the molecule to become unstable, aggregate or impede translation. It may be possible to perform northern blot analysis to examine mRNA size and, through probe hybridisation ability, mRNA structure.

Additionally, protein targeting may also be an issue. Defects of protein target to the secretion pathway have already been reported in the complement terminal pathway (Rameix-Welti et al., 2007).

Finally, it may be that a molecule is secreted by the gene which lacks C5 functionality thereby appearing to be deficient. The expression of both C5 and the C5 A252T defect recombinant protein *in vitro* may indicate any inhibition of protein production or secretion.

C7

An initial selected exon sequencing method for investigating C7 deficiency may not be appropriate. Although some founder effects are known, in which defects segregate to specific populations (Fernie et al., 1997b), in general the number and geographical distributions of C7 deficiency are too great to allow such an approach. Therefore, it is recommended that initially a whole gene exon sequencing approach be used.

The deletion defect reported here was previously proposed following a routine failure of PCR amplification and a failure of southern blot probe hybridisation. Neither of these methods can be used for clinical diagnosis using DNA techniques or for carrier detection, and the duplex PCR method presented here meets these requirements. It may be interesting in future studies to determine the frequency with which the defect occurs, in particular, the Irish population.

The deletion defect presented here appears to be Alu mediated. Slipped mis-pairing of Alu sequences during DNA replication is a commonly known mechanism of gene rearrangement (Ball et al., 2005). The hypothesis discussed here is more complicated, and may illuminate the way in which a

failure of the cell's normal DNA repair mechanism can be responsible for a protein deficiency.

C8B

All defects identified in the C8B gene result in either premature stop codons or frame shifts. We have confirmed that the 1289C>T mutation is a common cause of deficiency, and in cases of C8β deficiency it may be advisable to screen for this defect alone in an initial selected exon screen.

There is an intrinsic value in sequence analysis. Taking C5 as an example, there are 9 synonomous and 12 non-synonomous sequence variants within the coding region of the gene that do not cause C5 deficiency. In addition there are over 320 polymorphisms located within the introns. The advent of high through-put single genome sequencing will provide greater insight into the mechanisms of protein production and the interaction of protein variants with other terminal pathway components and complement regulators, thereby illuminating variation in the functional capacity of the complement cascade.

Finally, in order to progress from a patient sample to the identification of a molecular defect a rigorous process must be employed. Initially DNA or RNA technologies may not be appropriate. Regardless of the specific molecular defect it is the absence of functional protein that produces a clinical pathology. As a first step a haemolytic assay or |ELISA may be employed to identify the complement component which is deficient. Further protein techniques such as western blotting may resolve whether non functional protein or no protein is being produced.

It is at this stage that a PCR based investigation of known genomic defects may be pursued. The may then require expanding PCR screening to include all exons, regions flanking the genes or intronic sequence.

It may also be prudent, particularly for larger genes such as C5, to examine mRNA as the absence of intronic sequence and potential quantification of mature mRNA may more rapidly identify novel defects.

References

Abeysinghe, S. S., N. Chuzhanova, et al. (2003). "Translocation and gross deletion breakpoints in human inherited disease and cancer I: Nucleotide composition and recombination-associated motifs." Hum Mutat **22**(3): 229-44.

Agostoni, A. and M. Cicardi (1992). "Hereditary and acquired C1-inhibitor deficiency: biological and clinical characteristics in 235 patients." Medicine (Baltimore) **71**(4): 206-15

Aguilar-Ramirez, P., E. S. Reis, et al. (2009). "Skipping of exon 30 in C5 gene results in complete human C5 deficiency and demonstrates the importance of C5d and CUB domains for stability." Mol Immunol **46**(10): 2116-23.

Altschul, S; Gish, W; Miller, W; Myers, E; Lipman, D (October 1990). "Basic local alignment search tool". Journal of Molecular Biology 215 (3): 403–410.

Amara, U., M. A. Flierl, et al. "Molecular intercommunication between the complement and coagulation systems." J Immunol **185**(9): 5628-36.

Arnold, D. F., A. G. Roberts, et al. (2009). "A novel mutation in a patient with a deficiency of the eighth component of complement associated with recurrent meningococcal meningitis." J Clin Immunol **29**(5): 691-5.

Ball, E. V., P. D. Stenson, et al. (2005). "Microdeletions and microinsertions causing human genetic disease: common mechanisms of mutagenesis and the role of local DNA sequence complexity." Hum Mutat **26**(3): 205-13.

Basta, M., P. Kirshbom, et al. (1989). "Mechanism of therapeutic effect of high-dose intravenous immunoglobulin. Attenuation of acute, complement-dependent immune damage in a guinea pig model." J Clin Invest **84**(6): 1974-81.

Basta, M., F. Van Goor, et al. (2003). "F(ab)'2-mediated neutralization of C3a and C5a anaphylatoxins: a novel effector function of immunoglobulins." Nat Med **9**(4): 431-8.

Bindon CI, Hale G, et al. (1988). "Human monoclonal IgG isotypes differ in complement activating function at the level of C4 as well as C1q." J Exp Med. 1988 Jul 1;168(1):127-42.

Blencowe, B. J. (2000). "Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases." Trends Biochem Sci **25**(3): 106-10.

Blom, A. M., T. Hallstrom, et al. (2009). "Complement evasion strategies of pathogens-acquisition of inhibitors and beyond." Mol Immunol **46**(14): 2808-17.

Blyth, M., M. Raponi, et al. "Expanding the tuberous sclerosis phenotype: mild disease caused by a TSC1 splicing mutation." J Neurol Neurosurg Psychiatry **81**(3): 350-2.

Bockow, B. and M. Mannik (1981). "Clearance and tissue uptake of immune complexes in complement-depleted and control mice." Immunology **42**(4): 497-504.

Boelaert, J., R. Joos, et al. (1985). "C5 deficiency in a white family." Arch Intern Med **145**(7): 1333.

Botto, M., D. Lissandrini, et al. (1992). "Biosynthesis and secretion of complement component (C3) by activated human polymorphonuclear leukocytes." J Immunol **149**(4): 1348-55.

Breslauer KJ, Frank R, Blöcker H, Marky LA. Predicting DNA duplex stability from the base sequence. Proc Natl Acad Sci U S A. 1986 Jun;83(11):3746-50.

Campbell, R. D., S. K. Law, et al. (1988). "Structure, organization, and regulation of the complement genes." Annu Rev Immunol **6**: 161-95.

Carney, D. F., D. L. Haviland, et al. (1991). "Structural aspects of the human C5 gene. Intron/exon organization, 5'-flanking region features, and characterization of two truncated cDNA clones." J Biol Chem **266**(28): 18786-91.

Cartegni, L., S. L. Chew, et al. (2002). "Listening to silence and understanding nonsense: exonic mutations that affect splicing." Nat Rev Genet **3**(4): 285-98.

Cartegni, L., J. Wang, et al. (2003). "ESEfinder: A web resource to identify exonic splicing enhancers." Nucleic Acids Res **31**(13): 3568-71.

Chan, W. K., L. Huang, et al. (2007). "An alternative branch of the nonsense-mediated decay pathway." Embo J **26**(7): 1820-30.

Chang, Y. F., J. S. Imam, et al. (2007). "The nonsense-mediated decay RNA surveillance pathway." Annu Rev Biochem **76**: 51-74.

Chen, J. M., D. N. Cooper, et al. (2007). "Gene conversion: mechanisms, evolution and human disease." Nat Rev Genet **8**(10): 762-75.

Chou Q, Russell M, Birch DE, Raymond J, Bloch W. Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications. Nucleic Acids Res. 1992 Apr 11;20(7):1717-23.

Cole, D.S. and Morgan, B.P. (2003). "Beyond lysis: how complement influences cell fate." Clin Sci (Lond). 2003 May;104(5):455-66.

Colten, H. R. and F. S. Rosen (1992). "Complement deficiencies." Annu Rev Immunol **10**: 809-34.

Cordaux M and Batzer M (2009). "The impact of retrotransposons on human genome evolution". *Nature Reviews Genetics* **10** (10): 691–703

Cullen, M., S. P. Perfetto, et al. (2002). "High-resolution patterns of meiotic recombination across the human major histocompatibility complex." <u>Am J Hum Genet</u> **71**(4): 759-76.

Dieffenbach CW and Dveksler GS. PCR primer : a laboratory manual. Cold Spring Harbor Laboratory Press, 1995

Delgado-Cervino, E., G. Fontan, et al. (2005). "C5 complement deficiency in a Spanish family. Molecular characterization of the double mutation responsible for the defect." <u>Mol Immunol</u> **42**(1): 105-11.

DiScipio, R. G. (1993). "The size, shape and stability of complement component C9." <u>Mol Immunol</u> **30**(12): 1097-106.

DiScipio, R. G., D. N. Chakravarti, et al. (1988). "The structure of human complement component C7 and the C5b-7 complex." <u>J Biol Chem</u> **263**(1): 549-60.

Discipio, R. G., L. Jenner, et al. (1998). "Crystallization of human complement component C5." <u>Acta Crystallogr D Biol Crystallogr</u> **54**(Pt 4): 643-6.

Egan, L. J., A. Orren, et al. (1994). "Hereditary deficiency of the seventh component of complement and recurrent meningococcal infection: investigations of an Irish family using a novel haemolytic screening assay for complement activity and C7 M/N allotyping." Epidemiol Infect **113**(2): 275-81.

Ember JA, MA Jagels, and TE Hugli. 1998 'Characterisation of complement anaphylatoxins and their biological responses'. In The human complement system in health and disease. Marcel Dekker Inc., New York, pp. 241 284.

Epstein, J., Q. Eichbaum, et al. (1996). "The collectins in innate immunity." Curr Opin Immunol **8**(1): 29-35.

Fernie, B. A., A. Orren, et al. (1997a). "DNA haplotypes of the complement C6 and C7 genes associated with deficiencies of the seventh component; and a new DNA polymorphism in C7 exon 13." Ann Hum Genet **61**(Pt 4): 287-98.

Fernie, B. A., A. Orren, et al. (1997b). "Molecular bases of C7 deficiency: three different defects." J Immunol **159**(2): 1019-26.

Fernie, B. A., R. Würzner, et al. (1996b). "Molecular bases of combined subtotal deficiencies of C6 and C7: their effects in combination with other C6 and C7 deficiencies." J Immunol **157**(8): 3648-57.

Figueroa, J. E. and P. Densen (1991). "Infectious diseases associated with complement deficiencies." Clin Microbiol Rev **4**(3): 359-95.

Fijen, C. A., E. J. Kuijper, et al. (1989). "2 families with meningococcal infection and a hereditary disorder of the 5th component of the complement system." Ned Tijdschr Geneeskd **133**(36): 1796-800.

Fosbrink, M., F. Niculescu, et al. (2005). "The role of c5b-9 terminal complement complex in activation of the cell cycle and transcription." Immunol Res **31**(1): 37-46.

Garred, P., F. Larsen, et al. (2003). "Mannose-binding lectin deficiency--revisited." Mol Immunol **40**(2-4): 73-84.

Gerard, C. and N. P. Gerard (1994). "C5A anaphylatoxin and its seven transmembrane-segment receptor." Annu Rev Immunol **12**: 775-808.

Graveley, B. R. (2008). "The haplo-spliceo-transcriptome: common variations in alternative splicing in the human population." Trends Genet **24**(1): 5-7.

Graveley, B. R., K. J. Hertel, et al. (1998). "A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers." Embo J **17**(22): 6747-56.

Gros, P., F. J. Milder, et al. (2008). "Complement driven by conformational changes." Nat Rev Immunol **8**(1): 48-58.

Haeney, M. R. (1998). "The role of the complement cascade in sepsis." J Antimicrob Chemother **41 Suppl A**: 41-6.

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 41:95-98.

Hobart, M. J., B. A. Fernie, et al. (1995). "Structure of the human C7 gene and comparison with the C6, C8A, C8B, and C9 genes." J Immunol **154**(10): 5188-94.

Hogasen, A. K., R. Würzner, et al. (1995). "Human polymorphonuclear leukocytes store large amounts of terminal complement components C7 and C6, which may be released on stimulation." J Immunol **154**(9): 4734-40.

Houts GE, Miyagi M, Ellis C, Beard D, Beard J W Reverse transcriptase from avian myeloblastosis virus. J Virol. 1979 February; 29(2): 517–522.

Hugli, T. E. and H. J. Muller-Eberhard (1978). "Anaphylatoxins: C3a and C5a." Adv Immunol **26**: 1-53.

Innis MA, Gelfand DH, Sninsky JJ. PCR applications : protocols for functional genomics. London : Academic Press, c1999

Kaufmann T, Hansch G, et al (1993). Genetic basis of human complement C8 beta deficiency. J Immunol. ;150:4943–7.

Kaufman, K. M. and J. M. Sodetz (1994). "Genomic structure of the human complement protein C8 gamma: homology to the lipocalin gene family." Biochemistry **33**(17): 5162-6.

Ki, C. S., J. W. Kim, et al. (2005). "Two novel mutations in the C7 gene in a Korean patient with complement C7 deficiency." J Korean Med Sci **20**(2): 220-4.

Koch-Brandt, C. and C. Morgans (1996). "Clusterin: a role in cell survival in the face of apoptosis?" Prog Mol Subcell Biol **16**: 130-49.

Kohler, P. F. and H. J. Muller-Eberhard (1967). "Immunochemical quantitation of the third, fourth and fifth components of human complement: concentrations in the serum of healthy adults." J Immunol **99**(6): 1211-6.

Krishnan, V., Y. Xu, et al. (2009). "The structure of C2b, a fragment of complement component C2 produced during C3 convertase formation." Acta Crystallogr D Biol Crystallogr **65**(Pt 3): 266-74.

Laarman, A., F. Milder, et al. "Complement inhibition by gram-positive pathogens: molecular mechanisms and therapeutic implications." J Mol Med **88**(2): 115-20.

Lachmann, P. J. and R. A. Thompson (1970). "Reactive lysis: the complement-mediated lysis of unsensitized cells. II. The characterization of activated reactor as C56 and the participation of C8 and C9." J Exp Med **131**(4): 643-57.

Lambris, J. D. (1998). The Human Complement System in Health and disease.

Law, S. K. and A. W. Dodds (1997). "The internal thioester and the covalent binding properties of the complement proteins C3 and C4." Protein Sci **6**(2): 263-74.

Law, S. K., A. W. Dodds, et al. (1984). "A comparison of the properties of two classes, C4A and C4B, of the human complement component C4." Embo J **3**(8): 1819-23.

Lipman NS., Jackson LR., et al (2005) "Monoclonal versus polyclonal antibodies: distinguishing characteristics, applications, and information resources." ILAR J. **2005**;46(3):258-68. Review.

Lopez-Lera, A., S. Garrido, et al. (2009). "Molecular characterization of three new mutations causing C5 deficiency in two non-related families." Mol Immunol **46**(11-12): 2340-7.

Manoharan, A. and M. P. Madaio "Biomarkers in lupus nephritis." Rheum Dis Clin North Am **36**(1): 131-43, ix. 2010

Matsushita, M. (1996). "The lectin pathway of the complement system." Microbiol Immunol **40**(12): 887-93.

Matsushita, M., Y. Endo, et al. (2001). "Activation of the lectin complement pathway by ficolins." Int Immunopharmacol **1**(3): 359-63.

Merkens LS, Bryan SK, Moses RE. Inactivation of the 5'-3' exonuclease of Thermus aquaticus DNA polymerase. Biochim Biophys Acta. 1995 Nov 7;1264(2):243-8.

Messer PW, Arndt PF (2007). The majority of recent short DNA insertions in the human genome are tandem duplications. Mol Biol Evol.;24:1190–7. doi:10.1093/molbev/msm035.

Michelotti, G. A., J. V. Snider, et al. (1995). "Genomic organization of human complement protein C8 alpha and further examination of its linkage to C8 beta." Hum Genet **95**(5): 513-8.

Migita, S., M. Nonaka, et al. (1990). "Biosynthesis and significance of acute phase proteins." Rinsho Byori **38**(5): 543-7.

S A Miller, D D Dykes, and H F Polesky (1988). "A simple salting out procedure for extracting DNA from human nucleated cells". Nucleic Acids Res. 1988 February 11; 16(3): 1215.

Morgan, B. P. and P. Gasque (1997). "Extrahepatic complement biosynthesis: where, when and why?" Clin Exp Immunol **107**(1): 1-7.

Morgan, B. P., K. J. Marchbank, et al. (2005). "Complement: central to innate immunity and bridging to adaptive responses." Immunol Lett **97**(2): 171-9.

Morgan, B. P. and A. Orren (1998). "Vaccination against meningococcus in complement-deficient individuals." Clin Exp Immunol **114**(3): 327-9.

Morgan, B. P. and M. J. Walport (1991). "Complement deficiency and disease." Immunol Today **12**(9): 301-6.

Myers, S., L. Bottolo, et al. (2005). "A fine-scale map of recombination rates and hotspots across the human genome." Science **310**(5746): 321-4.

Myers, S. R. and S. A. McCarroll (2006). "New insights into the biological basis of genomic disorders." Nat Genet **38**(12): 1363-4.

Noonan KE, Roninson IB. 'mRNA phenotyping by enzymatic amplification of randomly primed cDNA'. Nucleic Acids Res. 1988 Nov 11;16(21):10366.

O'Hara, A. M., B. A. Fernie, et al. (1998). "C7 deficiency in an Irish family: a deletion defect which is predominant in the Irish." Clin Exp Immunol **114**(3): 355-61.

Ooi, Y. M. and H. R. Colten (1979). "Biosynthesis and post-synthetic modification of a precursor (pro-C5) of the fifth component of mouse complement (C5)." J Immunol **123**(6): 2494-8.

Orren, A., D. A. Caugant, et al. (1994). "Characterization of strains of Neisseria meningitidis recovered from complement-sufficient and complement-deficient patients in the Western Cape Province, South Africa." J Clin Microbiol **32**(9): 2185-91.

Pallares DE, Figueroa JE, et al (1996) Invasive Haemophilus influenzae type b infection in a child with familial deficiency of the beta subunit of the eighth component of complement. J Pediatr.;128:102–3. doi:10.1016/S0022-3476 (96)70436-5.

Pangburn, M. K. (1983). "Activation of complement via the alternative pathway." Fed Proc **42**(1): 139-43.

Pfarr, N., D. Prawitt, et al. (2005). "Linking C5 deficiency to an exonic splicing enhancer mutation." J Immunol **174**(7): 4172-7.

Platonov, A. E., I. V. Vershinina, et al. (2003). "Antibody-dependent killing of meningococci by human neutrophils in serum of late complement component-deficient patients." Int Arch Allergy Immunol **130**(4): 314-21.

Podack, E. R. (1984a). "Molecular composition of the tubular structure of the membrane attack complex of complement." J Biol Chem **259**(13): 8641-7.

Podack, E. R. and J. Tschopp (1984b). "Membrane attack by complement." Mol Immunol **21**(7): 589-603.

Purwar, R., T. Werfel, et al. (2006). "IL-13-stimulated human keratinocytes preferentially attract CD4+CCR4+ T cells: possible role in atopic dermatitis." J Invest Dermatol **126**(5): 1043-51.

Rameix-Welti, M. A., C. H. Regnier, et al. (2007). "Hereditary complement C7 deficiency in nine families: subtotal C7 deficiency revisited." Eur J Immunol **37**(5): 1377-85.

Renart J, Reiser J, Stark GR. (1979). "Transfer of proteins from gels to diazobenzyloxymethyl-paper and detection with antisera: a method for studying antibody specificity and antigen structure". Proceedings of the National Academy of Sciences USA 76 (7): 3116–20.

Rosa, D. D., A. C. Pasqualotto, et al. (2004). "Deficiency of the eighth component of complement associated with recurrent meningococcal meningitis--case report and literature review." Braz J Infect Dis **8**(4): 328-30.

Rychlik, W.; Spencer, W. J.; Rhoads, R. E. (1990). "Optimization of the annealing temperature for DNA amplification in vitro.". N*ucleic Acids Res.* **18** (21): 6409–6412.

Sambrook J and Russell D (2001). *Molecular Cloning: A Laboratory Manual.* New York: Cold Spring Harbor Laboratory Press

Sanger, F.; Nicklen, S.; Coulson, A.R. (1977), "DNA sequencing with chain-terminating inhibitors", *Proceedings of the National Academy of Sciences USA* **74** (12): 5463–5467

Saucedo L, Ackermann L, et al (1995). Delineation of additional genetic bases for C8 beta deficiency. Prevalence of null alleles and predominance of C>T transition in their genesis. J Immunol.;155:5022–8.

Scheele G, Blackburn P. Role of mammalian RNase inhibitor in cell-free protein synthesis. Proc Natl Acad Sci U S A. 1979 October; 76(10): 4898–4902.

Seluanov A, Mittelman D, et al. (2004) "DNA end joining becomes less efficient and more error-prone during cellular senescence". Proc Natl Acad Sci U S A. 2004 May 18;101(20):7624-9.

Shapiro AL, Viñuela E, Maizel JV Jr. (September 1967). "Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels". Biochem Biophys Res Commun. 28 (5): 815–820

Sim RB, Reid KB. (1991) "C1: molecular interactions with activating systems." Immunol Today. 1991 Sep;12(9):307-11.

Takabayashi, T., S. Shimizu, et al. (2004). "Interleukin-1 upregulates anaphylatoxin receptors on mononuclear cells." Surgery **135**(5): 544-54.

Thomas AD, Orren A, et al (2011). "Characterization of a large genomic deletion in four Irish families with C7 deficiency." Mol Immunol. 2012 ;50(1-2):57-9.

Veres G, Gibbs RA, Scherer SE, Caskey CT. The molecular basis of the sparse fur mouse mutation. Science. 1987 Jul 24;237(4813):415-7.

Wang, X., D. T. Fleischer, et al. (1995). "Inherited human complement C5 deficiency. Nonsense mutations in exons 1 (Gln1 to Stop) and 36 (Arg1458 to Stop) and compound heterozygosity in three African-American families." J Immunol **154**(10): 5464-71.

Wetsel, R. A., R. S. Lemons, et al. (1988). "Molecular analysis of human complement component C5: localization of the structural gene to chromosome 9." Biochemistry **27**(5): 1474-82.

Würzner, R., A. E. Platonov, et al. (1996). "How partial C7 deficiency with chronic and recurrent bacterial infections can mimic total C7 deficiency: temporary restoration of host C7 levels following plasma transfusion." Immunology **88**(3): 407-11.

Würzner, R., A. Orren, et al. (1992). "Inherited deficiencies of the terminal components of human complement." Immunodefic Rev **3**(2): 123-47.

Zerzri, Y., M. Kallel-Sellami, et al. "[Hereditary complement C5 deficiency: study of 3 Tunisian adult cases and literature review]." Tunis Med **88**(4): 269-76.

Appendix 1 – Buffers and Reagents

A number of buffers and reagents were used during this study, the compositions of which are listed here.

Lysate Buffer

Used generally for the lysis of cells

0.1 M NaCl, 50mM Tris pH8, 10 mM EDTA

For 100 ml, add 2ml of 5 M NaCl, 5ml of 1 M Tris pH8, 2 ml 0.5M EDTA and make up to 100ml with distilled water. This should be autoclaved.

5% Nonfat Dry Milk

The preferred method is that this should be prepared using Marval nonfat Dry Milk in TBST.

10% Ammonium Persulfate (APS) Solution

0.12 g APS

1.2 ml H_2O

TBST

This can also be made without azide, which can be added back for longer incubations such as overnight blocking.

10mM Tris, 0.15 M NaCl, 8mM sodium azide, 0.05% tween-20, pH 8.0

Chemical	For 4 litres	For 500 ml
Tris	4.84 g	0.61 g Tris
NaCl	35.06 g	4.38 g NaCl
NaN$_3$	2.0 g	500 µl 10% NaN$_3$
Tween-20	2.0 ml	250 µl Tween-20

Adjust pH to 8.0 with HCl.

SDS PAGE Tank Buffer

25mM Tris, 0.2 M glycine, 0.35% SDS

Chemical	For 4 litres	For 500 ml
Tris	12.11g	1.51 g Tris
Glycine	57.6 g	7.2 g
SDS	4.0 g	17.5 ml 10% SDS

Wet blot transfer buffer (3L)

25mM Tris, 0.2M glycine, 20% methanol.

9.08 g Tris

43.24 g glycine

600 ml methanol

Make up to 3L with water.

Lower Gel Buffer

For 100 ml:

18.17 g Tris

4 ml 10% SDS

Adjust pH to 8.8 with HCl. Make up to 100 ml with water.

Upper Gel Buffer

For 100 ml:

6.06 g Tris

4 ml 10% SDS

Adjust pH to 6.8 with HCl. Make up to 100 ml with water.

SDS 4x Reducing Sample Buffer

This is made from stock sample buffer which contains0.24 M Tris, 0.24 M

SDS, 40% glycerol, 20% 2-mercaptoethanol, pH 6.8

8 ml of stock sample buffer

2 ml of concentrated 2-mercaptoethanol

Transfer to 1 ml aliquots and store at -20°C.

2x Loading Buffer

This contains 125mM Tris, 4% SDS, 20% glycerol, 10% 2-mercaptoethanol, 0.004% bromophenol blue.

2.5 ml 0.5 M Tris-HCl, pH6.8

4.0 ml of 10% SDS

2.0 ml glycerol

1.0 ml concentrated 2-mercaptoethanol (reducing only)

0.4 mg bromophenol blue

Make up to 10 ml with water. Transfer to 1 ml aliquots and store at -20°C.

Stock Sample Buffer

0.312 M Tris-HCl, 0.346 M SDS, 50% glycerol, pH 6.8

3.03 g Tris

8.0 g SDS

40 ml glycerol

Adjust pH to 6.8 with HCl. Make up to 80 ml with water.

ACD

This is an anticoagulation reagent

1.32g of sodium citrate

0.48g of citric acid

1.47g of dextrose

Add distilled water to 100ml

The filter sterilize through 0.2um filter

TAE

This is the standard running buffer for agarose gels and the base liquid to which agarose is added to make a DNA gel

For 1 litre of 50XTAE stock solution

242g Tris

87.1mls of glacial acetic acid

100mls of 0.5M EDTA pH8.0

Appendix II – Genomic Primers Details

Sequence	Primer	Sequence	Primer	Anneal (°C)
Complement component C5				
AGATCATCTTGTGAGCCCTGTC	Promoter F	GCTGTTCTCAGAAGTAGCAAG	Promoter R	61
ATGGAATCTCTCAAGTGAGAG	2 F	CATCTAATAATCCCACAGTGG	2 R	58
ACTGAAGGGCTTCGGGCAAAATTT	3 F	GAAACCGATCTCCACTCAATTTCA	3 R	58
GGAATCCCTGAGGATTCTCTT	4 F	AGTGTAGTGTGAGGACAGGGT	4 R	58
GACATGGGTATTTGTCTTC	5 F	GCTTCTTCACCAAAAGTTC	5 R	58
GGAATTTATTAGATCTCCTC	6 F	GGAACAGCGAGACTCCA	6 R	54
CCAGAAAGTCTAGCTGTTTTGCTA	7 F	CAATTAGTCCCCAAATCATGGCA	7 R	58
CTGAATCCTGTTATTGTAACCAT	8 F	GTAACTTCTCATCTCTCACCTCT	8 R	58
GGACTCGGTTATCTACTTTACTTTTGACAC	9 F	GACTAAAAAGAAGCAGAGCCCTACTGCC	9 R	58
ACTCCAACCTGGGCAACGTAG	11 F	CACTAGCCAAAACACATGGTC	11 R	60
GTAATGGAGGGTTTTTTCCCCTTT	12 F	TGCTCTTGAAAGTGCTTTGGGAAT	12 R	58
AGCTCCAGGCTGCATCCTAAAGCACA	13+14 F	CTTCCTCCATATCCTAGCCTGGTGGC	13+14 R	60
GGAATGACTTGTGAAGGCTCTAAATAC	16 F	CCTTAGATTAAGGCTATGGCAAATG	16 R	58
GGTATCCCTTGAGATTCCTTCTAAG	17 F	GCATAAAATTACACAAAATTCCCCA	17 R	58
CCAAAGTGCTGGGATTACAGG	18 F	AGTGTATCGCACATAGTTGTG	18 R	58
GATGAGGTTTTCCTTAATGATGTG	19 F	CTACCATTTAAGTCCTGGGTAAAC	19 R	58
GTGAGCTGAGATCATGCCGC	21 F	TTACACTATTGTTTCTCTCC	21 R	58
TGCTCTGGTCAATATGCAA	22 F	CCAGAAGTGGATACTACTGA	22 R	58
TAATCTGCAAACAGTCTTCC	23 F	GCTACTGAACACGGAAGTG	23 R	56
CTCTAGATGGTTGTCTTC	24 F	GTTGAGCCAAAATGCTACAG	24 R	54
CAGAGCATATGCTTGCTGCC	25 F	CCTGAGCAACTGGAGGAATG	25 R	54
AGATGTGCTGTTCACATTTG	26 F	AGATGTGCTGTTCACATTTG	26 R	52
CCAGCCTGGTCAATATGGTG	27 F	GGTGGTCTCTAATATCTATC	27 R	58
GCCATCCATTTCCATTAGGT	28 F	GTTCACTGTCACCACCCA	28 R	58
TGCATTGGTCATCTTCAGGA	29 F	GAGGCCAGATGAGTGTGGTG	29 R	58
TCATCTCCAGACAATGCTGC	30 F	GAGTGGAACAGATGATTTCA	30 R	54
CAGTGTCTCATCCAGTTACTC	31 F	GAGGATAGTTACAAAGATTGC	31 R	58
GTACACCATTTGTGCTGGAT	32 F	CTTGGATGTCAACCTGCTA	32 R	54
GATTCTGTTCTGACAGGTTC	33 F	TTGTATCTTCTGGACATACC	33 R	52

Appendix III – PCR additives

PCR enhancement materials	Chakrabarti Nucleic Acids Res 2001 Jun 1;29(11):2377-81
2-pyrrolidone	1st: most effective because it give the highest level of amplification of the target band and has one of the highest specificities and has the widest range
acetamide	2nd: good amplification, highest specificity, not quite as effective as pyrrolidone
N-methylpyrrolidone (NMP)	3rd place
B-hydroxyethylpyrrolidone (HEP)	3rd place
propionamide	4th place
NN-dimethylacetamide (DMA)	5th place
N-methylformamide (MMP)	last
NN-dimethylformamide (DMF)	last
formamide	Formamide is generally used at 1-5% and 10% formamide is reported (Gelfand 1989) to have no effect on the activity of Taq polymerase, however, Sarker et al. (1990) (see table for ref.) found that 1.25% formamide worked as well as 2.5% and 5%, and no amplification was seen at 10% so it seems prudent not to use concentrations of formamide greater than strictly necessary for optimal amplification. Nucleic Acids Research 18: 7465
N-methylacetamide (MMA)	last
	http://www.micro.msb.le.ac.uk/pcr/additives.html
DMSO	DMSO at 2-10% may be necessary for amplification of some templates, however 10% DMSO can reduce Taq polymerase activity by up to 50% (Gelfand 1989) so it should not be used routinely. DMSO is thought to reduce secondary structure and is particularly useful for GC rich templates. Amplifications 5: 16
polyethylene glycol	
Betaine	Biochemistry 32: 137 A number of PCR additives are now comercially available, however the identities of these agents are not usually revealed by their suppliers. Frackman et al.(1998) have demonstrated (using NMR analysis) that the PCR additive provided by QIAGEN in their PCR core kit (Q-Solution) and that provided by CLONTECH in the Advantage-GC cDNA PCR kit is in fact Betaine which is available at a fraction of the cost as a 5M solution from Sigma-Aldrich (cat. # B 0300), but be sure to use Betaine or Betaine (mono)hydrate and not Betaine HCl. Other products suspected of consisting largely of Betaine include the "GC-RICH solution enhancer" from Roche, "TaqMaster enhancer" from Eppendorf, "GC-melt" from Clontech and "FailSafe enhancer" (formerly "MasterAmp PCR Enhancement Technology") from Epicentre (Weissensteiner, pers. comm.). Betaine is generally used at a final concentration of 1.0-1.7M. Frackman, S., Kobs, G., Simpson, D. and Storts, D. 1998. Betaine and DMSO: enhancing agents for PCR. Promega Notes 65: 27.
TMAC	(tetramethylammonium chloride)Nucleic Acids Research 18: 4953 TMAC is generally used at a final concentration of 15-100mM to eliminate non-specific priming. TMAC has is also used to reduce potential DNA-RNA mismatch (Proceedings of the National Academy of Sciences of the United States of America 82: 1585) and improve the stringency of hybridization reactions (Nucleic Acids Research 16: 4637). TMAC with formamide (TMACIDE) offers both an enhancement in specificity and an increase in the yield of PCR products. Employing TMACIDE in the reaction mix effectively eliminates the need for extensive optimization of the PCR and will certainly ease future cross-species primer screening projects
7-deaza-2'-deoxyguanosine	(dC7GTP)Nucleic Acids Research 16: 3360 The base analogue 7-deaza-2'-deoxyguanosine may facilitate amplification of templates with stable secondary structures when used in place of dGTP in a ratio of 3: 1, 7-deaza-2'-deoxyguanosine: dGTP.
BSA	(bovine serum albumin)BioTechniques 23:504 BSA has proven particularly useful when attempting to amplify ancient DNA or templates which contain PCR inhibitors such as melanin.
T4 gene 32 protein	Applied and Environmental Microbiology 62:1102-1106
glycerol	improves the amplification of high (G+C) templates (Smith et al., 1990) Smith KT, Long CM, Bowman B and Manos MM (1990). Using cosolvents to enhance PCR amplification. Amplifications 9/90 (5):16-17.
nonionic detergents	Triton X-100, Tween 20 or Nonidet P-40 Biotechniques 12: 332 Non-ionic detergents stabilise Taq polymerase and may also supress the formation of secondary structure. 0.1-1% Triton X-100, Tween 20 or NP-40 may increase yield but may also increase non-specific amplification. As little as 0.01% SDS contamination of the template DNA (left-over from the extraction procedure) can inhibit PCR by reducing Taq polymerase activity to as low as 10%, however, inclusion of 0.5% Tween-20 or -40 will effectively neutralise this effect (Gelfand 1989).
new web site	http://www.alkami.com/reviews/idxrvws.htm

Appendix IV - DNA recombination motifs

Motif description	Motif sequence(s)
Vertebrate/plant topoisomerase I consensus cleavage sites	CAT,CTY,GTY, RAT
Vaccinia topoisomerase I consensus cleavage site	YCCTT
Vertebrate topoisomerase II consensus cleavage site	RNYNNCNNGYNGKTNYNY
Drosophila topoisomerase II consensus cleavage site	GTNWAYATTNATNNR
Heptamer recombination signal	CACAGTG
Nonamer recombination signal	ACAAAAACC
Immunoglobulin heavy chain class switch repeats	GAGCT,GGGCT,GGGGT,TGGGG,TGAGC
Translin target sites	ATGCAG,GCCCWSSW
X element	