MONASH University

A biochemical characterisation of the MACPF/CDC pore assembly

Bradley A. Spicer

BSc (Hons)

A thesis submitted for the degree of Doctor of Philosophy at Monash University

Submitted October 2018

Biomedicine Discovery Institute

Copyright notice

Abstract

The Membrane Attack Complex (MAC) of the vertebrate immune system is the terminal, pore-forming, effector of the complement system. The MAC is responsible for mounting an immune response against a broad range of target such as bacteria, protozoa and parasites. Conversely aberrant MAC formation on host cells can contribute to inflammatory conditions.

The MAC is a member of an ancient family of pore forming proteins: the MACPF/CDC family. Structural studies on the archetypal members of this family (including perforin, CDCs and pleurotolysin) show that these pores assemble using three major steps. These steps include initial membrane binding, oligomerisation into a pre-pore structure and simultaneous penetration of each component into the lipid membrane. Conversely, formation of the MAC is strikingly different from these pores. The MAC assembles with several different components that lack membrane binding domains for initial interaction with the pathogen membrane. Furthermore, there is evidence for sequential insertion of the membrane inserting regions of these proteins which has an exquisite level of control. In this regard, the mechanism of MAC pore formation has been elusive.

Here, a comprehensive structural investigation of C9 in both the monomeric and oligomeric pore-like state was performed using cryogenic electron microscopy and X-ray crystallography. The structures of C9 highlight the role of the thrombospondin domain at the oligomer interface. This is postulated to overcome the need for a dedicated membrane binding domain.

Comparison of the monomer and oligomer C9 structures revealed how this component is sequentially recruited from solution to the growing membrane associated MAC in a well-controlled fashion, and how this molecule undergoes conformational

change to form the final pore. Further, the X-ray structure of monomeric C9, together with biochemical analyses, suggested that one of the membrane spanning regions functions to prevent premature self-oligomerisation of C9 in blood plasma by blocking the elongation interface.

Taken together, the structural data presented in this study reveals how the MAC has evolved to assemble *via* direct recruitment of the pore forming component from solution and how this process is controlled with respect to C9 assembly. This event proceeds without the need for a membrane binding domain, such as those observed in other MACPF/CDCs. This mechanism ensures that the MAC can function to assemble on a wide range of pathogens.

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Signature: █████████████

Print Name:…Bradley A. Spicer…

Date:…………23/10/2018………

IV

Thesis including published work declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes **2** original papers published in peer reviewed journals. The core theme of the thesis is MACPF/CDC proteins. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the student, working within the Monash Biomedicine Discovery Institute under the supervision of Associate Professor Michelle A. Dunstone.
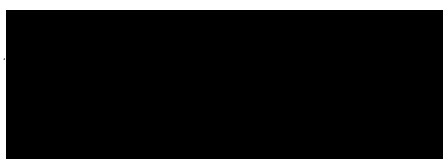
The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research.

In the case of **Chapter 3** and **Chapter 5** my contribution to the work involved the following:

Thesis Chapter	Publication Title	Status (published, in press, accepted or returned for revision)	Nature and % of student contribution	Co-author name(s) Nature and % of Co-author's contribution*	Co-author(s), Monash student Y/N*
Chapter 3	Structure of the poly-C9 component of the complement membrane attack complex	Published (Nat. comm)	34%. Protein purification (12%), Sample screening and optimisation for TEM experiments (11%), functional analysis (11%)	1) N.D., cryo-EM experiments, TEM map refinement, 33% 2) C.R., atomic modelling 33%	No Yes
Chapter 5	The first transmembrane region of complement component-9 acts as a brake on its self-assembly	Published (Nat. comm)	50%. cryo-EM optimisation, data collection, processing, and model building (25%), crystallography (crystal trials, optimisation of crystals, data collection, modelling) (25%)	R.H-P.L., Crystallography (data collection, processing and model refinement) 50%	No

I have not renumbered sections of submitted or published papers in-order-to generate a consistent presentation within the thesis for **Chapter 3**

Student signature   Date: 23/10/2018

V

The undersigned hereby certify that the previous page declaration correctly reflects the nature and extent of the student's and co-authors' contribution to this work. In instances where I am not the responsible author I have consulted with the responsible author to agree on the respective contributions of the authors.

Main Supervisor signature   Date: 23·10·2018

Acknowledgments

Dedication

This thesis has two dedications. First, to my Mum and Dad, who encouraged me to go as far with my education as possible without ever pushing me one way or the other. I love you both very much.

Next, this work is dedicated to my first science mentor and friend Prof. Kathleen (Kotty) Postle, from The Pennsylvania State University, who taught me to take my science seriously, and myself less so. I promise that I have not lost my kazoo.

Units of measurement

°	Degree
°C	Degree Celsius
Å	Angstrom (1 Å = 1 x 10^{-10} meters)
AU	Absorbance unit at a given wavelength
bp	Base pair
kb	Kilobase
kDa	KiloDalton
keV	Kiloelectron Volt
L	Liter
M	Molar (mole/liter)
MDa	MegaDalton
mg	Milligram
mL	Milliliter
mM	Millimolar
mS/cm	milliSiemen per centimetre (conductivity measurement)
MW	Molecular weight
rcf	Relative centrifugal force
V	Volts
v/v	Volume/volume
w/v	Weight/volume
w/w	Weight/weight
μL	Microliter
μM	Micromolar

X

Abbreviations

Amp	Ampicillin
APS	Ammonium persulfate
BDT	BigDye Terminator
β ME	β -mercaptoethanol
BSA	Bovine serum albumin
C1	Complement component 1
C2	Complement component 2
C3	Complement component 3
C4	Complement component 4
C5	Complement component 5
C5b6	Complement complex of C5 and C6
C5b7	Complement complex of C5, C6 and C7
C5b8	Complement complex of C5, C6, C7 and C8
C5b9	Complement complex of C5, C6, C7, C8 and C9
C6	Complement component 6
C7	Complement component 7
C8	Complement component 8
C9	Complement component 9
CAPS	N-cyclohexyl-3-aminopropanesulfonic acid
CDC	Cholesterol dependent cytolysin
CHT	Ceramic hydroxyapatite
Cm	Chloramphenicol
Cryo-EM	Cryogenic electron microscopy
CV	Column volume
DAMPs	Damage-associated molecular patterns
DMSO	Dimethyl sulfoxide
DTT	Dithiolthreitol
EAC1-8	Erythrocytes/Antibody/Complement components 1-8
EDTA	Ethylene diaminetetra-acetic acid
FEG	Field Emission Gun
FPLC	Fast protein liquid chromatography
Gent	Gentamycin
GPI	Glycosylphosphatidylinositol
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
MAC	Membrane Attack Complex
MACPF	Membrane attack complex perforin-like family
PAGE	Polyacrylamide gel electrophoresis
PAMPs	Pathogen-associated molecular patterns
PDB ID	Protein databank identification
PFP	Pore forming protein
PMSF	Phenylmethyl sulphonyl fluoride
PNH	Paroxysmal nocturnal haemoglobinuria
PRR	Pattern-recognition receptor
SDS	Sodium dodecyl sulphate
SEC	Size exclusion chromatography
sRBC	Sheep red blood cells
TAE	Tris Acetate EDTA

| TBS | Tris-buffered saline |
| TEM | Transmission electron microscope |

Table of Contents

List of Tables (not including chapter 3 and 5)

List of Figures not including chapter 3 and 5

Other publications arising from this thesis: Chronological order

[1] Perforin—A key (shaped) weapon in the immunological arsenal

Spicer, B.A., Conroy P.J., Law, R.H.-P., Voskoboinik, I., Whisstock, J.C. 2017

Published review – Seminar in cell & developmental biology

[2] Delivery of femtolitre droplets using surface acoustic wave based atomisation for cryo-EM grid preparation

Ashtiani, D., Venugopal, H., Belousoff, M. Spicer, B.A., Mak, J., Neild, A., de Marco, A. 2018

Published – Journal of structural biology

[3] Serum glycoprotein biomarker validation for esophageal adenocarcinoma and application to Barrett's surveillance

Shah, A.K., Hartel, G., Brown, I., Winterford, C., Na, R., Le Cao, K-A., Spicer, B.A., Dunstone, M.A., Phillips, W.A., Lord, R.V., Barbour, A.P., Watson, D.I., Joshi, V., Whiteman, D.C., Hill, M.M. 2018

Published – Molecular & Cellular Proteomics

1.1 Discovery of the complement system

Over 100 years ago, it was observed that the fluid portion of the blood (plasma) had lytic properties against red blood cells of different species (1). Further characterisation classified plasma into two components: a heat stable component that contains antibodies, and a heat labile fraction (>56 °C). The interaction between the heat labile fractions and antibodies is key to the observed lytic function and the ability to damage membranes (2). The heat labile portion contains around 35 protein effectors, now collectively known as the complement system. Complement is a part of the vertebrate immune system and the components responsible for forming pores have been identified as belonging to the Membrane Attack Complex / PerForin / Cholesterol Dependent Cytolysin superfamily (MACPF/CDCs) (3). This introduction describes the complement pathway and how the terminal membrane attack complex (MAC) has been harnessed as a lytic effector of the pathway.

1.1.1 Overview of the complement system

The vertebrate immune system comprises two parts that function to defend the body against microbes and pathogens. Firstly, innate immunity functions as a constitutive immune surveillance system and responds rapidly to foreign threats. Part of the innate system includes complement and cells such as macrophages and natural killer cells. Secondly, the adaptive immune system, which includes B-cells and T-cells, contributes to long-lasting and specific immunity against microorganisms (4). The effects of the innate immune system are fast acting (minutes-days), which contrasts with adaptive immunity that can take days to mount an immune response.

The complement pathway functions in innate immunity and comprises three parts; the activation, amplification and downstream cascades (4) (**Figure 1-1**). Overall, when

the complement pathway is activated and amplified there is formation of downstream effectors that function to mark the foreign targets for phagocytosis (opsonization), chemoattraction of neutrophils to the site of infection, and assembly of the membrane attack complex (MAC). Other effects include activation of B cells and general inflammation.
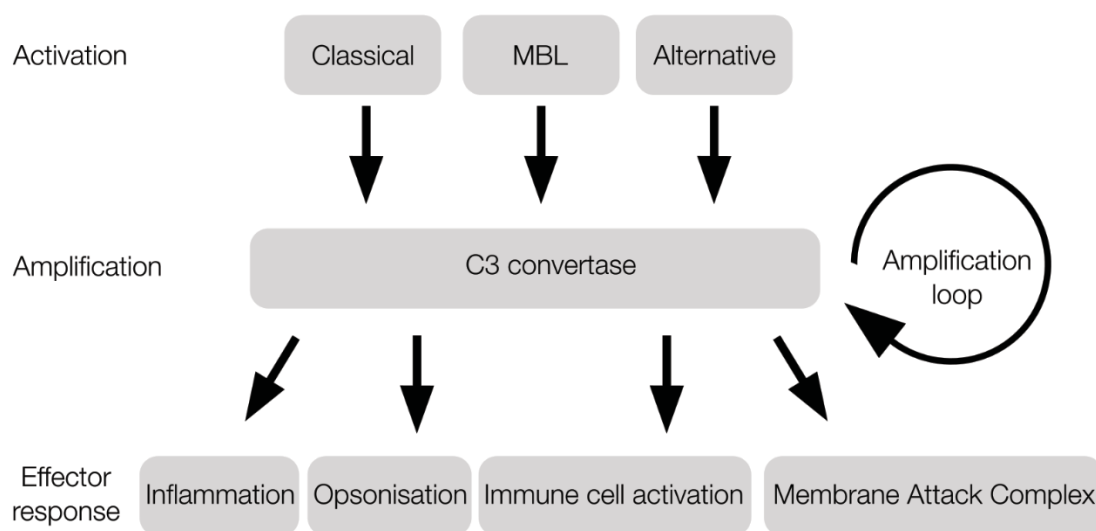
Figure 1-1 Schematic of complement pathways.

Activation of complement pathways converge at an amplification loop which leads to the effector responses.

1.1.2 Activation and amplification of complement

There are three activation pathways of the complement system: the classical, the mannose-binding lectin and alternative (**Figure 1-1**). These systems represent the earliest stage of interaction with potential targets that lead to complement amplification and immune responses. The classical and mannose binding lectin pathways are examples of pattern-recognition receptors that can associate with antibody-antigen complexes and pathogen associated molecular patterns (PAMPs), respectively (5,6). Conversely, the alternative pathway has a continuously low level of spontaneous activation that does not

rely on specific recognition of any foreign factors (7). Together, these activation mechanisms lead to a rapid immune response.

Activation of the classical pathway utilises the 766 kDa C1 complex, an assembly containing six C1q trimers together with two copies of C1r and C1s (see **Table 1-1**). The C1q units recognise the constant Fc region of antibody-antigen complexes, which in turn activates the associated protease C1r which then activates C1s (8). The structure of a C1-Fc complex has recently been determined and suggests a model of activation where the C1r of one complex activates an adjacent molecule of C1s to propagate the signal, as depicted by the schematic in **Figure 1-2** (9). The amplification of C1 in turn leads to downstream activation of several effectors, detailed in **Table 1-1**.

Table 1-1 Complement proteins and their respective functions.

For simplicity only select complement proteins are shown.

Protein	Pathway	Molecular weight (kDa)	Immunologic function
C1q	Activation	460	Binds C1r/C1s, recognition of Fc region of antibodies
C1r	Classical	80	Serine protease, cleaves C1s
C1s	Classical	80	Serine protease, cleaves C2 and C4
C2	Classical/MBL	108	Serine protease
C2a	Classical/MBL	74	Binds C4 and C3, cleaves C5
C2b	Unknown	34	Unknown, possibly enhanced vascular permeability
C3	Amplification	187	Precursor of C3a and C3b
C3a	Effector	9	Anaphylatoxin, binds C3a receptor (C3AR1)
C3b	Amplification	178	Opsonisation, Formation of C3 and C5 convertase
C4	Amplification	203	Precursor of C4a and C4b
C4a	Classical/MBL	9	Anaphylatoxin
C4b	Classical/MBL	194	Binds to C2a to form C3 convertase
MBL	Activation	750	Recognition of terminal sugars on foreign targets
MASP-1	MBL	90	Serine protease, cleaves MASP-2
MASP-2	MBL	76	Serine protease, cleaves C2 and C4
MASP-3	MBL	82	Serine protease, Isoform of MASP-1, cleaves pro-factor D
C5	Effector	190	Precursor of C5a and C5b
C5a	Effector	11	Anaphylatoxin, binds C5a receptor (CD88)
C5b	Effector	179	Formation of terminal MAC
Factor B	Alternative	93	Binds $C3_{(H2O)}$ in alternative activation
Bb	Alternative	63	Serine protease, Cleaves C3 and C5
Factor D	Alternative	24	Serine protease, Cleaves Factor B

Upon C1 complex activation, C4 is cleaved into two fragments: C4a and C4b. C4b then undergoes a conformational change that exposes a reactive thioester bond (**Figure**

1-2). The reactive thioester group covalently attaches to the microbial membrane. C4b furthermore binds to C2, whereupon the latter protein is cleaved by C1s to form C2a and C2b. C4b2a is a C3 convertase, and cleaves multiple copies of C3 into C3a and C3b. This catalytic activity can also form a separate C3 convertase by cleaving factor B, to make C3bBb. The latter events are central to the amplification pathway (**Figure 1-1**).
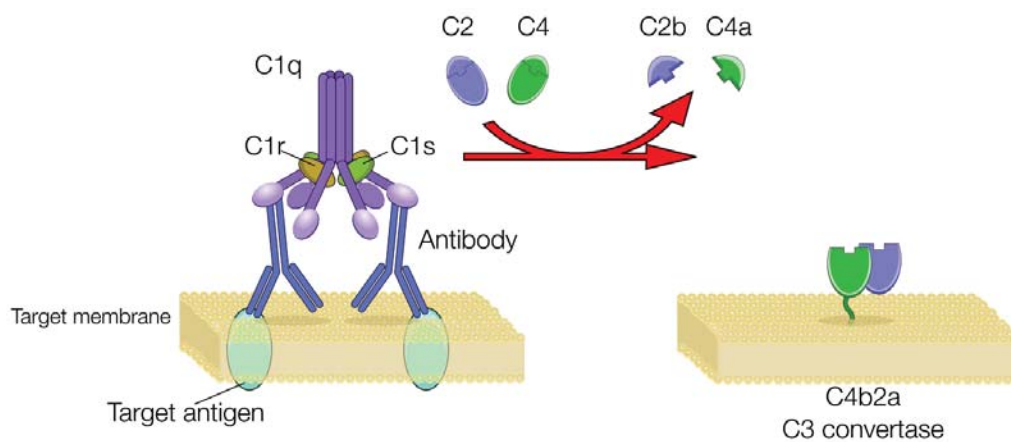
Figure 1-2 Classical pathway activation by antibody-antigen complex formation.

C1q recognises the constant Fc region of antigen bound antibodies. The association causes activation of C1r, which cleaves C1s. The latter protein cleaves C2 and C4 that form a membrane bound C3 convertase.

Analogous to the classical pathway, the mannose binding lectin (MBL) pathway utilises MBL, ficolins and collectin as pattern recognition receptors (6). These receptors specifically recognise and bind PAMPs on the pathogen surface, including lipopolysaccharide (LPS, endotoxin), lipoteichoic acid and β-glucans (10) (**Figure 1-3**). In the case of the MBL pathway, there is recognition of the terminal mannose of glycans found only on foreign surfaces of pathogens such as bacteria and yeast. Self-recognition of host tissue is excluded due to the terminal sialic acid sugars capping the N-glycans on mammalian glycoproteins, thus the MBL pathways is not activated (11).

12
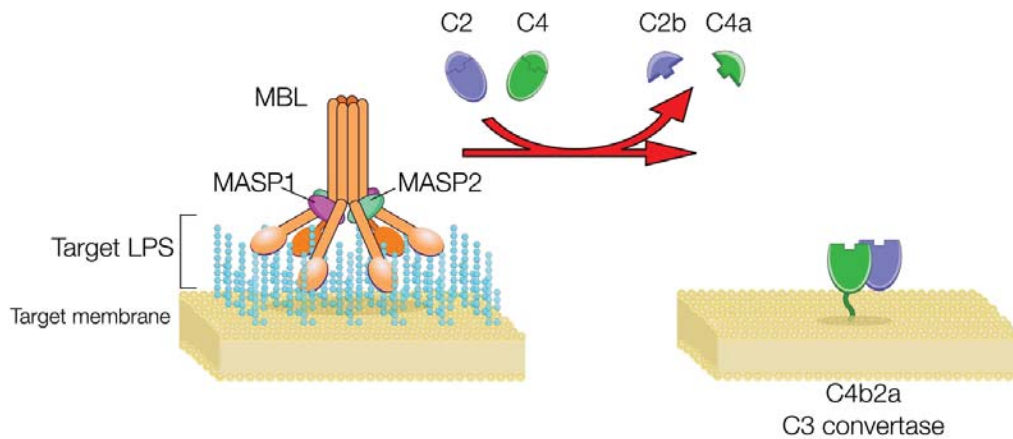
Figure 1-3 Mannose binding lectin activation.

MBL recognises PAMPS on foreign surface membranes. The association causes MASP-1 to cleaves MASP-2. Similar to C1s, MASP-2 cleaves C2 and C4 to form a C3 convertase.

In the MBL pathway, the pattern recognition receptors are in complex with MBL-associated serine proteases (MASPs): either MASP-1, MASP-2 or MASP-3 (12), to form 450 - 670 kDa structures that are analogous to C1 (13) (**Table 1-1**). Indeed, MASP proteases are homologous to the C1r and C1s proteases (14). The products of MASP activation are the same as the classical pathway, where MASP-1 activates MASP-2 (analogous to C1r (15)) and MASP-2 cleaves C4 and C2 to form a C3 convertase (**Figure 1-3**). At this point the classical and MBL pathways converge. MASP-3, on the other hand, cleaves the zymogen form of factor D to its active form, which is involved in the alternative pathway (16). In addition MASP-3 is a competitive inhibitor of the MBL pathway (17).

The alternative pathway is evolutionarily ancient compared to the classical and MBL pathways and features low-level spontaneous hydrolysis of C3 (18,19) (**Figure 1-4**). Following its auto-hydrolysis C3 can either return to its native conformation, or be irreversibly changed to C3$_{(H2O)}$ (20). C3, like C4, contains a reactive thioester that is exposed upon conformational change which can then form a covalent bond with molecules on target membranes (21). The latter protein can bind to factor B which is in

turn cleaved by factor D to form the soluble C3 convertase, C3$_{(H2O)}$Bb (20) (**Figure 1-4**).

C3$_{(H2O)}$Bb has a short half-life, however it can cleave additional C3 molecules to form

C3b. Overall this results in positive feedback and amplification of complement function

(22).
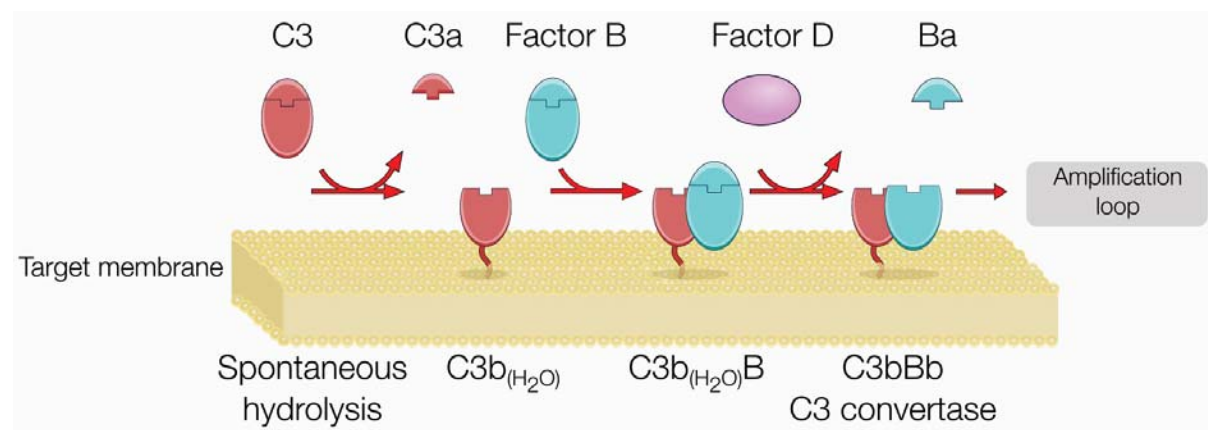
Figure 1-4 Alternative pathway activation.

C3 undergoes low-level of spontaneous activation by hydrolysis and can form C3b on a target membrane. Factor B can bind C3b (H2O) and is cleaved by factor D. C3bBb is a C3 convertase and the Bb component can cleave C3 or C5.

1.1.3   Production of C5 convertase

Each of the three activation pathways lead to the C3 convertase positive feedback

loop (**Figure 1-5**). If the amplification of C3 convertase on pathogen membranes reaches

a high concentration of C3b, then a second C3b binds to a C3 convertase thereby making

a C5 convertase. Specifically, the factor Bb or C2a of the C5 convertases (C3b**Bb**C3b or

C4b**2a**3b respectively) cleaves C5 to C5a and C5b. C5a is a potent anaphylatoxin that

binds the C5a receptor CD88 leading to downstream signalling. C5b forms the first

component of the membrane attack complex (MAC) (23,24). The MAC is the terminal

killing structure of the complement system and assembles on target lipid membranes to

form pores.

14
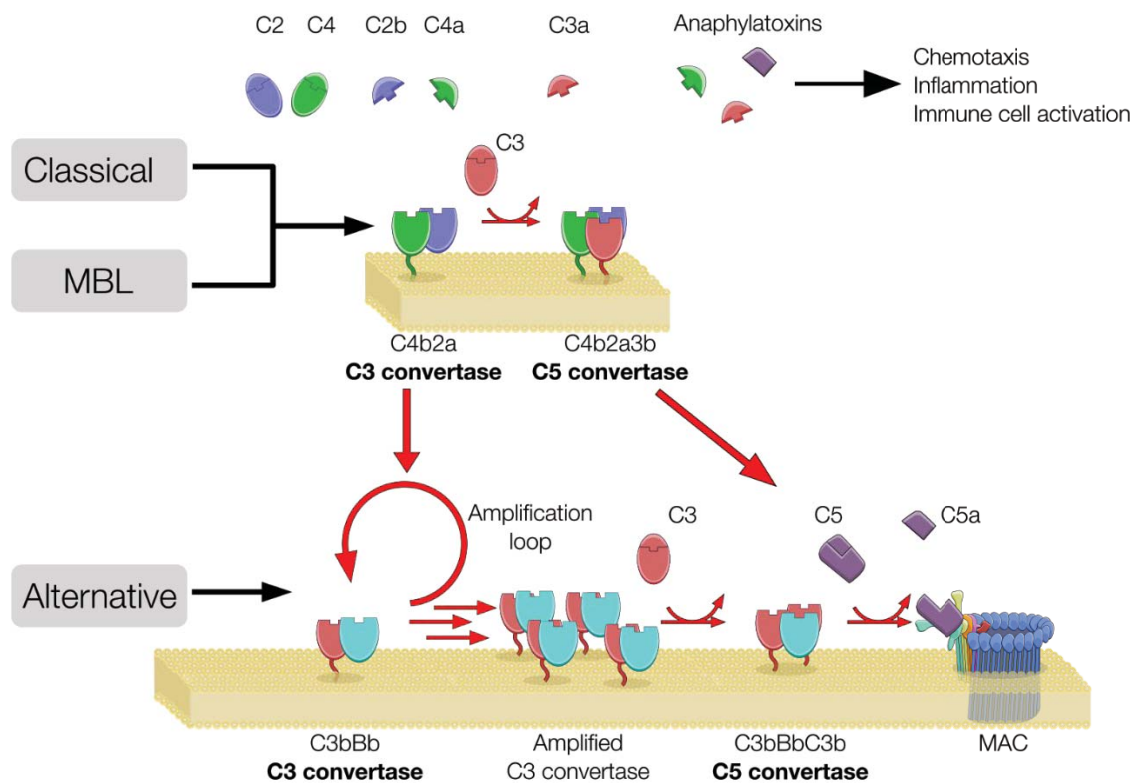
Figure 1-5 Formation of C5 convertases and effectors of complement.

C3 convertases are a central tenant of the complement amplification pathway. C3 can associate with these complexes to form a C5 convertase. The latter convertase is the first component in formation of the lytic membrane attack complex (MAC). Notably, the anaphylatoxins are formed in each of these pathways.

1.2   The membrane attack complex (MAC)

Activation and amplification of complement leads to the production of downstream effectors and initiates the assembly of the terminal MAC (25,26). The MAC was first observed using Electron Microscopy (EM) by Borsos, Dourmashkin and Humphrey who showed that it forms circular structures on erythrocyte membranes (27) (**Figure 1-6**). These EM studies suggested that the MAC formed $8 - 10$ nm pores. The MAC was subsequently shown to contain five complement components: C5b, C6, C7 C8 and C9. Furthermore, the MAC was confirmed to be the lytic component of the complement pathway (28,29). This finding formed the foundation for several decades of research and these efforts have recently culminated in a detailed molecular and structural picture of the MAC.
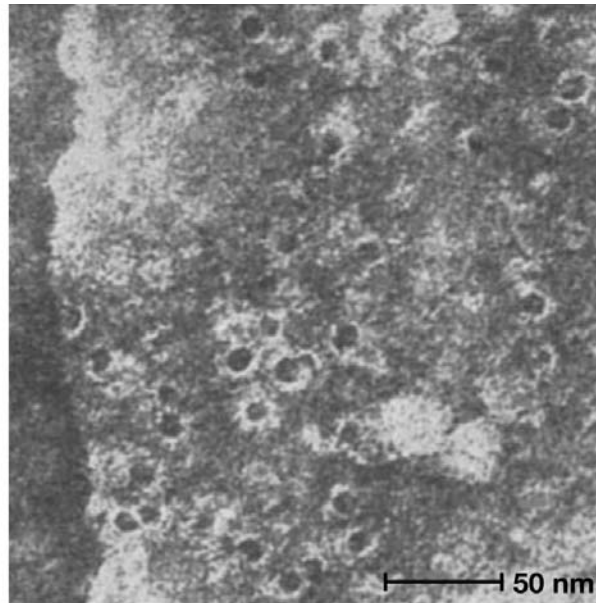
Figure 1-6 MAC pores formed on erythrocyte membranes. Electron micrograph of sheep erythrocyte treated with antibody and complement.

This research was originally published in Nature and adapted for this figure. T.Borsos, R. Dourmashkin and J. Humphrey. Pores in Erythrocyte Membranes Caused by Immune Haemolysis Nature. 1964 Vol. 202, pp. 251-252.

In vitro experiments have suggested that the MAC can assemble and form pores in the outer membrane of *E. coli* and other gram negative bacteria (30). *E.coli* treated with the MAC components have reduced viability (31). This killing effect may be the consequence of osmotic flux (30) and is also enhanced by the presence of serum lysozyme an enzyme that hydrolyses peptidoglycan located in the periplasm (31). Taken together these data suggest a possible synergistic effect between the MAC and lysozyme (32), and it is accordingly hypothesised that the MAC may function as a channel for lysozyme delivery (32).

The MAC is known to be important for providing protection against *N. meningitidis* (a causative bacterium of meningitis infections) (33). People who lack one or more components of the MAC are not susceptible to bacterial disease in general, due to redundancies of immune effectors. However, MAC deficient individuals suffer from persistent infection from *Neisseria* species (34,35). These data suggest that *Neisseria* may be predominantly killed by the MAC over other effectors of the immune system.

16

Conversely, there are many examples of bacteria that have evolved mechanisms to avoid being killed by the MAC (36). For example, the outer membrane proteins BGA66 and BGA71 from *Borrelia bavariensis* bind C7 and C9 to inhibit pore formation (37).

1.2.1 Control of the MAC and Paroxysmal nocturnal haemoglobinuria (PNH)

Aberrant assembly of the MAC on host membranes plays a significant role in inflammation and disease. The off target deposition of the MAC on host cells results in a range of different immune driven diseases with paroxysmal nocturnal hemoglobinuria (PNH) identified as the most acute MAC related disorder (38). MAC associated inflammation occurs when the MAC pores form on bystander host tissue and trigger intracellular signalling pathways. MAC induced signalling pathways involve events that can cause an influx of calcium (39), the coupling of pores to G-protein coupled receptor (through a poorly understood mechanism) (40) and activation of protein complexes (inflammasomes) that promote synthesis of inflammatory cytokines (41).

Individuals with PNH have a reduced level of the MAC inhibitor, CD59 (42). CD59 is a membrane bound inhibitor of the MAC that is found on most cell types. This protein interacts with C8 and C9 on host tissue and prevents the MAC from forming a pore (43).

CD59 is a small glycoprotein (~10 kDa) that contains a C-terminal glycosylphosphatidylinositol lipidation (GPI anchor) (44). The processing of GPI anchored proteins occurs in the lumen of the endoplasmic reticulum and involves a complex biosynthesis pathway whereby a pre-formed GPI anchor attaches to the protein via a signal peptide (45). The enzymes involved in catalysing several of the steps in GPI anchor processing are phosphatidylinositol N-acetylglucosaminyl-transferases (*PIG* gene products).

Spontaneous *PIG-A* mutations in clonal hematopoietic stem cells can result in erythrocytes that lack CD59 and the development of PNH (46,47). Uncontrolled MAC formation on CD59 deficient erythrocytes results in intravascular lysis of red blood cells, lethal thromboses and acute kidney failure (42,48,49).

Individuals with PNH typically have a poor prognosis and a median survival of 10 – 15 years after diagnosis, if left untreated (50). To date, two treatments for PNH are in clinical use – bone marrow transplantation therapy (51) and the monoclonal antibody Eculizumab (Solaris). The latter protein inhibits MAC formation through targeting C5 and preventing the formation of C5b (52).

1.2.2   Broad Structural features of the MAC

After the first observation of the membrane bound MAC in 1964 (27), several EM studies advanced our knowledge of the overall architecture of the MAC (53–56) (**Figure 1-7**). Critically, these experiments made use of individually purified complement components and artificial membranes that are more amenable to EM imaging as compared to the original erythrocyte membrane studies. This work includes the first example of EM projections of the MAC where the complex was observed from several different angles (53–56). These approaches contrast with the erythrocyte assembled pores which were top-down views only. Concomitant with this work ultracentrifugation studies suggested that the complete MAC comprises a ~1.8 MDa assembly (57).
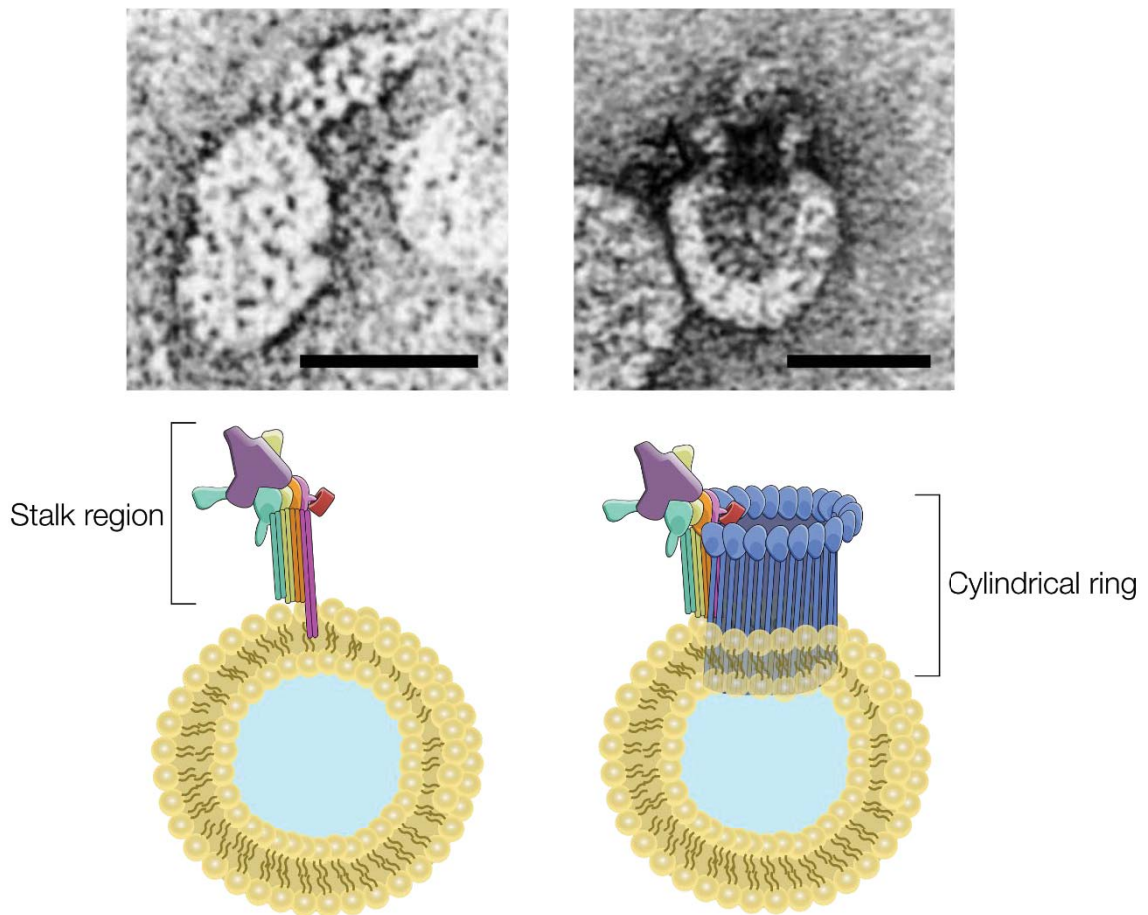
Figure 1-7 Cartoon schematic of MAC rings formed on a liposome.

Left, the stalk-like region of the MAC containing C5b, C6, C7 and C8. Right, the entire MAC showing the cylindrical ring portion that mostly comprises C9 (56). Scale bars in the top panels represent 30 nm.

The micrographs in the top row were originally published in the Journal of Biological Chemistry Jürg Tschopp, Ultrastructure of the Membrane Attack Complex of Complement (Heterogeneity of the Complex Caused by Different Degree of C9 Polymerization) J. Biol. Chem. 1984 Vol. 259, pp. 7857-7863.

Collectively, these imaging studies suggested that the MAC is comprised of two basic units: a stalk region and a larger cylinder that extends 12 nm above the membrane surface (54,55) (**Figure 1-7**). Immune gold-labelling and other biochemical characterisation revealed that the stalk region consists of a globular section of: C5b, C6, C7 and C8 (54). The complete cylindrical ring unit forms only upon addition of several C9 subunits, but also includes parts of C6, C7 and C8 (53,58).

1.3   The Molecular mechanism of MAC assembly

The molecular basis detailing how the five MAC components assemble into the final MAC pore and interact with their lipid surface has been rigorously investigated (summarised by **Figure 1-8**) (58). It is well established that the MAC assembles through sequential formation C5b $\Rightarrow$ C5b-C6 $\Rightarrow$ C5b-7 $\Rightarrow$ C5b-8 $\Rightarrow$ C5b-9 (56). Furthermore, radiolabelling and lipid binding experiments have revealed that the individual components C5b-6, C7, C8 and C9 have low affinity for phospholipids compared to the growing complexes C5b-7, C5b-8 and C5b-9 (59,60).
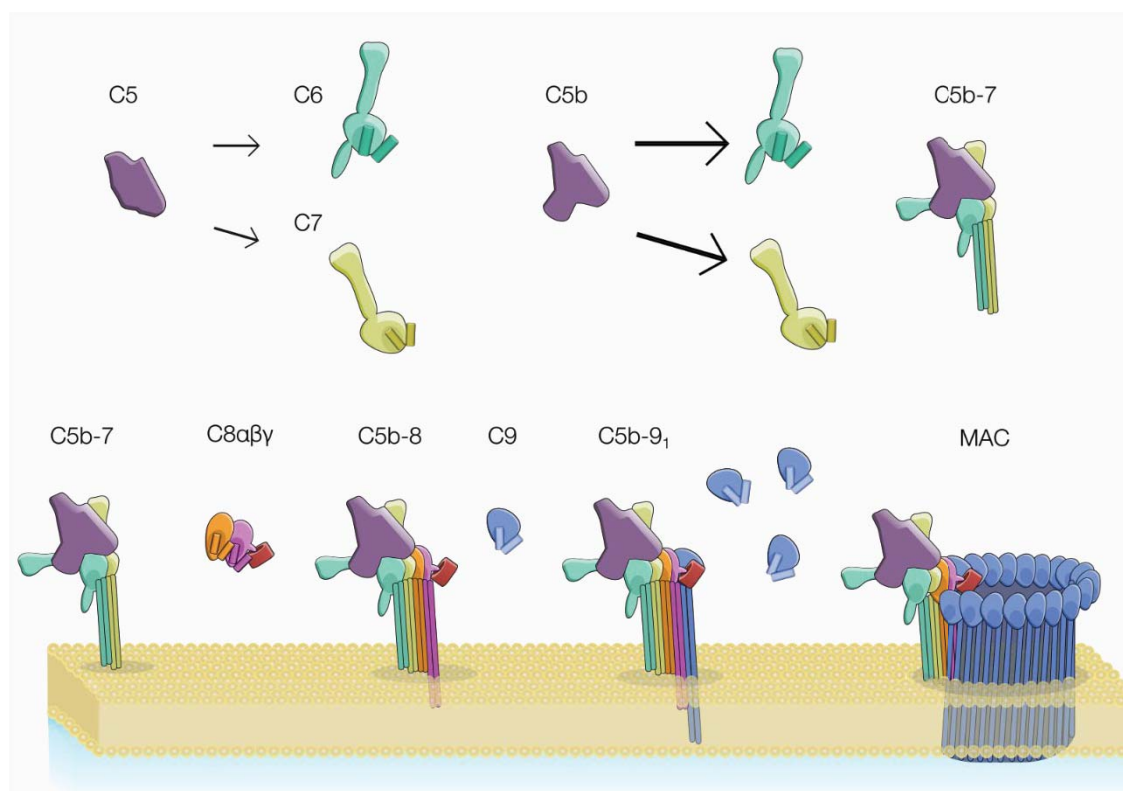
Figure 1-8 Cartoon schematic of MAC assembly and relative interaction with lipid bilayer.

The top section illustrates the solution-based interactions between C5 and C6/C7. C5 has low affinity for C6 and C7 (small arrows) compared to C5b (large arrows). The bottom section represents the assembly of MAC on the membrane by sequential assembly of the pore components.

Prior to cleavage, C5 has relatively low affinity for C6 and C7 (**Figure 1-8 top**) (61). However, upon C5 being cleaved into C5a and C5b, the affinity of the latter protein for

C6 increases and results in the formation of the C5b-6 complex. C7 is then able to bind to C5b-6 to form C5b-7 (**Figure 1-8 bottom**) (62).

Upon formation of C5b-7, a conformational change occurs that results in initial membrane binding. Specifically, C7 undergoes a conformational rearrangement that allows partial incorporation of this subunit into hydrophobic regions of the lipid bilayer (63,64). The lipid bound C5b-7 complex is more resistant to proteolysis, supporting the idea that a stable lipid interaction occurs upon insertion in the membrane (65). In the context of pathogens encountered by the immune system, C5b-7 represents the first point that the MAC strongly interacts with the target membrane. It was also shown that soluble C5b-7 complex is metastable and degrades if it is not subsequently bound by C8 to form the nascent MAC (63). Although a precise understanding of the C5b-7-lipid interaction is unclear, this complex has a preference for negatively charged membranes, such as those on bacteria (66).

C5b-7 binding to the lipid surface facilitates the recruitment of C8 to form C5b-8, a complex known as the nascent MAC. In isolation, C5b-8 can make a small ~1 nm membrane puncture causing slow leakage of radiolabelled sucrose (~0.9 nm diameter) from C5b-8 treated erythrocytes (67). The C8 is distinct from the other MAC subunits in that it is a heterotrimer of C8α, C8β and C8γ (68,69). In this trimer, C8α and C8β are bound by non-covalent interactions, whereas C8α and C8γ are covalently bound via disulphide bond. C8β is essential for binding of C8 to C5b-7 (70). Similarly, the C8α is essential for exclusively recruiting the first C9 monomer. C9 can also bind soluble, unbound C8 in solution to make a weak reversible C8-C9 (1:1) complex (61). However, when C9 binds to C8 in the nascent complex it binds irreversibly to form $C5b-9_1$ (71). $C5b-9_1$ acts as a template for the final addition of 17-19 subunits of C9 to form the

complete MAC pore. Like C7, the C8 and C9 undergo substantial conformational changes allowing insertion into the membrane (**Figure 1-8**).

In the final pore, the C8 and C9 span the lipid bilayer. This was determined through the use a photoactivatable lipid probe that interacts with integral membrane proteins (72). In these experiments, the probe was found to react with the MAC proteins only when C8 or C9 were present (72). By comparison, the C8 (and C9) binds more phospholipid than C5b, C6 or C7 a finding that suggests a higher degree of membrane insertion (59,73,74).

1.4   *Domain structure of the components of the MAC*

The domain composition of the proteins that comprise the MAC (C5b, C6, C7, C8 and C9) is illustrated in **Figure 1-9**. C5 is homologous to C3 and C4 proteins and comprises several macroglobulin domains (MG), a **C**omplement C1r/C1s/ **U**egf/ **B**mp1 (CUB) domain, a C5d and a C-terminal netrin module (C345C) (**Figure 1-9**).

 C6, C7, C8 and C9 all include a central Membrane Attack Complex PerForin-like / Cholesterol Dependent Cytolysin (MACPF/CDC) domain and several common ancillary domains. The latter includes thrombospondin-like (TSP), low-density lipoprotein receptor class A (LDL), and epidermal growth factor (EGF) domains (**Figure 1-9**). C6 and C7 (of the human form) also contain an extra set of C-terminal complement control protein (CCP) and Factor I Module (FIM) (**Figure 1-9**). C8α also binds C8γ, a lipocalin family protein. Finally, human C9 is the smallest of the MAC proteins, as it lacks the C-terminal TSP domain.
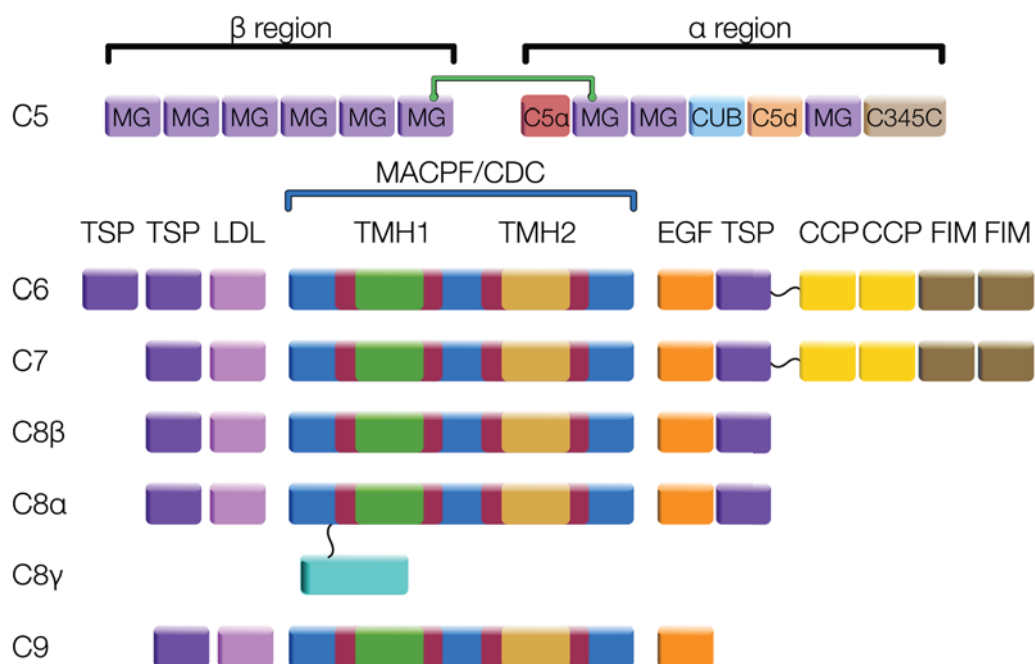
Figure 1-9 Domain layout of MAC components.

The two C5 regions contain a disulphide bond between the discontinuous MG 6 (green). The C6, C7, C8 and C9 each contain a MACPF/CDC domain (blue and red) with two transmembrane hairpins (TMH1, green and TMH2, gold). The MACPF/CDC components also contain: thrombospondin like domains (TSP, purple); low density lipoprotein receptor class A (LDL, pink); and EGF-like domain (EGF, orange); complement control protein (CCP, bright yellow) and factor I module (FIM, brown). The C8α contains a disulphide link to C8γ (cyan).

1.5   Structural studies of the MAC components

In this section, the structures of each MAC component are briefly described.

Table 1-2 Structures of terminal complement precursors subunits

*The structure of C9 (shaded row) is the subject of chapter 5. Complexes are listed in **Table 1-3**.*

Protein/complex	Accession code	Method/resolution	Reference
C5	PDB-3CU7	X-ray 3.1 Å	(75)
C5-CVF	PDB-3PVM	X-ray 4.3 Å	(76)
SSL7-C5-CVF	PDB-3PRX	X-ray 4.3 Å	(76)
C6	PDB-3T5O	X-ray 2.9 Å	(77)
C5b6	PDB-4E0S	X-ray 4.2 Å	(78)
	PDB-4A5W	X-ray 3.5 Å	(79)
C8α MACPF	PDB-2QQH	X-ray 2.5 Å	(80)
C8α MACPF/C8γ	PDB-2RD7	X-ray 2.15 Å	(81)
C8αβγ	EMD-1805	SP Cryo-EM 25 Å	(82)
C8αβγ	PDB-3OJY	X-ray 2.5 Å	(83)
C9	PDB-6CXO	X-ray 2.2 Å	Chapter 5

1.5.1 Complement C5

The mature form of C5 (190 kDa) is homologous to C3 (84) and C4 (85). C5 does not, however, contain the thioester bond found in C3 and C4 that permits covalent binding to target membranes. The structure of C5 is heart-shaped and contains 15 domains that can be broken into two regions, the α region and β region (**Figure 1-10**). These two regions are linked by a disulfide bond formed between Cys567-Cys810 (75).

The β region forms a rigid super helix structure forming a protein scaffold. The MG 6 domain contains the C5a fragment that is cleaved by C5 convertase between Arg751 and Leu752 (24). C5a dissociates from the rest of C5 and functions as a potent anaphylatoxin (section 1.1.3). The C-terminus α region undergoes a conformational change upon cleavage from C5 to C5b.
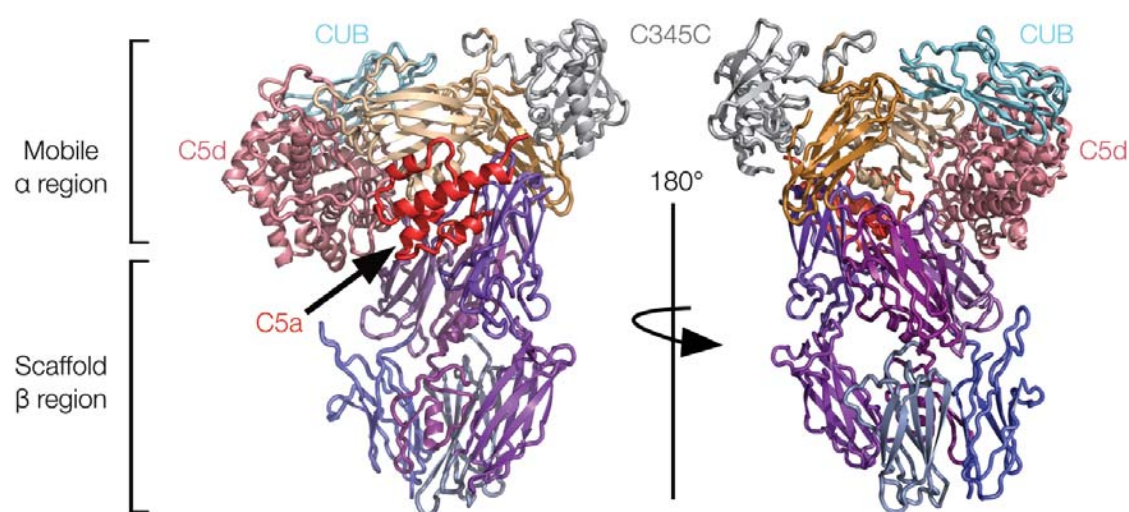
Figure 1-10 The precursor structure of complement component 5 (C5).

The scaffold region containing various macroglobulin domains is shown in different shades of purple. The mobile region containing C5d (pink), macroglobulin domains in this regions (orange), CUB (cyan), C345C (grey) and C5a fragment (red).

This includes a dramatic displacement of the C5d and CUB domains and exposes the binding site for C6 (**Figure 1-11**). In the C5b-6 structures, C6 occupies the hydrophobic residues from the C5d domain (78,79).

24
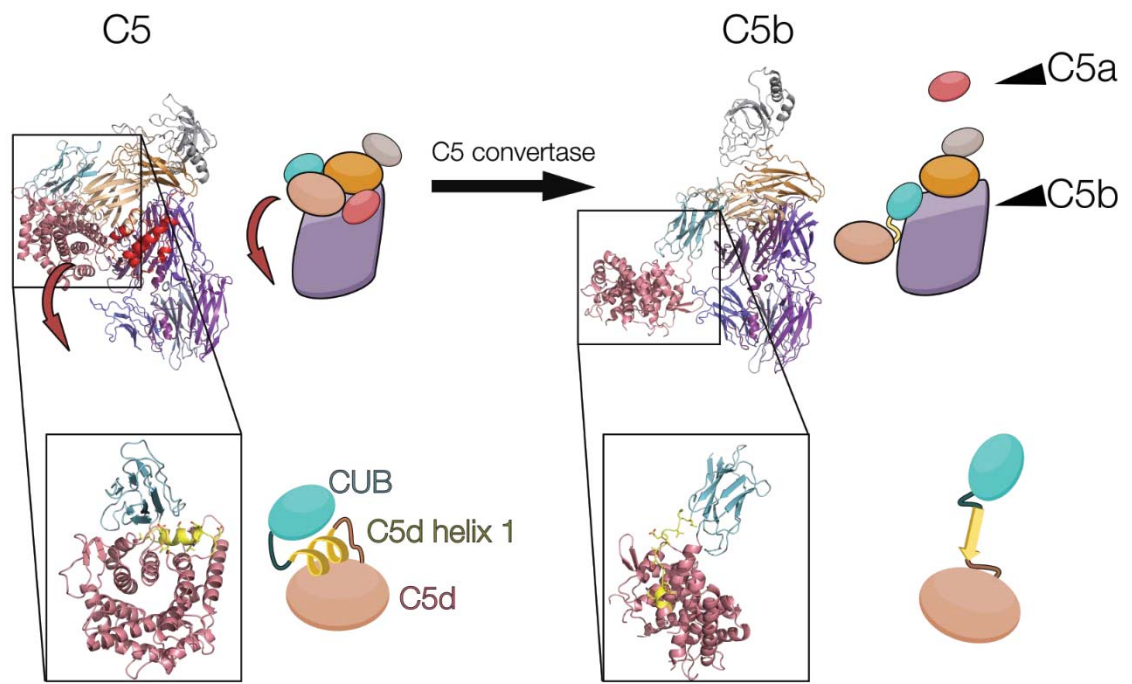
Figure 1-11 Conformational change from C5 to C5b.

Top, the cartoon structure of the C5 precursor (PDB 3CU7); and C5b from the C5b6 structure (PDB 4A5W) with corresponding schematics. The schematics show the relative domain movement of the C5d region (pink oval) and CUB (cyan). Bottom, the conformational change of the C5d domain from C5 to C5b, with the unfurling of the C5d helix (yellow) and schematic of this rearrangement.

1.5.2   Complement C6

C6 is a seahorse shaped molecule with body, tail, head and neck regions (77) (**Figure 1-12 a**). The body and tail of the C6 structure contains the MACPF, three TSPs, an LDL and an EGF domain. The head and neck region of C6 comprises two CCP and two FIM domains.
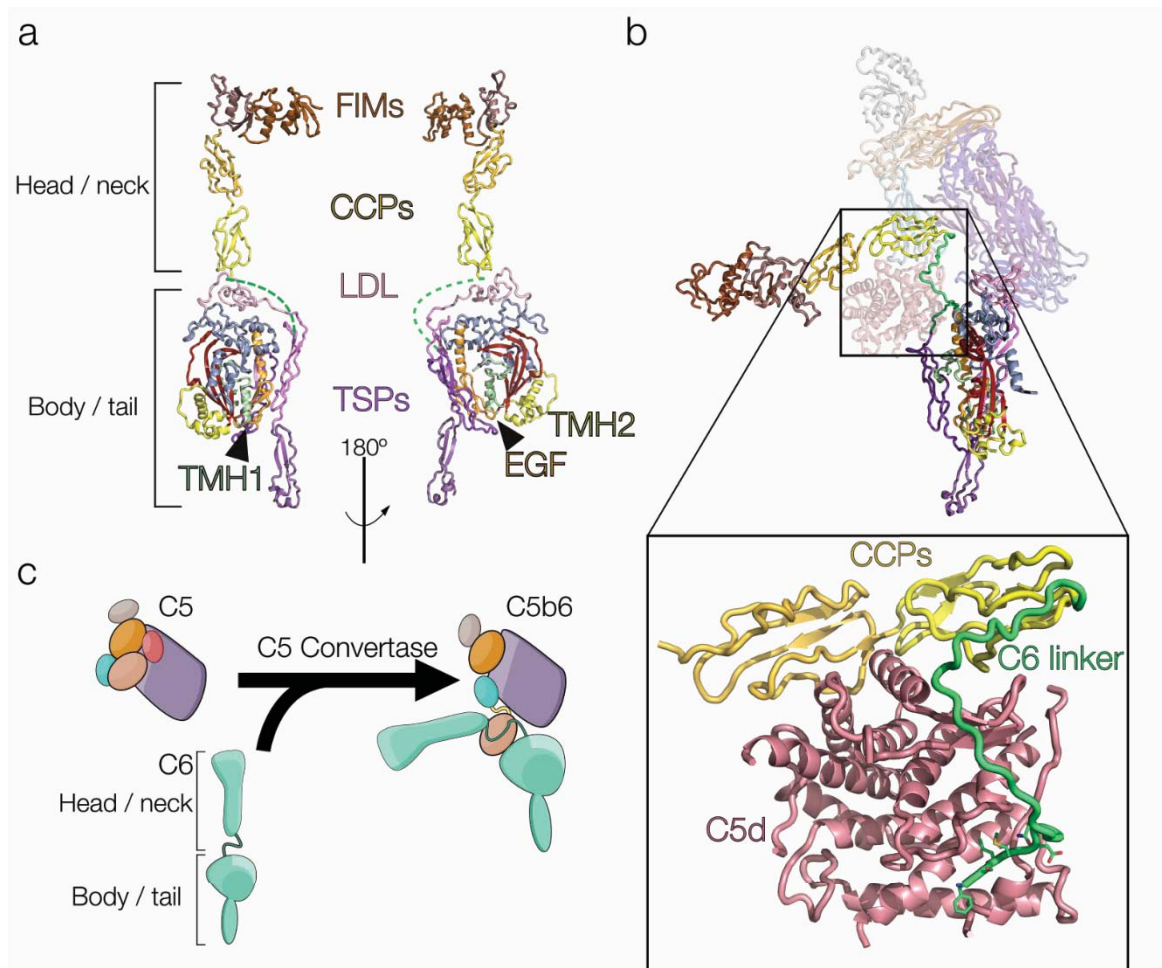
Figure 1-12 The structure of complement component 6 (C6) and rearrangement to C5b-6.

a) Crystal structure of the precursor form of C6 (PDB 3T5O) showing the seahorse structure. The C-terminal linker is shown with a dashed line (green) b) The structure of C5b6 (PDB 4A5W). For clarity, the C5 is transparent. Inset, showing interactions of the C6 linker (green, residues 612-643) with the C5d (pink) and CCP domains (yellow). c) Cartoon schematic of C6 and C5 conformational change and interaction.

Upon engagement of C6 with C5b, several conformational changes take place that render the C5b6 complex competent to recruit C7. First, the C6 head and neck region binds the C5d domain of C5b (**Figure 1-12 b**) (78,79). A C-terminal linker of C6 makes the most intimate interactions with C5b and wraps around a large portion of the hydrophobic C5d domain. Finally, two CCP domains contact the C5d domain (**Figure 1-12 b** inset). The result of the C6 interactions with C5 is that the TSP3, EGF and MACPF domains are shifted with respect to the rest of the molecule. These changes

further result in a modest conformational change in C6 MACPF domain, an event that presumably permits binding to C7.

1.5.3 Complement C7

No high-resolution structure of C7 is available. However, C7 is closely homologous to C6 and possesses a similar domain makeup (86,87). NMR studies reveal that the C-terminal FIM domains may interact with C5 in its precursor form (88).

1.5.4 Complement C8

C8 comprises a heterotrimer of C8αβγ (53,54). C8α and C8β both contain the core features observed in C6 including TSP, LDL, MACPF and EGF domains (83) (**Figure 1-13**). The complex comprises a globular structure which presents the interaction surfaces of C5b-7 and the first C9 on the C8β and C8α regions, respectively (**Figure 1-13**). The C8γ subunit is positioned such that it is in the lumen of the MAC.
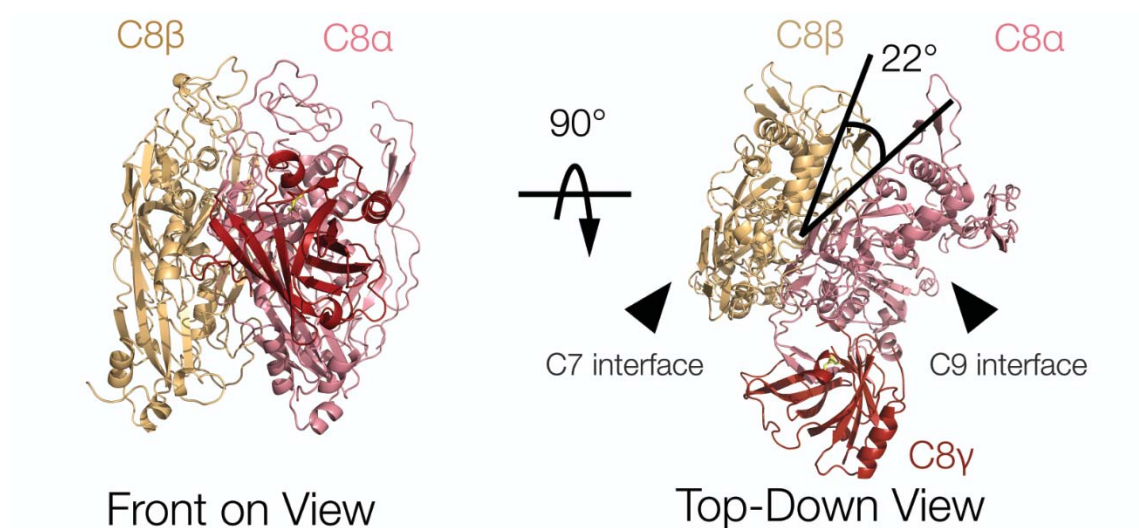
Figure 1-13 Overall structure of C8 shown in two different views (PDB 3OJY).

Left, the side-by-side globular structure of C8 as observed from the luminal side of the MAC. Right, a top-down view shows a 'MAC-like' curvature of C8 and the angular relation between the subunits. The relative position of the C7 and C9 interface sides are indicated by arrows.

1.5.5 Complement C9

The final component to be recruited to the MAC is C9, the only MAC component that self-associates and that is critical for forming a complete MAC pore. The structure of this component is the subject of **Chapter 3** and **Chapter 5** of this thesis.

1.6 Mechanism of pore formation by MACPF/CDC proteins

Bioinformatics and structural studies reveal that C6, C7, C8β, C8α and C9 all contain a MACPF/CDC domain. Thousands of MACPF/CDC proteins have been identified across all kingdoms of life (89). The role of these MACPF/CDC proteins include functions in pathogen virulence (including bacterial CDCs), immunity, venom, pathogen invasion and egress, developmental biology and neural guidance. Many of these proteins have also been shown to form large membrane-spanning pores.

The basic mechanism of pore formation by MACPF/CDC proteins has been elucidated through studies of the CDCs. These data reveal that CDCs are secreted as soluble monomers that initially bind to target lipid membranes through a domain (typically called domain 4) that is ancillary to the MACPF/CDC domain (**Figure 1-14 b**) (90). The monomers then self-associate via lateral diffusion on the membrane surface using the flat surfaces of the MACPF/CDC domain to form a large, circular pre-pore (91).
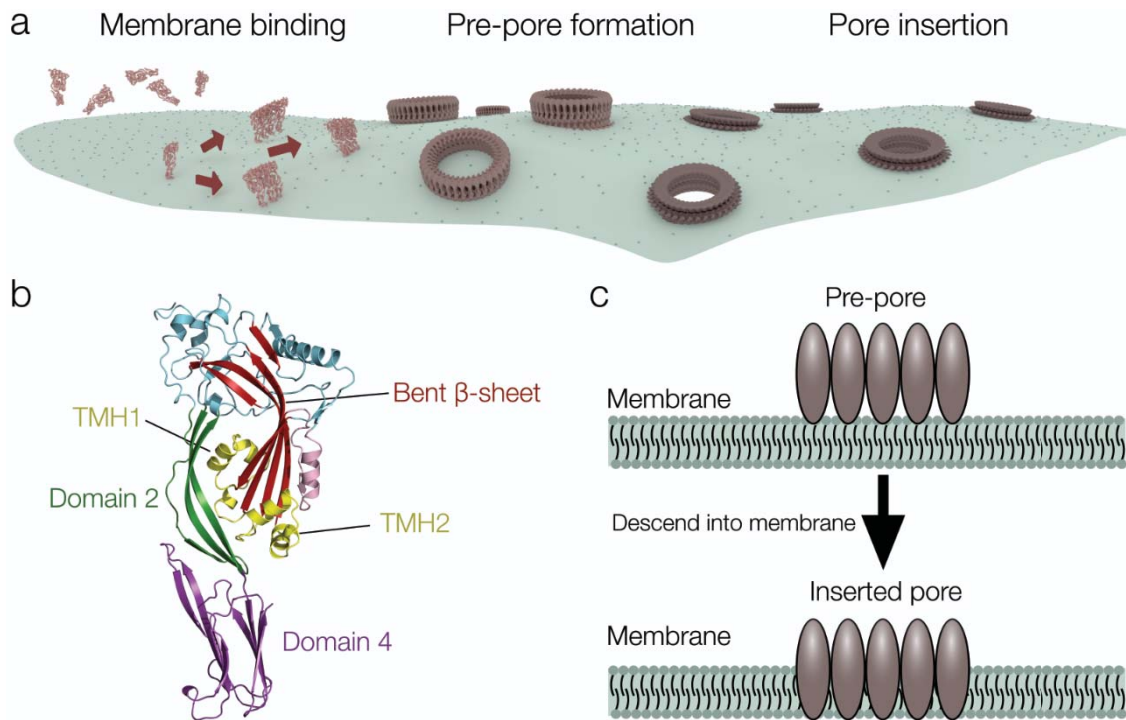
Figure 1-14 Cartoon schematic detailing the general mechanism of pore assembly by MACPF/CDCs.

a) the model of pore formation includes membrane binding, oligomerisation of subunits into a pre-pore and insertion into the membrane in a concerted manner (1PFO) and EM structures of the pneumolysin pre-pore and pore (EMDB 1106 and 1107). b) X-ray structure of perfringolysin O (PDB 1PFO). c) schematic diagram of pore insertion by descending CDC pore into the membrane.

Crystallographic studies reveal that the core of the MACPF/CDC domain comprises a central twisted four-stranded β-sheet along with two clusters of α-helices (termed Transmembrane Hairpin [TMH1] and [TMH2]) that are located at the distal end of the β-sheet (**Figure 1-14 b**) (92,93). In CDCs central β-sheet of the MACPF/CDC domain is proposed to drive assembly of the pre-pore, predominantly through the formation of β-strand interactions formed between adjacent monomers (94).

Once the pre-pore is formed it is suggested that all the TMH1 and TMH2 regions simultaneously unravel to form a large membrane inserted β-barrel (94). A typical CDC pore includes between 35-50 monomers in the assembly (95). Concomitant with membrane insertion, the entire assembly descends ~40Å towards the membrane surface (**Figure 1-14 c**) (96). The latter conformational change is required in order to bring the

(relatively short) membrane spanning TMH regions in close proximity to the membrane surface (97).

Like the gram-positive bacterial CDCs, the MACPF branch has been observed to assemble by a multi-step process as elucidated by the fungal toxin pleurotolysin (**Figure 1-15**) (98). Unlike CDCs, that use a common β-sandwich structure (domain 4) to bind membranes, MACPF proteins have adapted to use a wide range of membrane binding domains (93,98–100). Subsequently, a pre-pore is assembled (**Figure 1-15**) (98). Finally, the TMH1 and TMH2 insert into the membrane. In contrast to CDCs, the MACPFs contain longer TMH regions that can span the membrane without a vertical collapse.
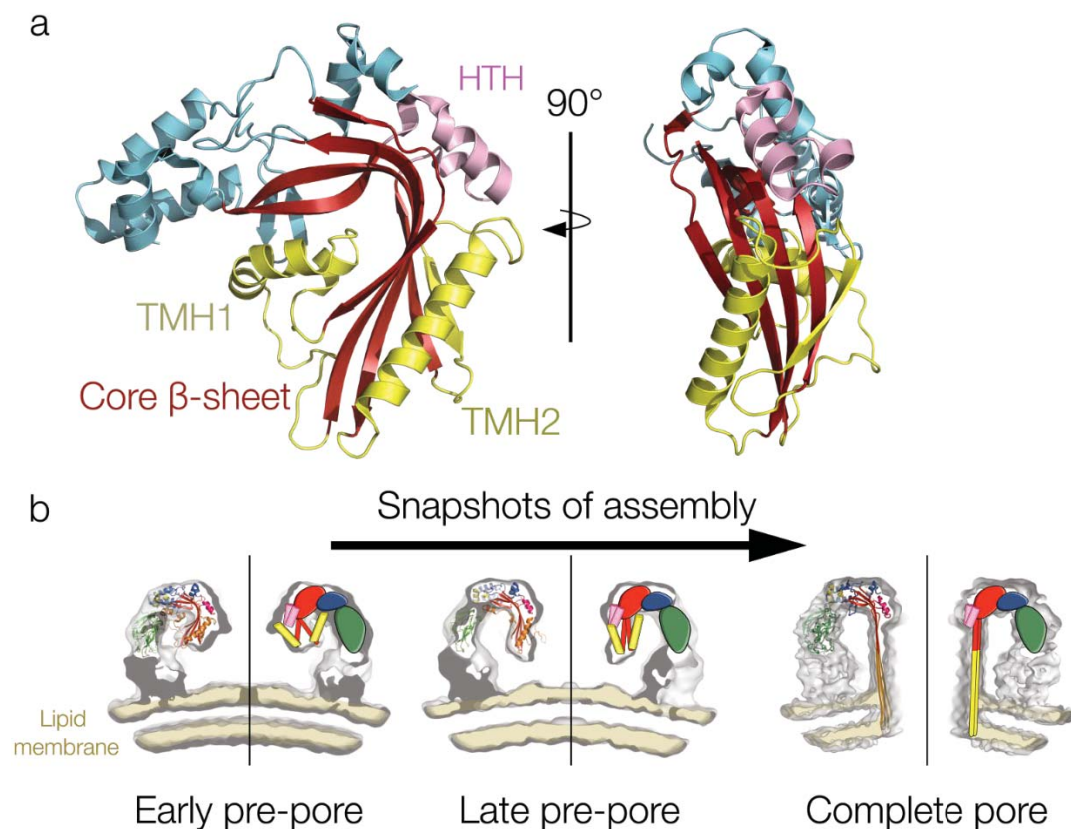
Figure 1-15 MACPF domain and assembly by pleurotolysin.

a) The crystal structure of the MACPF/CDC domain from pleurotolysin B (PDB 4OEJ). Two views are shown, highlighting the secondary-structure elements (left) and the relatively flat surface (right). The core β-sheet region (red), with TMH1 and TMH2 (yellow) and HTH (pink). b) Cryo-EM structures of pleurotolysin shown in several stages of the assembly (EMD 2795, 2794, and 2793 respectively).

1.6.1 Structures of MAC assemblies

Several recent studies provide substantial new insight into assembly of the MAC. These insights were made possible by improvements in resolution (reviewed in (101–103)). These include the structures of a soluble incomplete MAC (sC5b9) of C5b-9$_1$ (79); the MAC assembled on liposomes and then solubilised (104); liposome assembled MAC (105); and the MAC analogue, polyC9 (**Chapter 3** and **Chapter 5**) (106) (**Table 1-3**).

Table 1-3 Structural assemblies of the MAC.

The structures of polyC9 (shaded rows) are the subjects of chapter 3 and chapter 5. Abbreviations: SP – single particle analysis; STA – sub-tomogram averaging.

Protein/complex	Accession code	Method/resolution	Reference
sC5b9	EMD-1991	SP Cryo-EM 24 Å	(79)
MAC MAC with masked flexible regions	EMD-3134 EMD-3135	SP Cryo-EM 8.5 Å SP Cryo-EM 7.3 Å	(104) (104)
PolyC9	EMD-3235 PDB-5FMW	SP Cryo-EM 6.7 Å Model from EMD-3235	Chapter 3 (106) Chapter 3
PolyC9 PolyC9 atomic model	EMD-7773 PDB-6DLW	SP Cryo-EM 3.9 Å Model from EMD-7773	Chapter 5 Chapter 5
MAC on liposome	EMD-3289	STA Cryo-ET 23 Å	(105)

It is now understood from recent structures of the MAC, that the cylindrical portion of the pore contains a corkscrew shape (**Figure 1-16**) (104,105). In this way, the final C9 monomer in the MAC pore does not complete the ring by contacting C5b or C6 (the first components of the assembly). Additionally, tomography of the MAC formed on liposomes confirms that only a portion of the complex inserts into the membrane to make the final pore (105). It was also observed that two or more MAC pores can concatenate to make a heterogeneous assembly, which is also postulated to occur in MAC formed with scarce amounts of C9 (55,105).
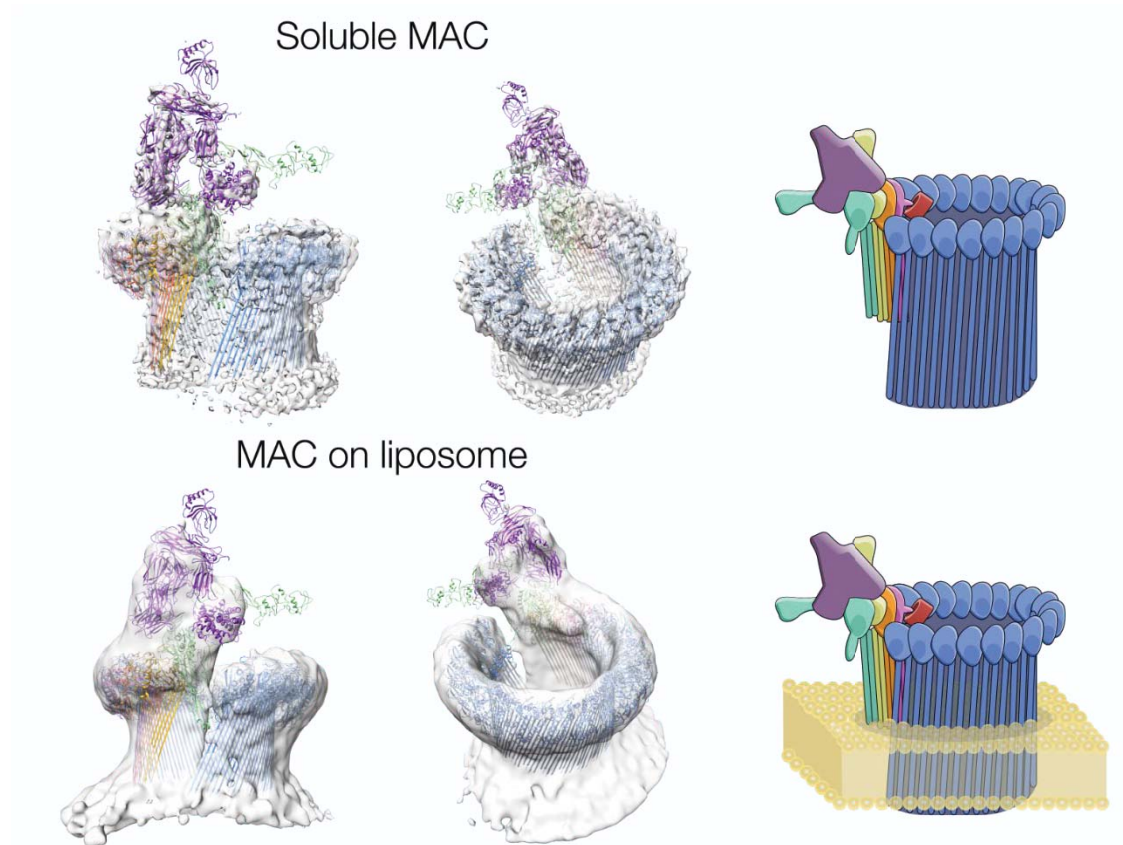
Figure 1-16 Corkscrew structures of MAC.

Top panel, the cryo-EM reconstructions of the MAC formed on liposome formed then solubilised and determined by single particle analysis (EMDB 3134). Bottom panel, the cryo-EM reconstruction of a MAC formed on liposomes determined by sub tomogram averaging (EMDB 3289). Both cryo-EM maps contain an atomic model of the MAC.

1.6.1 The MAC analogue - polyC9

Another avenue used to study MAC formation, is the structural analogue polyC9. C9 can self-polymerise into a cylindrical ring with the same features and dimensions as the MAC, but lacking the stalk unit (107–109) (**Figure 1-17**) Typically, polyC9 formation occurs at high protein concentrations (>1 mg/mL) and has a reduced haemolytic activity once formed (110). The loss in haemolytic activity of polyC9 can be attributed to its inability to insert into membranes once the irreversible polymerisation has occurred in solution. Consequently, as the proportion of polyC9 increases, there is a decrease in the amount of C9 monomers that can insert into a lytic pore.

Notably, polyC9 was shown to form in a range of conditions all of which are independent of the presence of membranes. These conditions include: incubation in low ionic buffers at 37 °C, addition of zinc, copper, or cadmium, or limited proteolysis with trypsin (110,111). PolyC9 rings have not been observed *in vivo* and it remains to be understood whether one or more specific inhibitors prevent such structures from forming in plasma (112).
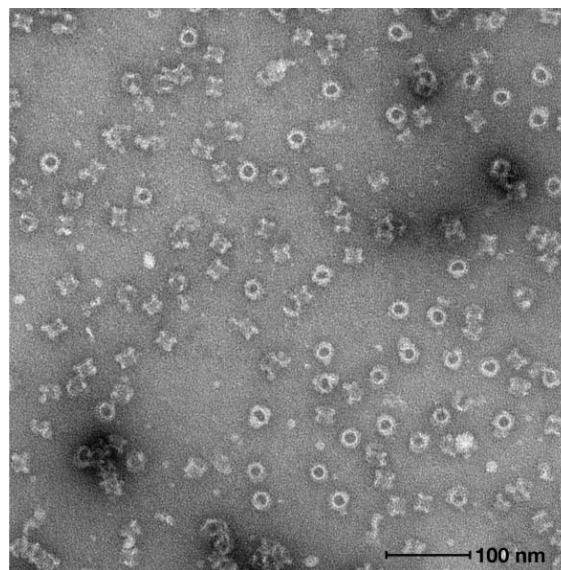
Figure 1-17 PolyC9 rings formed from purified C9 protein.

The polyC9 rings were formed by incubation at 37 °C, in a low ionic buffer. Scale bar represents 100 nm. The micrograph was acquired on a Technai T12, (Unpublished, Spicer 2015).

1.7 Aims

Understanding the mechanisms and nuances of formation of the MAC has been an area of intense investigation since its discovery in the 1950s. Numerous studies have aimed to elucidate the structural basis of assembly of the MAC, as it has implications in understanding how the immune system rapidly responds to infections and how off target effects lead to a hyper-inflammatory state. Structural research of the MAC components, including C5b and C6, suggest that these proteins undergo substantial conformational rearrangement during assembly. Accordingly, this research investigates approaches to understand this process using the following aims:

Aim 1) Determine the structure of the polyC9 as a surrogate of the MAC structure by using plasma derived human C9 (Chapter 3).

Aim 2) Develop and improve the methods of expressing various forms of C9 protein to comprehensively interrogate the structure of both C9 monomer and polyC9 (Chapter 4).

Aim 3) Determine the structure of the monomer C9 and the structure of polyC9 at improved atomic resolution (Chapter 5).

2.1 Chemicals, reagents, media and equipment

Unless stated otherwise, all solutions prepared in this study were made using ultrapure water from a Millipore system (MQ) EMD Millipore Corporation ©, (Billerica, MA, USA).

2.1.1 Chemicals

All chemicals used in this study are listed in **Table 2-1**.

Table 2-1: List of chemicals used in this study.

Abbreviations for chemicals (where appropriate) based on their respective IUPAC designation.

Abbreviation	Chemical name	Chemical provider
96% EtOH	Absolute ethanol (96%)	Merck
	40% (w/v) acrylamide/bis-acrylamide (37.5:1)	Amresco
	Agarose	Amresco
	Ammonium hydroxide	Sigma-Aldrich
APS	Ammonium persulfate	Amresco
Amp	Ampicillin	Astral
β ME	2-mercapto ethanol	Sigma-Aldrich
	Benzamidine	Sigma-Aldrich
	Bromophenol blue	Sigma-Aldrich
CAPS	3-(cyclohexylamino)-1-propanesulfonic acid	Astral
Cm	Chloramphenicol	Amresco
	D-Glucose	Amresco
DMSO	Di-methyl sulfoxide	Sigma-Aldrich
EDTA	Ethylenediaminetetra acetic acid	Amresco
	Endotoxin-Free Ultra Pure Water	Sigma-Aldrich
Gent	Gentamycin	Sigma-Aldrich
	Glacial acetic acid	Merck
	Glycerol	Amresco
HEPES	4-[2-hydroxyethyl)-1-piperazineethanesulfonic acid	Amresco
	Isopropanol	EMD Chemicals Inc.
	Imidazole	Sigma-Aldrich
	Magnesium chloride	Amresco
MeOH	Methanol	Merck
TEMED	N,N,N',N'-tetramethylethane-1,2-diamine	Sigma-Aldrich
PMSF	Phenylmethanesulfonyl fluoride	Sigma-Aldrich
PEG 3350	Polyethylene glycol (mw ~3350)	Sigma-Aldrich

PEG 4000	Polyethylene glycol (mw ~4000)	Sigma-Aldrich
Tween20	Polyoxyethylene (20) sorbitan monolaurate	Amresco
NaMalon pH 4.0	3.4 M Sodium malomate pH 4.0	Hampton
NaMalon pH 8.0	3.4 M Sodium malomate pH 8.0	Hampton
Azide	Sodium azide	Sigma-Aldrich
NaCl	Sodium chloride	Amresco
SDS	Sodium dodecyl sulphate	Sigma-Aldrich
	Sodium phosphate dibasic	Merck
	Sodium phosphate monobasic	Merck
Tet	Tetracycline	Sigma-Aldrich
TCA	Trichloroacetic acid	Sigma-Aldrich
Tris	Trisma base	Amresco
	Zinc chloride	Fluka

2.1.2 Reagents

All reagents used in this study are listed in **Table 2-2**.

Table 2-2 List of reagents used in this study.

Abbreviation	Reagent description	Reagent provider
A8-35	Amphipol A8-35	Anatrace
	Blue/Orange Loading dye, 6x	Promega
BSA	Bovine serum albumin	Sigama-Aldrich
C9-dpl	C9-depleted serum	Complement tech
anti-C9	Goat anti-human C9	Complement tech
anti-goat	Rabbit anti-goat	Millipore
dNTPS	di-nucleotide tri-phosphates (2 mM each: dATP, dCTP, dGTP, dTTP)	Sigma-Aldrich
Gelatin	Gelatin from porcine skin	Sigma-Aldrich
sRBCs	Sheep red blood cells	Monash Animal Research Platform
	SYBR Safe™ Nucleic acid stain	Invitrogen
	Human plasma	Australian Red Cross
	Cellfectin	ThermoFisher Scientific
PI tablet	cOmplete EDTA-free protease inhibitor cocktail	Roche
	Rabbit anti-sheep IgG	Rockland Immunochemicals
PEI	Polyethyleneimine	Polysciences, Inc.
Lupin	Lupin peptone	Biotech Solabia
100x GlutaMAX™	L-alanyl-L-glutamine dipeptide in 0.85% NaCl	ThermoFisher Scientific

A list of all the crystal screening kits that were trialled in this study can be found in **Table 2-3**.

Table 2-3 Crystal screening kits

Crystal screening kit	Kit provider
Ammonium sulphate screen	Hampton Research
CrystalScreen HT	Hampton Research
CrystalScreen Lite	Hampton Research
Index	Hampton Research
JBScreen HTS I	Jena Bioscience
JCSG+	Molecular Dimensions
MemGold™	Molecular Dimensions
Midas	Molecular Dimensions
MbClass Suite 1 and 2	Qiagen
Morpheus 1 and 2	Molecular Dimension
PACT suite	Qiagen
PEG/Ion screen	Hampton Research
Proplex	Molecular Dimensions
Shotgun Top 96	Molecular Dimensions
Wizard 1 and 2	Molecular Dimensions
Additive screening kit	Hampton Research

2.1.3   Enzymes

The enzymes for DNA manipulation were purchased from New England Biolabs, (Connecticut, USA) or Invitrogen (California, USA). The following restriction endonucleases were used for cloning purposes: *Sfi*I, *Not*I, *Eco*RI. Antarctic phosphatase (New England Biolab) was used to dephosphorylate restriction digest products, and T4 ligase (Promega) was used for DNA ligation. The PCR mutagenesis was performed using KOD DNA polymerase (Sigma-Aldrich).

DNA sequencing was performed by dye terminator sequencing reactions. Big dye terminators (BDT), containing dNTPs, terminating nucleotides and polymerase were purchased from Micromon DNA sequencing platform (Monash University). The purified and concentrated BDT reaction products were submitted to the Micromon facility for capillary analysis.

2.1.4 Protein purification equipment

Protein purifications columns used in this study are listed in **Table 2-4**. A detailed description for each of the protein purification techniques can be found in section 2.7. Protein purification techniques were performed using an ÄKTA purifier (GE Healthcare Life Sciences). The HiTrap columns were loaded using a peristaltic pump (P1 or P3 (Pharmacia)). Size exclusion columns were purchased as pre-packed columns of the following two dimensions (bed diameter x height): 16 mm x 600 mm (column volume = 120 mL) and 26 mm x 600 mm (column volume = 320 mL) (GE Healthcare Life Sciences).

Table 2-4 Protein purification columns used in this study.

Abbreviation	Chemistry/Support resin	Column provider
CHT type I	Ceramic hydroxyapatite	Bio-Rad
DEAE	Diethyleaminoethyl sepharose	GE Healthcare Life Science
Lysine	Lysine sepharose 4b	GE Healthcare Life Science
S200	Superdex 200	GE Healthcare Life Science
MonoQ	Quaternary amine	GE Healthcare Life Science
Ni-NTA	Nitrilotriactic acid	Qiagen

2.1.5 Electron microscopy equipment

For negative stain TEM imaging, either carbon coated grids containing a layer of formvar with a copper support were used (ProSciTech) or carbon only grids on a copper support were used (Agar Scientific). Several grids were tested for cryo-EM studies, however the most suitable ones for high resolution imaging were R1.2/1.3 (Quantifoil) (1.2 μm hole size/ 1.3 μm hole period). A complete description of the EM techniques can be found in section 2.8.

The initial assessment of TEM sample quality used a T12 (Thermofisher) microscope with a LaB_6 filament accelerated to 120 keV and equipped with an Eagle 4k CCD detector (FEI Thermo). High resolution data was collected using a Titan Krios, which

40

contained a field emission gun (FEG) accelerated to 300 keV and a GIF quantum energy filter (Thermofisher) positioned between the stage and detector. Images were acquired using a K2 Summit direct electron detector (Gatan) (See also Chapter 5 Materials and Methods).

2.2 Plasmids, oligonucleotides and strains

A list of the plasmids used in this study is described in **Table 2-5**, and the DNA oligonucleotides can be found in **Table 2-6**. The parent plasmid for baculovirus expressed human C9 was pFastbac1 (Thermofisher Scientific). The parent plasmids used for mammalian cell expression of human C9 was pSectag2A (Thermofisher Scientific). The mammalian expression system codon optimised mouse C9 gene was synthesised in the vector pCDNA3.1 hygro+ (GenScript). All plasmids used in this study contain ampicillin resistance genes.

Table 2-5. List of plasmids used in this study.

✳*The honeybee melittin (HBM) signal sequence (MKFLVNVALVFMVVYISYIYA) was used in insect expression.*
✚*The IgK signal peptide (METDTLLLWVLLLWVPGSTGD) used in HEK293F cell expression.*

Species	Construct name (mutations)	Description	Parent vector	Expression system
Human	$C9_{[bac]}$	Wild-type C9 containing ✳	pFastbac1	Insect cells
	$C9_{[N\text{-his bac}]}$	C9 with N-terminal 6x his tagged	$C9_{[bac]}$ pFastbac1	Insect cells
	$C9_{[aglyco]}$ (T258M/T396M)	Double glycosylation mutant C9 with His tagged	$C9_{[N\text{-his bac}]}$ pFastbac1	Insect cells
	$C9_{[HEK]}$	Wild-typeC9 containing ✚	pSectag2a	HEK293F cells
	C9 (N256D)	Single glycosylation mutant	$C9_{[HEK]}$ pSectag2a	HEK293F cells
	C9 (T258M)	Single glycosylation mutant	$C9_{[HEK]}$ pSectag2a	HEK293F cells
	C9 (N394D)	Single glycosylation mutant	$C9_{[HEK]}$ pSectag2a	HEK293F cells
	C9 (T396M)	Single glycosylation mutant	$C9_{[HEK]}$ pSectag2a	HEK293F cells
	$C9_{[aglyco-1]}$ (T258M/T396M)	Double glycosylation mutant	C9 (N256D) pSectag2A	HEK293F cells
	$C9_{[aglyco-2]}$ (N256D/N394D)	Double glycosylation mutant	C9 (T258M) pSectag2A	HEK293F cells
	$C9_{[SERp]}$ ($_{127}$EESE$_{130}$➔AASA)	Surface entropy reduction mutant	$C9_{[HEK]}$ pSectag2A	HEK293F cells
	$C9_{[\Delta N]}$ (Δ1-15)	N-terminal truncation mutant	$C9_{[HEK]}$ pSectag2a	HEK293F cells
	$C9_{[\Delta C]}$ (Δ527-538)	C-terminal truncation mutant	$C9_{[HEK]}$ pSectag2a	HEK293F cells
	$C9_{[\Delta N/\Delta C]}$ (Δ1-15; 527-538)	N-/C-terminal truncated C9	$C9_{[HEK]}$ pSectag2a	HEK293F cells
	$C9_{[aglyco\ \Delta N/\Delta C]}$ (N256D/N394D Δ1-15; 527-538)	Double glycosylation; N-/C-terminal truncation mutant	$C9_{[aglyco-2]}$ pSectag2a	HEK293F cells
	C9 (T258M/T396M Δ1-15; 527-538)	Double glycosylation; N-/C-terminal truncation mutant	$C9_{[aglyco-1]}$ pSectag2a	HEK293F cells
Murine	$C9_{[murine]}$	Wild-type C9 ✚	pCDNA3.1 hygro+	HEK293F cells
	$C9_{[aglyco\ murine]}$ (N28E/N243D/N397D)	Triple C9 glycosylation mutant ✚	pCDNA3.1 hygro+	HEK293F cells
	$C9_{[aglyco\ \Delta N\ murine]}$ (N28E/N243D/N397D Δ1-15)	Triple glycosylation mutant C9 with N-terminal truncation	$C9_{[aglyco\ murine]}$ pCDNA3.1 hygro+	HEK293F cells

Table 2-6. List of oligonucleotides used in this study.

The oligonucleotides are listed from 5'-3'. The underlined nucleotides represent codons that were mutated from the C9 sequence to obtain the desired mutations.

Oligo	Base pair sequence	Description
oBS003	CATATTCCAAAAATGAA<u>ATG</u>TACCAACTATTTTTGTC	T258M forward human C9
oBS004	GACAAAAATAGTTGGTA<u>CAT</u>TTCATTTTTGGAATATG	T258M revers human C9
oBS005	GAGAGGGTAGAGCTGTAAACATC<u>ATG</u>AGTGAAAACCTCATAG	T396M forward human C9
oBS006	CTATGACGTTTTCACT<u>CAT</u>GATGTTTACAGCTCTACCCTCTC	T396M reverse human C9
oBS033	GACAGAGTGGTA<u>GCAGCG</u>TCT<u>GCG</u>CTGGCACGAACAG	[127]EESE[130]→AASA forward human C9
oBS034	CTGTTCGTGCCAGC<u>CGC</u>AGA<u>CGCTGC</u>TACCACTCTGTC	[127]EESE[130]→AASA reverse human C9
oBS069	GAGGCGGCCCAGCCGGCCTCTGCATCACACATAGACTGC	Δ1-15 truncation forward human C9
oBS086	GAGGCGGCCCAGCCGGCCCAGTACACGACCAGTTATGACCC	Cloning forward human C9
oBS087	AGAAATTTTTTGTTTACTGCGAAAGCGGCCGCTA	Δ528-538 truncation reverse human C9
oBS101	ACTTACCAACTA<u>TGT</u>TTGTCATATTCTTCAAAG	F262C forward human C9
oBS102	GAAGAATATGACAA<u>ACA</u>TAGTTGGTAAGTTTC	F262C reverse human C9
oBS105	CATAGATGATGTT<u>TGT</u>TCACTCATAAGAGGTGG	V405C forward human C9
oBS106	CTCTTATGAGTGA<u>ACA</u>AACATCATCTATGAG	V405C reverse human C9
oBS148	CATATTCCAAA<u>GAT</u>GAAACTTACCAAC	N256D forward human C9
oBS149	GTTGGTAAGTTTC<u>ATC</u>TTTGGAATATG	N256D revers human C9
oBS150	GGTAGAGCTGTA<u>GAC</u>ATCACCAGTGAAAACCTC	N394D forward human C9
oBS151	GTTTTCACTGGTGAT<u>GTC</u>TACAGCTCTACCCTC	N394D revers human C9
oBS173	CAACCGGCGATTACCCCATCCCTATCGACTG	Δ1-14 truncation forward mouse C9
oBS174	GGGATGGGGTAATCGCCGGTTGACCCAGGG	Δ1-14 truncation reverse mouse C9

Bacteria strains used in this study:

DH5α™: a K-12 derived strain of *E. coli* used for plasmid DNA amplification and subsequent purification, manipulation techniques and large scale transfections for transient mammalian expression; genotype – F- *fhu*A2, *lac* ΔU169, *pho*A, *gln*V44, φ80*lac*Z ΔM15, *gyr*A96, *rec*A1, *rel*A1, *end*A1, *thi*-1, *hsd*R17

Top10™: a strain of *E. coli* used for DNA amplification and subsequent purification. F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80lacZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara, leu*)7697 *gal*U *gal*K *rps*L (*Str*R) *end*A1 *nup*G

DH10bac™: a strain of E. coli used for production of recombinant bacmid DNA. F- mcrA Δ(*mrr-hsdRMS-mcrBC*) phi80*lac*ZΔM15 Δ*lac*X74 recA1 *end*A1 *ara*D139 Δ(*ara, leu*) 7697 *gal*U *gal*K λ- *rps*L *nup*G/ pMON14272/pMON7124

2.3 General methods

2.3.1 2xYT Medium

The 2xYT media was prepared in double-distilled water (ddH$_2$O) with the following final concentrations of solutes: 0.5% (w/v) NaCl, 1.0% (w/v) Bacto-yeast extract, and 1.6% (w/v) Bacto-tryptone. The media was stirred until all the media reagents dissolved, and then the volume adjusted to the final volume. The media was autoclaved for sterilisation and cooled to room temperature prior to use.

2.3.2 SOC Medium

The SOC medium was prepared in ddH$_2$O with the following final concentrations of solutes: 2% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl$_2$, 10 mM MgSO$_4$, and 20 mM D-glucose. After dissolving the media, the pH is adjusted to 7.0 with NaOH and filter sterilised with a 0.22 μm pore size filter.

2.3.3 Agar plates

Agar plates were prepared by adding 1.5% (w/v) of bacteriological agar to LB media containing 1% (w/v) tryptone, 0.5% (w/v) yeast extract and 1% (w/v) NaCl adjusted to pH 7.0 with NaOH. The agar was sterilised by autoclaving at 121 ºC and 100 kPa for 20 min, after which it was left to cool to below 50 °C. After cooling, the appropriate antibiotics or additives were added, and the agar poured into 10 cm diameter petri plates and allowed to cool until the agar had solidified.

2.3.4 Preparation of competent cells

Laboratory stocks of chemically competent cells used in all experiments were prepared in-house using methods adapted from the rubidium chloride competent cell

44

protocol (113). Briefly, *E. coli* cells were cultured in the appropriate media overnight at 37 °C with shaking, after which the stationary phase cells were sub-cultured into fresh growth media. The cells were grown to an optical density of 0.2 AU (OD 600 nm) and incubated at 4 °C for 10 min. The cells were harvested by centrifugation and incubated in ice-cold 0.1 M CaCl$_2$ for 30 min. The cells were harvested again by centrifugation and the pellet resuspended in ice-cold 0.1 M CaCl$_2$, after which time the cells were competent for DNA transformations.

2.3.5   Transformation of competent cells

Approximately 5-100 ng of plasmid DNA was mixed with 100 μL of competent *E. coli* cells and incubated at 4 °C for 30 min. The cells were heat-shocked at 42 °C for 45 sec before being incubated at 4 °C for another 2 min. Next, the cells were recovered by addition of 500 μL of SOC media. The DH5α or TOP10 cells were then incubated for 1 hour at 37 °C, whereas DH10bac were incubated at 37 °C for 5 hours. The cells were then used to inoculate liquid media or plated onto LB agar (containing antibiotics) where appropriate.

2.3.6   DNA purification techniques

All DNA was purified in endotoxin free ultra-pure water (Sigma-Aldrich) and stored either at 4 °C or -20 °C for long term storage.

For small scale purification of plasmid DNA, transformed DH5α or Top10 *E. coli* containing the desired plasmids were inoculated into 3 mL 2xYT broth with the appropriate antibiotics for selection of the plasmid. The cultures were incubated in a shaking 37 °C incubator, at 220 rpm O/N until cells had grown to stationary phase (14-18 h). The DNA was extracted and purified from the overnight cultures using the Wizard

Plus SV Miniprep DNA kit (Promega) according to the manufacturer's instructions. Bacmid DNA purifications from transformed DH10bac cells containing the desired recombinant bacmids were performed identically to the small-scale plasmid preps.

Large scale plasmid DNA purifications (for HEK293F cell transfections) were performed using either a Plasmid Maxi kit (Qiagen) or a plasmid Giga kit (Qiagen). Maxi DNA purifications were performed by inoculating 100 mL of 2xYT media plus antibiotics, with DH5α or Top10 cells containing the desired plasmid and incubating overnight at 37 °C, 220 rpm. The following day, cells were pelleted at 3220 rcf, 4 °C for 20 min and the DNA was extracted and purified according to the manufacturer's high yield protocol (Qiagen). The Plasmid Giga DNA purification was performed by inoculating 2 x 3 mL of 2xYT broth with appropriate antibiotics with a single transformed colony from an LB Agar selection plate and incubating overnight at 37 °C, 220 rpm (Day 1). The overnight cultures were used as the inoculum (Day 2) to scale up *E. coli* cells into 2 x 1 L of 2xYT broth with appropriate antibiotics, which were incubated for 12-16 h. On Day 3, the cells were centrifuged in a JLA 8.1 rotor centrifuge, 3993 rcf, 15 min 4 °C. The cell pellets were used for the DNA extraction and purification as per the manufacturer's instructions.

2.3.7   *Agarose gel electrophoresis*

Agarose was added to Tris acetate EDTA buffer (TAE) (40 mM Tris acetate, 1 mM EDTA, pH 8.0) at the appropriate percentage (w/v). The solution was heated until the agarose was completely dissolved, and the volume readjusted with ddH$_2$O. Then, SYBR Safe was added at a dilution of 1: 10,000 to allow the visualisation of DNA under UV light. The agarose solution was then poured into the appropriately sized tray containing a 10- or 15-tooth gel comb and left until completely solidified. The DNA gels were

resolved in TAE buffer in a Bio-Rad Mini-Sub Cell GT system. DNA samples were prepared by addition of blue/orange 6x loading dye (Promega) and electrophoresed at 120 V. The DNA was detected under UV-trans-illumination by the fluorescence of the DNA-bound SYBR Safe.

2.3.8 DNA Sequencing

DNA sequencing was performed using the Micromon Monash Research Technology Platform (Monash University, Australia) using an Applied Biosystem 3730S Genetic Analyser. The DNA sequencing reactions were prepared using the Applied Biosystems BigDye Terminator (BDT) mix. A 20 μL mixture containing 40 ng of DNA was combined in ddH$_2$O with BDT buffer, BigDye Premix (v3.1) and 5 pmoles of appropriate sequencing primer. The following thermal cycle for DNA sequencing was performed on a Bio-Rad MyCycler thermal cycler:

Initial denaturation	90 °C, 1 min	
Denaturation	90 °C, 1 min	
Annealing	50 °C, 10 sec	} Repeat for 25 cycles
Extension	60 °C, 4 min	
Final extension	60 °C, 4 min	

The BDT products were precipitated in 100 mM sodium acetate and 70% (v/v) ethanol (EtOH) at 4 °C for 20 min. The DNA was pelleted by centrifugation at 16,100 rcf for 20 min, and subsequently washed in 70% (v/v) EtOH. The EtOH was removed by aspiration and residual EtOH removed by heating at 95 °C for 1 min.

2.4 DNA Cloning and mutagenesis techniques

The cloning of various constructs for recombinant expression was performed by standard molecular biology techniques: including restriction digestion and DNA ligation and transformation into host *E. coli* cells. Mutagenesis of C9 was performed by

Quikchange mutagenesis. Briefly, the desired mutations were engineered into DNA primers, which were used in PCRs containing the target plasmid and standard PCR reagents. In these PCRs, the entire length of the plasmids (6-7 kb) were amplified. Subsequently, the plasmid mixture was digested with *Dpn*I to specifically digest methylated parental plasmid DNA. The plasmids were purified using a Wizard DNA Clean-Up kit according to the manufacturer's instructions (Promega). The undigested plasmids were transformed into competent DH5α or Top10 cells and plated onto LB agar containing ampicillin. Single colonies from the transformations were then re-streaked to ensure single isolates were obtained. All the plasmids produced were sequenced to ensure the correct sequences were obtained for the coding region of the gene (section 2.3.8.).

2.5   General Protein techniques

2.5.1   TCA precipitation

Expression media containing secreted C9 protein was obtained by centrifugation of cell culture at 2000 rcf for 5 min. The supernatant (500 µL) was transferred to an Eppendorf tube and 50% (w/v) trichloroacetic acid (TCA; 125 µL) was added to a final concentration of 10% (w/v). The precipitant was incubated on ice for 10 min and then centrifuged at 16,1000 rcf for 10 min and the supernatant was decanted. Protein pellets were washed twice by centrifugation (16,100 rcf, 5 min) with 500 µL cold acetone and then the supernatant was aspirated. After the final wash, the pellets were centrifuged to remove excess TCA and then boiled in 50 µL 1x LSB for 10 min prior to loading into SDS-PAGE.

2.5.2   SDS-Polyacrylamide Gel Electrophoresis (PAGE)

Stock solutions for SDS-PAGE:

4x Resolving gel buffer	1.5 M Tris · HCl, pH 8.8, 0.4% (w/v) SDS
4x Stacking gel buffer	0.5 M Tris · HCl, pH 6.8, 0.4% (w/v) SDS
Reservoir buffer	25 mM Tris HCl, 192 mM glycine, 0.1% (w/v) SDS, pH 8.3
5x Laemmli Sample Buffer (LSB)	50% (w/v) glycerol, 5% (w/v) SDS, 0.05% (w/v) Bromophenol blue, 300 mM Tris HCl pH 6.8
APS solution	20% (w/v) ammonium persulfate
Coomassie blue staining solution	0.025% (w/v) Coomassie brilliant blue R250, 40% (v/v) methanol, 7% (v/v) acetic acid
Coomassie de-stain solution	40% (v/v) methanol, 7% (v/v) acetic acid

Table 2-7 Preparation of a 12% SDS-Polyacrylamide Gel.

	Resolving gel	Stacking gel
40% (w/v) Bis-Acrylamide (mL)	1.5	0.375
4x Resolving buffer (mL)	1.25	-
4x Stacking buffer (mL)	-	0.75
ddH$_2$O (mL)	2.25	1.875
APS solution (μL)	50	30
TEMED (μL)	5.0	5.0

Typically, 12% (w/v) bis-acrylamide SDS gels were prepared according to **Table 2-7**. The protein purification samples (10 μL) were combined with 5x non-reducing LSB (2.5 μL) and boiled for 10 min, then pipetted into the wells of the stacking gel. A Precision Plus Protein Dual (Bio-Rad) protein standard was loaded on all SDS-PAGE to aid in protein identification. SDS-PAGE were electrophoresed in SDS reservoir buffer at 300 V for 30 min using a Bio-Rad minigel system. Proteins were visualised by either incubating with Coomassie blue staining solution or the gel was used to perform Western blotting analysis.

2.5.3 Western blot

Western blotting buffers:

CAPS Transfer buffer:	10 mM CAPS pH 11, 15% (v/v) methanol
Tris-buffered Saline (TBS):	20 mM Tris HCl, 200 mM NaCl, pH 7.4
Tris-buffered saline-Tween (TBST):	TBS containing 0.5% (v/v) Tween20
Blocking buffer:	TBST containing 5% (w/v) low-fat milk powder

The SDS-PAGE was resolved by electrophoresis at 300 V for 30 min and then the SDS-polyacrylamide gel sandwiched between the nitrocellulose membrane, Whatman filter paper, and two sponges that were pre-soaked in CAPS transfer buffer. The western sandwich was horizontally transferred using a Mini-Trans Blot Transfer Cell (Bio-Rad) in CAPS transfer buffer at 100 V for 60 min. During the transfer, the CAPS buffer was stirred and kept cold with ice as per the manufacturer's instructions. Following transfer, the nitrocellulose membrane was blocked with blocking buffer for 30 min, after which the membrane was probed with a dilution of primary antibody prepared in blocking buffer for 1 hour. The probed membrane was then washed 3 x 10 min with blocking buffer, and then probed with a dilution of HRP-conjugated secondary antibody prepared in blocking buffer for 1 hour. The membrane was then washed with TBST 4 x 5 min.

For detection of probed membranes, a chemiluminescent kit (GE Healthcare) was used. Equal volumes of detection reagent 1 and 2 were combined and immediately used to coat the membrane. Then, the membrane was exposed to CL-film in a darkroom and developed in an automatic Fujifilm processor (Minato, Tokyo, Japan).

2.5.4 Protein concentrating and buffer exchange

A 30 kDa molecular weight cut off (MWCO) protein concentrator was used to concentrate purified protein samples according to the manufacturer's instructions (EMD Millipore Corporation©, Billerica, MA, USA). The concentrators were also used to

50

buffer exchange proteins by repeated concentration and dilution of protein samples in the desired buffer.

2.5.5 Protein concentration determination

Protein concentrations were determined throughout the study by absorbance at 280 nm or by Bradford assay (Bio-Rad) using BSA protein standards. Extinction coefficients (Σ) for all C9 proteins were calculated by the following formula (114):

Σ = Number (Tyrosine) Ext(Tyrosine) + Number (Tryptophan) Ext(Tryptophan) + Number of (Cysteine) Ext(Cysteine)

Where Ext(Tyrosine) = 1490; Ext(Tryptophan) = 5500 and Ext(Cysteine) = 125

The Bradford assays were performed as per the manufacturer's instructions, using BSA as a protein standard (115). The values from both empirical and calculated quantification technique correlated well for all C9 proteins.

2.5.6 Titration based haemolytic assay of C9

Buffers used for haemolytic assays of C9:

DHB++	2.5% (w/v) D-glucose, 5 mM HEPES, 71 mM NaCl, 0.15 mM CaCl$_2$, 0.5 mM MgCl$_2$, pH 7.4
Human C9 Gel filtration buffer	10 mM HEPES pH 7.2, 200 mM NaCl
Mouse C9 Gel filtration buffer	10 mM HEPES pH 7.2, 100 mM NaCl

Sheep red blood cells (sRBC) were maintained in Celpresol™ at 4 °C for normal long term storage. Prior to assays, the sRBCs were washed by adding 0.5 mL cells to 2.5 mL of ice cold DHB++ and centrifuged at 3220 rcf for 2 min. The supernatant was decanted and the sRBCs resuspended with 2.5 mL of DHB++ and washed two more times. 6.5 x 10^8 cells/mL of sRBCs were sensitised by addition of anti-sheep antibody (0.75 mg/mL) at 30 °C. The cells were washed to remove excess antibody (3220 rcf, 2 min) and resuspended with fresh DHB++. Three independently prepared dilutions of C9 protein

(32 µg/mL) were made in the appropriate gel filtration buffer for each replicate, and 2-fold serial dilutions of protein were prepared. Reactions were performed in a final volume of 220 µL by addition of 3.75 x 10^6 sensitised cells to PCR strip tubes containing the dilutions of C9 (15 µL) and C9-depleted serum (1 µL). The reactions were immediately incubated on a PCR thermocycler block heated to 37 °C for 30 min. The assays were stopped by cooling to 4 °C and promptly centrifuged on a benchtop centrifuge for 20 sec. Supernatant from the assays (150 µL) was transferred to a 96-well round bottom plate and absorbance at 405 nm was measured.

All statistical measurements were performed in GraphPad Prism 7.0b under an education license provided by Monash University. The protein concentrations were transformed from [pMol] to a logarithmic scale with the equation: X=Log(X); and the absorbance data were normalised. A 0% untreated control and 100% containing the two component toxins (PlyA/B) was used. The graphs are reported as the raw unfitted curves with error reported as the standard error of the mean (SEM).

2.5.7 Quick haemolytic assay of C9

A quick haemolytic assay was developed as an alternative to the more time-consuming assays described in section 2.5.6. The sRBCs were washed and sensitised as described above. A concentration of serum derived C9 equivalent to ~90% lytic activity over 30 minutes was determined. A single dilution of recombinant C9 with an equivalent concentration to the 90% serum C9 activity was used to assess if the recombinant proteins were active.

2.6 Cell culture techniques

Recombinant C9 protein was produced from two sources: either baculovirus expression system (insect cells; Section 2.6.1 - 2.6.4) or mammalian expression system (HEK293F cell; Section 2.6.5 - 2.6.6).

2.6.1 Generation of P1 baculovirus

Baculoviruses containing the C9 gene were generated using the Bac-to-Bac™ (Invitrogen) method according to the manufacturer's instructions. Sf9 cells were diluted with Insect Xpress media (Lonza BioWhittaker) to a density of 8×10^5 cells/mL the day prior to transfections, and typically doubled in density over a period of 24-hours. A volume containing 2.0×10^6 cells was plated onto a Falcon 6-well tissue culture plate and incubated at 26 °C for 30-45 minutes. The recombinant bacmid DNA (1 μg) was added to 100 μL of insect media and combined with an equal volume of media containing Cellfectin II (16 μl) then allowed to incubate at room temperature for 45 min. Cells were washed twice after adherence to the plastic tissue culture plate, with insect media containing no gentamycin (which can interfere with the transfection efficiency). A prepared DNA/Cellfectin II mixture was added to the top of the cells and incubated in a stationary incubator at 26 °C for 5 h. Following transfection, the media containing the Cellfectin II was removed and replaced with fresh Insect Xpress media containing gentamycin. The cells were incubated for 4 days at 26 °C to generate the P1 baculovirus.

2.6.2 Generation of P2 and P3 baculovirus

After generation of the P1 virus, the second generation of viral stock was prepared (P2) typically by addition of 1 mL P1 stock to 50 mL of suspension growth Sf9 cells at a cell density of 1.0×10^6 cells/mL. Infected cells were incubated for 3 days at 27 °C at

110 rpm, then centrifuged at 600 rcf, RT 5 min. The supernatant containing the baculovirus was 0.45 μm filtered with a syringe filter to exclude any cells still in the supernatant. The P2 stock was stored at 4 °C. Baculovirus was typically propagated for a third generation (P3) in a larger volume (200 mL) using the method described for P2 stocks. These P3 stocks were used for large scale infections for protein purification.

2.6.3   Infection of insect cells

Large scale infections for protein purification of baculovirus expressed C9 were performed using Hi5 cells from the cabbage looper (*Tricopolusei ni*), as they have been demonstrated to be better for large scale secretion of protein. The baculovirus expressed protein contains an N-terminal secretion sequence (<u>H</u>oney <u>B</u>ee <u>M</u>ellitin [HBM]) for improved targeting to the ER for the secretion. The HBM sequence replaced the native secretion sequence for human C9. Large scale infections were performed in volumes ranging from 1 to 4 liters of Hi5 cells adapted to Insect Xpress media. The day prior to infections, the cells were diluted to a typical density of $5.0 – 8.0 \times 10^5$ cells/mL. The cells typically doubled in cell density over a period of 18 h incubation at 27 °C at 110 rpm. 1.0×10^6 cell/mL cells were infected with 20 mL of P3 stock per liter of cell culture and incubated at 27 °C, 110 rpm in a shaking incubator for 3 days. This optimized volume of P3 infection was determined by monitoring protein secretion in small scale experiments.

2.6.4   Acquiring insect media and dialysis

After the 3-day infection, the insect cells were separated from the conditioned media by centrifugation at 2000 rcf for 15 m at 4 °C using a JLA 8.1 rotor. The conditioned media, containing the secreted recombinant protein, was sequentially filtered using a vacuum trap filter of consecutively smaller pore sizes; i.e. from 0.8 μm then to 0.45 μm

54

and finally 0.22 μm. The filtered media was then dialyzed in buckets containing 20 L of 10 mM Na_2HPO_4/NaH_2PO_4 pH 7.4, 45 mM NaCl using a 28.6 mm width dialysis tubing. Several changes of dialysis buffer were performed over the course of 2 days and then the media was passed over chromotography columns such as DEAE (section 2.7.2) or Ni-NTA (section 2.7.5).

2.6.5 Mammalian C9 expression

Stock mammalian HEK293F cells were passaged and stored at Monash Protein Production Unit. All cell incubations were performed at 37 °C in a 5% (v/v) CO_2 shaking incubator, at 110-140 rpm. Transient transfections of plasmids with the C9 gene were performed over a 4-days period. The day before transfections, the HEK293F cells were split to a cell density of $1.0 - 1.2 \times 10^6$ cell/mL in 90% of the final transfection volume. The cultures were incubated for a period of 24 h at 37 °C whereupon they double in cell density. On the day of transfections, PBS equal to 10% of the final transfection volume was pre-warmed to room temperature, mixed with DNA (1 ug/mL of final transfection volume) and PEI (4 μg/ 1 μg DNA) in a Falcon tube. The DNA/PEI suspension was vortexed and allowed to incubate at room temp for $15 - 20$ min, and then added dropwise with constant swirling to the HEK293F cells. Finally, the glucose concentration of the HEK293F cell culture was analyzed using an Accu check meter and adjusted to 33 mM before placing in the shaking CO_2 incubator. After 18 hours, the cells were supplemented with lupin to a final concentration of 5 mg/mL and the cells were culture was continued to a total of 4 days.

2.6.6   Acquiring HEK293F media

After transfection for 4 days, HEK293F cell suspensions were centrifuged at 2000 rcf, 20 min 4 °C. The conditioned media, which contained secreted C9, was vacuum trap filtered through consecutively smaller pore sizes from 0.8 µm, 0.45 µm, and 0.22 µm. One aliquot of *cOmplete* EDTA free protease inhibitor was added per 100 mL of filtered media and the protein purified according to the native purification method in section 2.7.2.

2.7   Protein purification methods

Plasma-derived C9 was purified from expired apheresis, human plasma provided by the Australian Red Cross (section 2.7.1). Human plasma was stored at -80 °C until required for purification. Recombinant protein purification was from insect cells or mammalian cells (HEK293F) (sections 2.7.2 - 2.7.5). The various protein purification columns are listed in **Table 2-4**.

2.7.1   Purification of Plasma C9

Plasma C9 purification buffers:

PEG phosphate buffer	50 mM Na_2HPO_4/NaH_2PO_4 pH 7.4, 150 mM NaCl, 15 mM EDTA, 18% (w/v) PEG 4000
DEAE buffer A:	10 mM Na_2HPO_4/NaH_2PO_4 pH 7.4, 45 mM NaCl, 10 mM EDTA
DEAE buffer B:	10 mM Na_2HPO_4/NaH_2PO_4 pH 7.4, 500 mM NaCl, 10 mM EDTA
CHT Phosphate buffer:	10 mM Na_2HPO_4/NaH_2PO_4 pH 7.0
CHT wash buffer	10 mM Na_2HPO_4/NaH_2PO_4 pH 7.0, 100 mM NaCl
CHT buffer A:	45 mM Na_2HPO_4/NaH_2PO_4 pH 8.1
CHT buffer B:	350 mM Na_2HPO_4/NaH_2PO_4 pH 8.1
Gel filtration buffer:	10 mM HEPES pH 7.2, 200 mM NaCl, 10 mM EDTA

The purification of human plasma C9 followed previously published methods (116,117). On the first day of purification, the frozen plasma was thawed in a water bath at 37 °C until completely thawed. To this thawed plasma, a final concentration of 1 mM PMSF (made in suspension by 100% (v/v) isopropanol), 1 mM benzamidine, 10 mM EDTA and one *cOmplete* EDTA free protease inhibitor tablet (Roche) per 100 mL of plasma was added. Next the plasma was diluted with 0.4 x plasma volume of ultrapure water at 4 °C.

Initial purification was performed by PEG precipitation between 6% - 20% (w/v). In the first step of this precipitation, a volume of the PEG phosphate buffer equal to 1/2 x volume of the plasma suspension was added dropwise to a stirred solution to make a final percentage of PEG 6% (w/v) at 4 °C, and allowed to stir for 30 min. Following mixing of this suspension, the precipitated material was separated from the supernatant using centrifugation (16,000 rcf, 60 min, 4 °C, JA 14 rotor). The supernatant (containing C9 and other proteins) was then filtered by consecutively smaller pore size starting with a coffee filter, then 0.8 μm, 0.45 μm and 0.22 μm.

The filtered supernatant then underwent a second round of precipitation. Granular PEG 4000 was added to a final concentration equal to 20% (w/v) of the supernatant starting volume, at 4 °C, while stirring on a magnetic stir plate for 30 min. This suspension was centrifuged in a JLA 8.1 rotor (1 L bottles) at 2,500 rcf for 15 min, and the supernatant discarded. The pellet, containing C9 protein, was suspended with DEAE buffer A using 200 mL buffer per pellet.

The protein suspension was passed over a lysine resin which was loosely packed (ID 2.5 cm x 4 cm) for the specific removal of the protease plasminogen. The flow through was then passed over a DEAE column (ID 5 cm x 4 cm) equilibrated in DEAE buffer A. After binding overnight to DEAE, the resin was washed with 1 L DEAE buffer A, then

the protein eluted over a 6 CV linear gradient with DEAE buffer B. Fractions containing C9 from DEAE were determined by 12% (w/v) SDS-PAGE using either Coomassie stain or western blot analysis.

The pooled fractions from the DEAE chromatography step were prepared for the ceramic hydroxyapatite (CHT type I) step. To lower the conductivity of the DEAE fractions, CHT phosphate buffer was added to the pooled fractions while mixing, and the conductivity adjusted to below 8 mS/cm (the same as the CHT wash buffer). Following this adjustment, loading of C9-containing fractions to 10 mL CHT was performed in 10 mM Na_2HPO_4/NaH_2PO_4 pH 7.0, 100 mM NaCl (CHT wash buffer) via a peristaltic pump. The CHT resin was then washed with 4 CV of CHT wash buffer. After washing, the column was attached to an ÄKTA FPLC and washed with 4 CV CHT buffer A, then eluted over a 10 CV linear gradient with CHT buffer B. Fractions containing C9 from CHT were determined by using both Coomassie stained SDS-PAGE and western blot analysis.

The pooled CHT fractions were added to an equal volume of size exclusion buffer. Then the diluted protein was concentrated in a 30 kDa MWCO protein concentrator (Millipore) using centrifugation at 3220 rcf at 4 °C. Depending on the purity and yield of C9 from CHT, either an S200 16 mm x 600 mm or an S200 26 mm x 600 mm column was used equilibrated in gel filtration buffer. The injected concentrated protein was typically 1-3% of the total column volume.

2.7.2   Recombinant C9 purification by Native purification method

Purification buffers for recombinant C9 (<u>Note, these buffers did not contain EDTA</u>)

DEAE phosphate buffer:	10 mM Na_2HPO_4/NaH_2PO_4 pH 7.4, 20 mM NaCl
DEAE buffer A:	10 mM Na_2HPO_4/NaH_2PO_4 pH 7.4, 45 mM NaCl
DEAE buffer B:	10 mM Na_2HPO_4/NaH_2PO_4 pH 7.4, 500 mM NaCl
CHT Phosphate buffer:	10 mM Na_2HPO_4/NaH_2PO_4 pH 7.0
CHT wash buffer	10 mM Na_2HPO_4/NaH_2PO_4 pH 7.0, 100 mM NaCl
CHT buffer A:	45 mM Na_2HPO_4/NaH_2PO_4 pH 8.1
CHT buffer B:	350 mM Na_2HPO_4/NaH_2PO_4 pH 8.1
Human gel filtration buffer:	10 mM HEPES pH 7.2, 200 mM NaCl
Mouse gel filtration buffer:	10 mM HEPES pH 7.2, 100 mM NaCl
MonoQ buffer A:	10 mM Tris HCl pH 7.5, 50 mM NaCl
MonoQ buffer B:	10 mM Tris HCl pH 7.5, 350 mM NaCl

The native purification method for recombinant expressed C9 was similar to the plasma C9 purifications described in (Section 2.7.1). For mammalian expressed C9, the media was diluted with an equal volume of the DEAE phosphate buffer prior to binding to a HiTrap DEAE column. Dialysed insect media (section 2.6.4) was loaded onto a HiTrap DEAE without needing to be diluted with buffer. Following over-night binding, the DEAE column was washed with 4 CV of DEAE buffer A and then protein eluted with a linear gradient over 6 CV with DEAE buffer B. The C9-containing fractions were identified using Coomassie stained SDS-PAGE and pooled for the next chromatography step.

The pooled DEAE fractions were diluted with an equal volume of CHT phosphate buffer and then bound to a pre-packed CHT type I column equilibrated in CHT wash buffer. The CHT column was washed with 4 CV of CHT buffer A prior to eluting. Then the protein was eluted with a linear gradient over 10 CV with CHT buffer B. This CHT step utilised a mixed approach by simultaneously changing pH (from 7.0 → 8.1) and

increasing conductivity. The C9-containing fractions were confirmed by SDS-PAGE and pooled for either size exclusion chromatography or anion exchange using a monoQ column.

2.7.3   Size exclusion chromatography (SEC)

The pooled fractions from CHT were concentrated using a 30 kDa MWCO protein concentrator prior to size exclusion. For size exclusion of recombinant human C9, concentrated samples were injected onto an S200 column (with dimensions of either 16 mm x 200 mm or 26 mm x 200 mm) and resolved in the appropriate gel filtration buffer. The C9-containing fractions from size exclusion were confirmed by Coomassie stained SDS-PAGE and then pooled for use in structural biology techniques or functional assays.

2.7.4   MonoQ chromatography

High resolution anion exchange chromatography was performed using a MonoQ column to improve preparations that were not well resolved by CHT. The CHT fractions were diluted with MonoQ buffer A then bound to a 1 mL MonoQ column equilibrated in the same buffer. After binding, the protein was eluted with a linear gradient over 30 CV with monoQ buffer B. The C9-containing fractions from MonoQ were confirmed using Coomassie stained SDS-PAGE and pooled for use in structural biology techniques or functional assays.

2.7.5   Ni-NTA purification of insect expressed C9

Ni-NTA purification buffers

Ni-NTA binding buffer	10 mM Na_2HPO_4/NaH_2PO_4 pH 8.0, 500 mM NaCl, 10 mM imidazole
Ni-NTA wash buffer	10 mM Na_2HPO_4/NaH_2PO_4 pH 8.0, 500 mM NaCl, 50 mM imidazole
Ni-NTA elution buffer	10 mM Na_2HPO_4/NaH_2PO_4 pH 8.0, 500 mM NaCl, 500 mM imidazole

For purification of histidine tagged C9, 2.5 mL bed volume of Ni-NTA resin was used per litre of dialysed insect media. The insect media was dialysed as described in section 2.6.4, and bound to Ni-NTA using a gravity flow column overnight at 4 °C. All washes and elution of Ni-NTA resin were performed by pipetting the appropriate buffers onto the Ni-NTA resin. Following binding, the Ni-NTA resin was washed with 5 CV Ni-NTA binding buffer. Then the resin was washed with 5 CV of the Ni-NTA wash buffer. Finally, the protein was eluted with several CV additions of Ni-NTA elution buffer. Eluted fractions containing C9 were determined using Coomassie stained SDS-PAGE. Following Ni-NTA purification, the proteins were purified by size exclusion as described in (Section 2.7.3).

2.8   Electron microscopy imaging techniques

2.8.1   Glow discharging TEM grids

The surface of TEM grids were made hydrophilic by applying negative glow discharge for 45 sec in ambient atmospheric conditions at 38 mbar. Following glow discharge, the protein sample was applied to the TEM grids within 30 min to ensure the grids remained negatively charged.

2.8.2 Negative stain electron microscopy

Samples for negative stain TEM were prepared by applying 2.5 μL of protein onto the carbon side of freshly glow discharged carbon coated grids for 1 min. The grids were blotted onto Whatman paper by touching the edge of the grid perpendicular to the absorbent surface allowing slow drying of the sample on the grid surface. After blotting, 8 μL of either 2% (w/v) uranyl acetate or 1% (w/v) uranyl formate was applied to the carbon side and allowed to stain for 1 min, then blotted as described above. After staining, the grids were dried and then stored in TEM grid boxes prior to imaging using a transmission electron microscope.

2.8.3 Plunge freezing cryo-EM grids

All in-house cryo-EM preparations were performed at the Clive and Vera Ramaciotti Centre for Advanced Molecular Imaging (Monash, Clayton). Holey grids were glow discharged as described in Section 2.8.1. A protein sample (2.5 μL) was applied to the carbon side of the grid, then immediately blotted on both sides by the Vitrobot using Whatman blotting papers then plunge frozen in liquid ethane cooled by liquid nitrogen using the Vitrobot Mark IV (Thermofisher) according to the manufacturer's instructions. The blotting conditions were optimized to obtain the best vitreous ice. In a typical experiment, 100% humidity, 4 °C, with a blotting time of 4 sec, a blot force of -2 and drain time of 1 s was used. Vitrified cryo-EM grids were stored in a liquid nitrogen dewar until time of imaging.

2.8.4 Transmission electron microscopy (TEM)

TEM images were acquired using a Tecnai T12, with an acceleration voltage of 120 kV. For methods about TEM imaging for single particle analysis see specific Methods of **Chapter 3** and **Chapter 5**.

2.9 Crystallography methods

High throughput crystallisation screens were performed using purified recombinant C9 protein. The screening kits that were trialled in this study are listed in **Table 2-3**.

2.9.1 MMCF screening

The initial screening of crystals was performed by the Monash Molecular Crystallisation Facility (MMCF) platform using the sitting drop vapour diffusion method. The crystal trials were prepared using the CrystalMation™ (Rigaku) system with either the Phoenix, NT8 or Mosquito drop dispenser systems. The CrystalMation screens were prepared in a 96-well format with a reservoir containing 50 μL of the various crystallisation conditions dispensed into the 96-well plates. To each well, a volume of 0.1 μL of the crystal condition was added to 0.1 μL of concentrated protein combined on plastic sitting above the reservoir. The plates were sealed with adhesive tape and were inspected for crystal growth by light microscope, ultraviolet inspection (UV) and second order nonlinear imaging of chiral crystals (SONICC).

2.9.2 Fine screening crystal conditions

The crystal trials that produced successful crystal nucleation were repeated manually using the hanging drop vapour diffusion method, in a 24-well plate format. Typically, a reservoir volume of 450 μL was used, containing buffers and precipitants that were prepared in house, or commercially purchased. The hanging drops were set up in a final

volume of 2 μL that contained different ratios of the reservoir liquor: protein, as follows: 1.5 μL::0.5 μL; 1 μL::1 μL; or 0.5 μL::1.5 μL.

2.9.3 SDS-PAGE and mass spectrometry analysis of protein crystals

Protein crystals were washed several times in the reservoir liquor, then dissolved in buffer containing 10 mM HEPES pH 7.2, 100 mM NaCl. Non-reducing Laemmli sample buffer was added to the protein suspension and resolved by SDS-PAGE and Coomassie stained as in section 2.5.2. The protein band was extracted using a sterile scalpel blade and stored in a 1.7 mL Eppendorf tube at 4 °C for mass spectrometry analysis.

Briefly, the gel piece was washed and the protein was reduced and alkylated with DTT/iodoacetamide. The protein was digested overnight with trypsin in 50 μL of 20 mM ammonium bicarbonate buffer. Tryptic digests were analysed by LC-MS/MS at the Monash Biomedical Proteomics facility. Data from the LC-MS/MS experiment was exported to Mascot and searched against the swissprot database using the Mascot search engine version 2.4 (Matrix Science Inc., London, UK).

2.10 Visualisation of protein structures

MacPyMOL v1.7.4.5 (using an education licence provided by Monash University) was used to visualise structures (118). UCSF Chimera v1.11.2 was used to visualise EM maps and perform basic model fitting (119). All model building (both X-ray and cryo-EM) was performed as described in Chapter 5 methods.

2.11 Platforms and facilities

This study utilised several of the Monash University platform facilities including: the Clive and Vera Ramaciotti Centre for Cryo-Electron Microscopy (EM suite); Micromon

sequencing facility for DNA sequencing; the Monash Proteomic Facility for mass spectroscopy analysis (N-terminal Edman sequencing and LC-MS); Monash Micro Crystallisation Facility (MMCF) and Monash Protein Production Unit (PPU). X-ray diffraction data were collected at the Australian Synchrotron (AS).

CHAPTER 3: STRUCTURE OF THE POLY-C9 COMPONENT OF THE COMPLEMENT MEMBRANE ATTACK COMPLEX

Structure of the poly-C9 component of the complement membrane attack complex

Natalya V. Dudkina, Bradley A. Spicer[1,2], Cyril F. Reboul, Paul J. Conroy[1] , Natalya Lukoyanova, Hans Elmlund, Ruby H.P. Law, Susan M. Ekkel[1], Stephanie C. Kondos, Robert J.A. Goode, Georg Ramm, James. C. Whisstock[1,2], Helen R. Saibil and Michelle A. Dunstone[1]*

[1]Department of Biochemistry and Molecular Biology, School of Biomedical Sciences, Monash University, Clayton, Victoria, Australia, 3800.

[2]Centre for advanced molecular imaging – An ARC centre of Excellence, Monash Node, Clayton, Victoria, Australia, 3800.

✱To whom correspondence should be addressed: Michelle A. Dunstone, Department of Biochemistry and Molecular Biology, School of Biomedical Sciences, Monash University, Clayton, Victoria, 3800, Australia. Tel: +61 3 9902 9269; Email: michelle.dunstone@monash.edu

Structure of the poly-C9 component of the complement membrane attack complex

Natalya V. Dudkina[1,*], Bradley A. Spicer[2,3,*], Cyril F. Reboul[2,3,*], Paul J. Conroy[2,3], Natalya Lukoyanova[1], Hans Elmlund[2,3], Ruby H.P. Law[2,3], Susan M. Ekkel[2,3], Stephanie C. Kondos[3], Robert J.A. Goode[3], Georg Ramm[2,3], James C. Whisstock[2,3,**], Helen R. Saibil[1,**] & Michelle A. Dunstone[2,3,4,**]

The membrane attack complex (MAC)/perforin-like protein complement component 9 (C9) is the major component of the MAC, a multi-protein complex that forms pores in the membrane of target pathogens. In contrast to homologous proteins such as perforin and the cholesterol-dependent cytolysins (CDCs), all of which require the membrane for oligomerisation, C9 assembles directly onto the nascent MAC from solution. However, the molecular mechanism of MAC assembly remains to be understood. Here we present the 8 Å cryo-EM structure of a soluble form of the poly-C9 component of the MAC. These data reveal a 22-fold symmetrical arrangement of C9 molecules that yield an 88-strand pore-forming β-barrel. The N-terminal thrombospondin-1 (TSP1) domain forms an unexpectedly extensive part of the oligomerisation interface, thus likely facilitating solution-based assembly. These TSP1 interactions may also explain how additional C9 subunits can be recruited to the growing MAC subsequent to membrane insertion.

[1] Department of Crystallography, Institute of Structural and Molecular Biology, Birkbeck College, London WC1E 7HX, UK. [2] ARC Centre of Excellence in Advanced Molecular Imaging, Clayton Campus, Monash University, Melbourne, Victoria 3800, Australia. [3] Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Clayton Campus, Monash University, Melbourne, Victoria 3800, Australia. [4] Department of Microbiology, Biomedicine Discovery Institute, Clayton Campus, Monash University, Melbourne, 3800 Victoria, Australia. * These authors contributed equally to this work. ** These authors jointly supervised this work. Correspondence and requests for materials should be addressed to J.C.W. (email: James.Whisstock@monash.edu) or to H.R.S. (email: h.saibil@mail.cryst.bbk.ac.uk) or to M.A.D. (email: Michelle.Dunstone@monash.edu).

Paul Ehrlich originally characterized the haemolytic properties of human blood over 100 years ago[1]. Subsequent work revealed that the terminal 'membrane attack complex' (MAC) portion of complement represents the lytic, pore-forming part of the system[2,3]. This structure is responsible for eliminating Gram-negative bacteria and other pathogens.

The MAC comprises seven components: C5b, C6, C7, C8 (a heterotrimer composed of C8α, C8β and C8γ) and multiple copies of C9 (Supplementary Fig. 1). *In vitro* studies reveal that multiple C9 subunits are recruited to the C5b678 complex, whereupon it self-assembles to form large, ring-shaped pores with a lumen over 100 Å in diameter embedded in the membrane of target cells[4]. C9 can also be induced to form poly-C9, pore-like structures in solution that closely resemble the MAC pore[5]. C6, C7, C8α, C8β and C9 all belong to the MAC/perforin-like (MACPF)/CDC superfamily[6,7] and include a common set of four core domains; a N-terminal thrombospondin-1 (TSP1) domain followed by a low-density lipoprotein receptor-associated (LDLRA) domain, a MACPF domain and an epidermal growth factor (EGF) domain (Supplementary Fig. 1).

Much of our understanding of the MACPF/CDC superfamily comes from studying CDCs[8–10]. Briefly, soluble CDC monomers bind to and then oligomerise on the membrane surface to form a prepore intermediate[10,11]. Next the assembly undergoes a concerted conformational change that involves significant opening and untwisting of a central, four-stranded β-sheet. This event permits two helical regions (termed transmembrane hairpins TMH1 and TMH2) to unravel and insert into the membrane as amphipathic β-hairpins (Supplementary Fig. 2).

Studies on the MAC have revealed mechanistic distinctions from other family members. For example, perforin, pleurotolysin and CDCs bind to membrane lipids or membrane-associated proteins via ancillary domains before oligomerisation[9,12,13]. In contrast, C9 does not contain any obvious membrane-binding domain. Thus even when the nascent MAC (C5b678) is associated with the target cell, the assembly process must include the recruitment of C9 from solution (that is, from plasma, Supplementary Fig. 2b). Consistent with this, a soluble form of the MAC can also assemble independently of the membrane and be detected in blood plasma (Supplementary Fig. 2b).

To understand the mechanism of MAC assembly, we determined the sub-nanometer resolution single-particle EM structure of C9 in a polymerized pore-like form. These data reveal the unexpected finding that the TSP1 domain forms a significant portion of the interface between interacting C9 monomers. This finding may explain why the MAC, in contrast to related molecules such as perforin and the CDCs, is able to assemble from monomers directly recruited from the soluble phase. The additional interactions mediated by the TSP1 domain may also explain previous observations[14], where C9 monomers are recruited to a MAC that has already entered the target cell membrane.

Results

The structure of poly-C9. To understand MAC assembly we determined the 8 Å single-particle cryo-EM reconstruction of soluble poly-C9 from 5,000 particles (Fig. 1a–d, Supplementary Figs 3–6). These data revealed a symmetrical assembly of 22 C9 monomers (Fig. 1a–c) that closely resembles the MAC[4]. The structure comprises a ring-shaped assembly of globular domains atop a large β-barrel (Fig. 1a,b). The latter part of the structure is flexible and is less well resolved than the top half of the structure. However, the diameter of the β-barrel (120 Å) is consistent with

the predicted 88-stranded structure and is of sufficient size to permit passage of proteins such as lysozyme[15]. We further observed density, consistent with two N-glycosylation sites, one on each TMH sequence (Supplementary Figs 1 and 7). We observe a bulbous feature at the base of the β-barrel and suggest that this may be a consequence of structural rearrangements to protect the hydrophobic surface that ordinarily contacts the membrane (Fig. 1b). Higher resolution data will be required to validate this suggestion.

In the top, better-resolved portion of the map, the position of each of the four domains in C9 can be unambiguously assigned. Although no crystal structure of C9 is available, we were able to interpret the poly-C9 structure using the core TSP1-LDLRA-MACPF-EGF assembly from the crystal structure of C6 (refs 16,17) (Fig. 1e,f; Supplementary Fig. 1). Indeed, only minor changes in domain orientation are required to dock the C6 structure into the bulk of the poly-C9 density (Fig. 1e,f).

The TSP1 domain forms part of the oligomer interface. Structural studies on other MACPF/CDC proteins reveal that most interactions within the prepore or pore assembly appear to be formed between the relatively flat faces of the MACPF domain[8,11,13]. In contrast the poly-C9 structure reveals that the TSP1 domain packs against the C-terminal α-helix of the MACPF domain of an adjacent monomer and forms an additional and significant portion of the oligomer interface (Fig. 2). Thus in the pore form, each TSP1 domain is wedged between two C-terminal α-helices—one contributed in *trans* from an adjacent monomer and one in *cis*. This interaction at the outer edge of the ring-like assembly forms a quarter ($\sim 690\,\text{Å}^2$) of the total ($\sim 3,000\,\text{Å}^2$) surface buried in the globular, non-barrel region (Fig. 2). The remainder of the interacting surface is contributed by interactions between MACPF domains.

In the MAC it is anticipated that the MACPF domain of the related complement components C6, C7 and C8 form part of the overall circular assembly[3]. Like C9, C6–C8 all contain an analogous TSP1 domain that is functionally important (Supplementary Fig. 1)[14]. It is therefore suggested that the TSP1 domain of each protein in the complete MAC will be positioned at the subunit interface. Indeed, we suggest that the specialized TSP1/MACPF interactions likely explain the unusual ability of the nascent MAC to recruit components directly from solution. In contrast, proteins such as perforin, pleurotolysin and CDCs lack a TSP1 equivalent and do not readily self-assemble in solution. Instead, they require membrane anchoring via ancillary domains in order to oligomerise. Indeed, it is known from the study of receptors that restriction to the membrane plane can favour oligomerisation through weak protein–protein interactions[18].

Conformational transitions during pore formation. We next examined the conformational changes that take place in the transition from the soluble monomer to the pore form. Comparison with C6 suggests that the largest conformational rearrangements during the transition from the monomer to the pore form take place within the MACPF domain[19,20]. The bottom half of the central β-sheet is rotated by $\sim 10°$ relative to its position in C6. This movement shifts the lower part of the β-sheet laterally by ~ 5.5 Å (Fig. 3a,b). Concomitantly with this change, TMH1 and TMH2 unravel to form the β-barrel (Fig. 1b).

The lateral movement in the central sheet of the MACPF domain repositions the conserved helix-turn-helix (HTH) region that sits on top of TMH2 in the soluble monomeric form. Consistent with this, the top of the poly-C9 pore lumen is lined by pairs of α-helices (Fig. 3c). Previous mutagenesis and

69
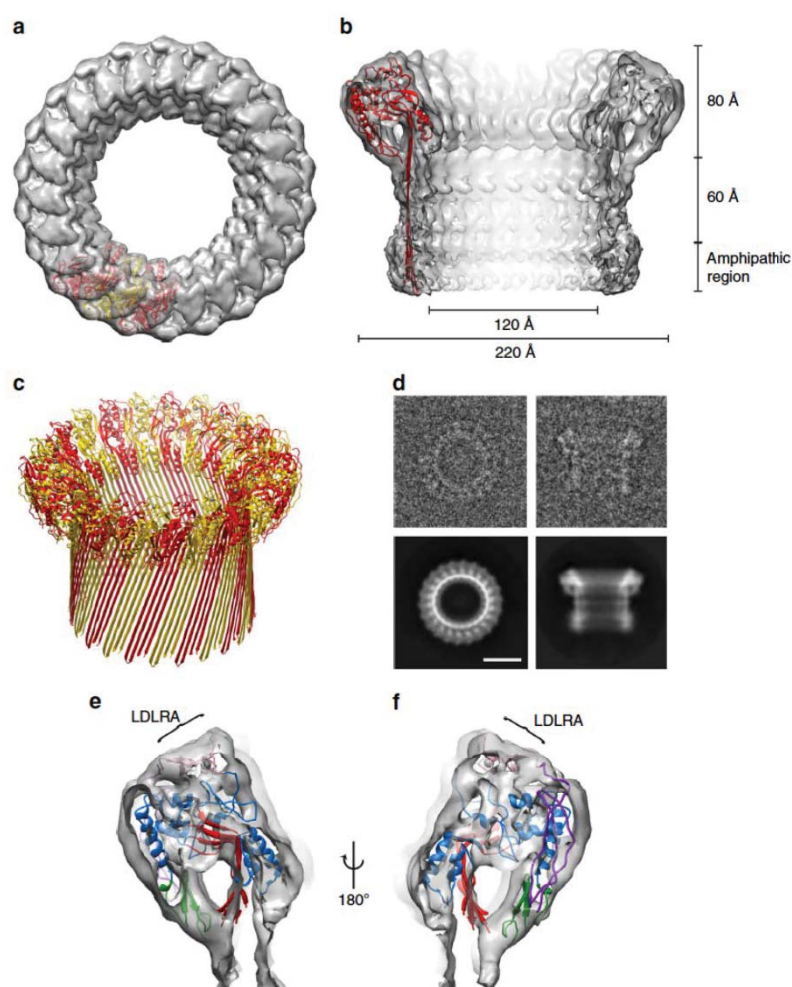
Figure 1 | The structure of poly-C9. (a) Top–down view of a C9 trimer in the poly-C9 map and **(b)** cut through of the poly-C9 map with cartoon (red) of the poly-C9 model. Approximate dimensions and the predicted amphipathic region are indicated. **(c)** Cartoon of the full poly-C9 pore (alternating red and yellow monomers). The barrel is best modelled with the architecture $S = n/2$ (ref. 42). **(d)** Cryo-EM end and side views of poly-C9 in individual images (top) and class averages (bottom). **(e,f)** With the exception of the mobile region of the MACPF domain (which in poly-C9 has rearranged in order to form the barrel), the crystal structure of C6 (PDB ID: 3T50) fits well into the map, with TMH1 and TMH2 omitted for clarity. In this figure the conserved β-sheet of the MACPF domain is in red, the body of the MACPF domain is in blue, the EGF domain in green, the TSP1 domain in purple and the LDLRA domain in pink (labelled).

structural studies on the fungal MACPF protein pleurotolysin, as well as the CDC suilysin, suggest a role of the HTH region in pre-pore assembly and in controlling the transition to the pore[13,21].

Discussion

The structure of poly-C9 provides mechanistic insight into how components of the MAC may assemble through additional interactions mediated via the TSP1 domain. Furthermore, the structure provides insights into self-association by MACPF domain-containing proteins more generally. In particular, our present poly-C9 structure may resolve the controversy regarding the orientation of perforin in the pore assembly. Our previous

analysis of the low-resolution EM structure of the perforin pore suggested that perforin monomers are orientated in the pore assembly opposite to the CDCs and pleurotolysin[8,11]. The latter two proteins, however, share very limited (<10%) sequence identity in the MACPF domain with perforin, whereas C9 is more closely related (~25% identity). Accordingly, we superposed the perforin structure onto the poly-C9 model. This suggests that perforin most likely oligomerises similarly to C9, following minor rearrangements of the TMH2 and HTH domains (Supplementary Fig. 8). We note that residues shown through mutagenesis studies to interact at the pore interface are brought into close proximity with one another[22]. Further, the absence of the TSP1 domain in perforin at the outer edge of the pore assembly may explain the heterogeneity in perforin pore size and shape. We thus conclude

70
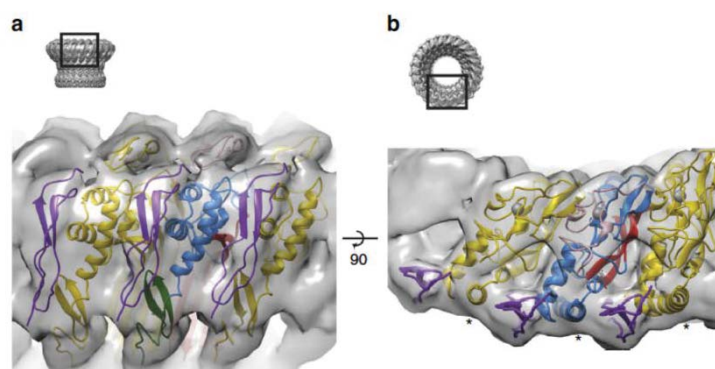
Figure 2 | Interactions made by the TSP1 domain. (**a**) A view of the outside of the globular portion of the poly-C9 map showing the TSP1 domain (purple) located at each subunit interface. The central C9 monomer is coloured as in Fig. 1, with the monomers each side in dark yellow and purple (TSP1 domain). (**b**) A view from the top showing placement of the TSP1 domain between the C-terminal helix (marked with *) of each MACPf domain.
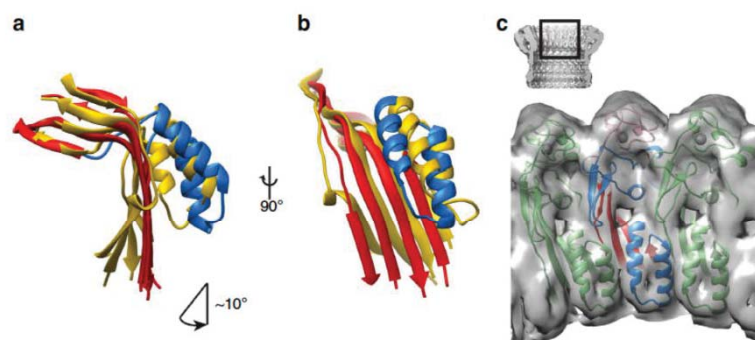
Figure 3 | Comparison between the structure of C6 (PDB ID 3T5O; yellow) and model of poly-C9 (red/blue). The shift of the central bent β-sheet (red) shows (**a**) an ~10° rotation of the bottom half of the sheet together with (**b**) an ~5.5 Å lateral movement. (**c**) The HTH region (a pair of α-helices) lines the pore lumen. A trimer is shown with the central monomer coloured red, blue and pink.

that the present 8-Å-resolution poly-C9 map thus provides a better model for the perforin assembly.

Finally, the new structural insights may help explain how the MAC assembles with respect to target cell membranes. In the current view, C7 and C8 sequentially insert into the membrane, anchoring it in place before the recruitment of multiple copies of C9. However, this mechanism contrasts with the current view of the MACPF/CDC pore formation, in which the amphipathic hairpins are proposed to be inserted in a concerted fashion in the context of a complete or incomplete ring[11,23]. The latter mechanism seems more plausible because the conformational change in the MACPF domain during membrane insertion is extensive and would be predicted to disfavour the addition of new subunits. The poly-C9 structure provides new insights into this problem. The additional TSP1/MACPF interactions involve regions of the molecule that do not undergo significant conformational change. We therefore hypothesize that the TSP1-mediated interactions may permit addition of C8 and C9 to a nascent MAC that has already entered the target membrane.

To conclude, we have determined the structure of poly-C9 at a resolution sufficient to confidently position individual domains and to resolve helical features in density. Our data further reveal an unexpected contribution of domains ancillary to the MACPF domain that likely function to stabilize the overall assembly and the top half of the β-barrel pore.

Methods

The protein underwent an additional chromatography step using MonoQ ion-exchange chromatography (GE Healthcare Life Sciences). Here, pooled fractions from the size exclusion chromatography step were concentrated using a 30 kDa MWCO centricon concentrator and buffer exchanged 2–3 times into 10 mM Tris-HCl pH 7.2, 100 mM NaCl. Buffer-exchanged protein was loaded onto a 1 ml MonoQ column and eluted over a linear gradient from: 10 mM Tris-HCl pH 7.2, 100 mM NaCl to 10 mM Tris-HCl pH 7.2, 350 mM NaCl. Purified protein that was shown to be haemolytically active and able to assemble into a complete MAC was used for EM experiments.

Haemolytic assays and MAC assembly on ghost cell membranes. Sheep red blood cells (sRBC) were washed with DGHB^{++} pH 7.4 (Dextrose Gelatin HEPES Buffer; with 2.5% (w/v) D-glucose, 0.1% (w/v) gelatin, 5 mM HEPES pH 7.4, 71 mM NaCl, 0.15 mM CaCl$_2$, 0.5 mM MgCl$_2$). In all, 6.5×10^8 sRBC were sensitized with an equal volume of anti-sheep antibody to a concentration of 0.75 mg ml^{-1} and incubated at 30 °C for 30 min to activate the classical pathway. Excess antibody was washed off before reactions. The lysis reactions were set up in triplicate with 0.2 ml sRBC (3.75×10^6 cells), 1 µl of C9-depleted serum (Complement Technology, USA) and 15 µl of C9 (initial concentration 8 µg ml^{-1}) and twofold dilutions of purified C9 in thin-walled PCR tubes. Reaction tubes were incubated at 37 °C on a PCR heat block for 30 min and immediately placed at 4 °C, then centrifuged for 20 s. The supernatant (150 µl) of the lysis reactions was transferred to a 96-well plate and absorbance at 405 nm was measured. The reactions were normalized to 0% lysis with a buffer control or to a reaction containing PlyA and PlyB[13] for 100% lysis.

Ghost membranes were prepared with rabbit red blood cells (rRBC) washed with DGHB^{++}. Packed red blood cells were pre-incubated with C9-depleted serum for 15 min at 37 °C. The rRBCs were washed of excess sera and re-suspended in DGHB^{++}. Purified C9 protein was added to rRBCs and incubated at 37 °C for 15 min. Reactions were immediately centrifuged for 30 s at 16,100 r.c.f. and the supernatant transferred to new tubes. The supernatant was further centrifuged for 10 min at 16,100 r.c.f. and the pellet containing membranes was washed once with DGHB^{++} and then resuspended in 10 mM phosphate buffer pH 8.0, 50 mM NaCl to make ghosts. Carbon-coated copper grids containing formvar were glow discharged, then floated over 8 µl of re-suspended ghosts followed by staining with 2% (w/v) uranyl acetate for 1 min. Pores were examined on a Hitachi H7500 TEM at 80 kV.

Characterization of the glycosylation state. Purified C9 was reduced with 2 mM DTT, alkylated with 5 mM iodoacetamide and digested with trypsin (Promega) in 1:40 ratio at 37 °C overnight. The digest was desalted with C18-packed tips (OMIX, Agilent) before nanoLC-MS/MS (Dionex Ultimate 3000 LC coupled to QExactive Plus, Thermo). Peptides (~1 µg) were loaded on a 2 cm trap column (100 µm ID, Acclaim PepMap 100, Thermo Scientific) in 2% (v/v) acetonitrile, 0.1% (v/v) trifluoroacetic acid and resolved on a 50 cm column (75 µm ID, Acclaim PepMapRSLC, Thermo Scientific) with a non-linear 25 min gradient from 2% (v/v) to 34% (v/v) acetonitrile in 0.1% (v/v) formic acid. Spectra were acquired in a Top12 strategy with full scans (375–1,800 m z^{-1}) acquired at 70,000 resolution and data-dependent HCD MS2 spectra acquired at 17,500 resolution. Peptide assignment was performed with the Preview and Byonic software (Protein Metrics[26]) utilizing Preview-determined modifications and mass tolerances, a focused human database from an initial Byonic search and N- and O-glycosylation databases for assignment of glycan compositions. All glycan composition assignments were manually validated. Skyline software (University of Washington) was used for semi-quantitative assessment of site-specific glycan compositions, using parent scan extracted ion chromatograms[27] (Supplementary Fig. 7).

Cryo-EM sample preparation and data acquisition. Monomeric C9 was polymerized by overnight incubation at 1 mg ml^{-1} and 37 °C. The resulting poly-C9 was applied to lacey carbon-coated copper grids (Agar, UK) and frozen with a FEI Vitrobot Mark III (FEI, Eindhoven) at 22 °C and 100% humidity. Images were recorded manually on a Tecnai G2 Polara microscope (FEI) operating at 300 kV with a Quantum energy filter and K2 Summit detector (Gatan, UK) in counting mode, at a pixel size of 2.76 Å. Exposures were recorded at 1.2 electrons (Å^2)$^{-1}$ s^{-1} for 25 s, with defocus values ranging from 1.2 to 4.9 µm (Supplementary Fig. 3).

3D reconstruction of poly-C9. The detector movies were aligned using IMOD[28]. CTF parameters were determined with CTFFIND4 (ref. 29). A total of 10,800 particles were extracted manually using Boxer (EMAN 1.9) (ref. 30). Classification and refinement were performed using RELION[31]. 2D classification in IMAGIC[29] revealed mainly end views with 22-fold symmetry, with a small fraction of particles having 21- or 23-fold symmetry (Supplementary Fig. 3). The initial model with 22-fold symmetry was created by angular reconstitution from 2D class averages of particles with all orientations in IMAGIC[32] and refined by projection matching using SPIDER[33]. A subset of ~5,000 particles in side and tilted views (homogeneous with respect to diameter of the wide part of the ring, corresponding to the 22-mers) was refined with RELION using the initial model from SPIDER filtered to 20 Å. Twenty-two-fold symmetry was applied during refinement. The

final map was corrected to the modulation transfer function of the detector and sharpened by applying a B-factor[34] determined by RELION. The final resolution calculation based on gold-standard FSC was estimated at 0.5 and 0.143 FSC in RELION. Local resolution was estimated using the ResMap program[35] (Supplementary Fig. 4).

Determination of handedness. In order to determine the absolute hand of the 3D reconstruction, the crystallographic structure of C6 was fitted into the map as well as into the map with opposite handedness. Although the fit of the C6 conformation was found to favour one hand over the other, the differences in cross-correlations were too small to conclusively assign the hand of the map (C6: 0.63 versus 0.59; calculated using the Chimera software[36]).

To resolve this issue, we examined the fit to both maps of the conserved structural features of the MACPF domain. The C-terminal α-helical bundle of the MACPF domain (Supplementary Fig. 5) is likely to be clearly discernable in an 8-Å-resolution cryo-EM density. Its characteristic arrangement of α-helices is asymmetric and highly conserved in all the crystallographic structures of MAC components[17]. We therefore expected that it should be possible to identify the correct hand from analysis of the fit of this structural motif in the enantiomeric maps (Supplementary Fig. 5b,c).

Accordingly, we found that the map in Supplementary Fig. 5b showed distinct density corresponding to the C-terminal α-helical bundle. The region of the map identified by rigid body fitting excellently reproduces the topology and length of the α-helices. Conversely, the map in Supplementary Fig. 5c produces a comparatively poor agreement with the fitted position of C6 (Supplementary Fig. 5c). We concluded that the map in Supplementary Fig. 5b represents the correct hand.

Fitting of atomic models. A homology model of C9 was fitted into the EM map by using a combination of manual, rigid body and flexible fitting. The C9 homology model was generated using the crystallographic structures of C6 (PDB IDs: 3T5O, 4A5W) and C8 (2RD7, 3OJYA, 3OJYB) and Modeller 9.14 (ref. 37). The TMH1/2 regions were discarded because these regions form a β-barrel in poly-C9.

Five symmetry-related monomers were then subjected to flexible fitting (MDFF methodology as implemented in NAMD 2.10 (ref. 38) using symmetry restraints[39]. The protein secondary structure was restrained to avoid overfitting. Oligomeric main chain hydrogen bonds between the β-sheets forming the top of the β-barrel were also restrained to reproduce the pattern conserved in the MACPF/CDC superfamily[13,40]. Two independent 5-ns simulations were performed in vacuo at 310 K (γ = 0.3; 1 fs time step; 12 Å cutoff for long-range interaction) using the CHARMM36 force field[41] and followed by 5,000 steps of energy minimization (γ = 0.5). The resulting model with the highest CC (0.93; Molprobity score of 1.15) was replicated with C22 symmetry and combined with a structural model of the 88-stranded β-barrel (architecture S = n/2 (ref. 42) using Modeller, thus extending the β-strands of the central MACPF β-sheet as performed in Leung et al.[11] Lukoyanova et al.[13] and Reboul et al.[43] The final poly-C9 22-mer model (CC of 0.94) is shown in Fig. 1.

In order to assess the reliability of the fitting procedure, the flexible fitting step was repeated using cryo-EM maps calculated from randomly partitioned half-sets, independently refined using RELION and used to determine the resolution of the final cryo-EM map (see EM methods). Individual residue RMSDs of both fitted models were calculated with respect to the structural model obtained from the whole data set (Supplementary Fig. 6). Structural elements displaying an overall high RMSD (for example, not fitted in a reproducible manner; >3.5 Å) were not included in the final structural model.

References

1. Kaufmann, S. H. E. Immunology's foundation: the 100-year anniversary of the Nobel Prize to Paul Ehrlich and Elie Metchnikoff. *Nat. Immunol.* **9**, 705–712 (2008).
2. Walport, M. J. Advances in immunology: complement (first of two parts). *N. Engl. J. Med.* **344**, 1058–1066 (2001).
3. Podack, E. R. Molecular mechanisms of cytolysis by complement and by cytolytic lymphocytes. *J. Cell Biochem.* **30**, 133–170 (1986).
4. Tschopp, J., Podack, E. R. & Müller-Eberhard, H. J. Ultrastructure of the membrane attack complex of complement: detection of the tetramolecular C9-polymerizing complex C5b-8. *Proc. Natl Acad. Sci. USA* **79**, 7474–7478 (1982).
5. Tschopp, J., Muller-Eberhard, H. J. & Podack, E. R. Formation of transmembrane tubules by spontaneous polymerization of the hydrophilic complement protein C9. *Nature* **298**, 534–538 (1982).
6. Rosado, C. J. et al. A common fold mediates vertebrate defense and bacterial attack. *Science* **317**, 1548–1551 (2007).
7. Hadders, M. A., Beringer, D. X. & Gros, P. Structure of C8alpha-MACPF reveals mechanism of membrane attack in complement immune defense. *Science* **317**, 1552–1554 (2007).
8. Tilley, S. J., Orlova, E. V., Gilbert, R. J. C., Andrew, P. W. & Saibil, H. R. Structural basis of pore formation by the bacterial toxin pneumolysin. *Cell* **121**, 247–256 (2005).

9. Shepard, L. A. *et al.* Identification of a membrane-spanning domain of the thiol-activated pore-forming toxin *Clostridium perfringens* perfringolysin O: An alpha-helical to beta-sheet transition identified by fluorescence spectroscopy. *Biochemistry* **37,** 14563–14574 (1998).

10. Czajkowsky, D. M., Hotze, E. M., Shao, Z. & Tweten, R. K. Vertical collapse of a cytolysin prepore moves its transmembrane beta-hairpins to the membrane. *EMBO J.* **23,** 3206–3215 (2004).

11. Leung, C. *et al.* Stepwise visualization of membrane pore formation by suilysin, a bacterial cholesterol-dependent cytolysin. *Elife* **3,** e04247 (2014).

12. Law, R. H. P. *et al.* The structural basis for membrane binding and pore formation by lymphocyte perforin. *Nature* **468,** 447–451 (2010).

13. Lukoyanova, N. *et al.* Conformational changes during pore formation by the perforin-related protein pleurotolysin. *PLoS Biol.* **13,** e1002049 (2015).

14. Scibek, J. J., Plumb, M. E. & Sodetz, J. M. Binding of human complement C8 to C9: role of the N-terminal modules in the C8 alpha subunit. *Biochemistry* **41,** 14546–14551 (2002).

15. Martinez, R. J. & Carroll, S. F. Sequential metabolic expressions of the lethal process in human serum-treated *Escherichia coli*: role of lysozyme. *Infect. Immun.* **28,** 735–745 (1980).

16. Hadders, M. A. *et al.* Assembly and regulation of the membrane attack complex based on structures of C5b6 and sC5b9. *Cell Rep.* **1,** 200–207 (2012).

17. Aleshin, A. E. *et al.* Structure of complement C6 suggests a mechanism for initiation and unidirectional, sequential assembly of membrane attack complex (MAC). *J. Biol. Chem.* **287,** 10210–10222 (2012).

18. Wu, Y., Vendome, J., Shapiro, L., Ben-Shaul, A. & Honig, B. Transforming binding affinities from three dimensions to two with application to cadherin clustering. *Nature* **475,** 510–513 (2011).

19. DiScipio, R. G. The size, shape and stability of complement component C9. *Mol. Immunol.* **30,** 1097–1106 (1993).

20. DiScipio, R. G. & Berlin, C. The architectural transition of human complement component C9 to poly(C9). *Mol. Immunol.* **36,** 575–585 (1999).

21. Ramachandran, R., Tweten, R. K. & Johnson, A. E. Membrane-dependent conformational changes initiate cholesterol-dependent cytolysin oligomerization and intersubunit beta-strand alignment. *Nat. Struct. Mol. Biol.* **11,** 697–705 (2004).

22. Baran, K. *et al.* The molecular basis for perforin oligomerization and transmembrane pore assembly. *Immunity* **30,** 684–695 (2009).

23. Sonnen, A. F.-P., Plitzko, J. M. & Gilbert, R. J. C. Incomplete pneumolysin oligomers form membrane pores. *Open Biol.* **4,** 140044 (2014).

24. Biesecker, G. & Müller-Eberhard, H. J. The ninth component of human complement: purification and physicochemical characterization. *J. Immunol.* **124,** 1291–1296 (1980).

25. Biesecker, G., Lachmann, P. & Henderson, R. Structure of complement poly-C9 determined in projection by cryo-electron microscopy and single particle analysis. *Mol. Immunol.* **30,** 1369–1382 (1993).

26. Bern, M., Kil, Y. J. & Becker, C. Byonic: advanced peptide and protein identification software. *Curr. Protoc. Bioinformatics* **40,** 13.20.1–13.20.14 (2012).

27. Schilling, B. *et al.* Platform-independent and label-free quantitation of proteomic data using MS1 extracted ion chromatograms in skyline: application to protein acetylation and phosphorylation. *Mol. Cell Proteomics* **11,** 202–214 (2012).

28. Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116,** 71–76 (1996).

29. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192,** 216–221 (2015).

30. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128,** 82–97 (1999).

31. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180,** 519–530 (2012).

32. van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. *J. Struct. Biol.* **116,** 17–24 (1996).

33. Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116,** 190–199 (1996).

34. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333,** 721–745 (2003).

35. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11,** 63–65 (2014).

36. Pettersen, E. F. *et al.* UCSF chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).

37. Eswar, N. *et al.* Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci.* **50,** 2.9.1–2.9.31 (2007).

38. Trabuco, L. G., Villa, E., Mitra, K., Frank, J. & Schulten, K. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16,** 673–683 (2008).

39. Chan, K.-Y. *et al.* Symmetry-restrained flexible fitting for symmetric EM maps. *Structure* **19,** 1211–1218 (2011).

40. Sato, T. K., Tweten, R. K. & Johnson, A. E. Disulfide-bond scanning reveals assembly state and β-strand tilt angle of the PFO β-barrel. *Nat. Chem. Biol.* **9,** 383–389 (2013).

41. Huang, J. & MacKerell, A. D. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J. Comput. Chem.* **34,** 2135–2145 (2013).

42. Reboul, C. F., Mahmood, K., Whisstock, J. C. & Dunstone, M. A. Predicting giant transmembrane β-barrel architecture. *Bioinformatics* **28,** 1299–1302 (2012).

43. Reboul, C. F., Whisstock, J. C. & Dunstone, M. A. A new model for pore formation by cholesterol-dependent cytolysins. *PLoS Comput. Biol.* **10,** e1003791 (2014).

Author contributions

M.A.D., H.R.S. and J.C.W. conceived the study, co-led the work and co-wrote the paper. N.V.D. collected data, determined the structure and co-wrote the paper. N.L. performed the initial EM of the C9 preparations. B.A.S., P.J.C., R.H.P.L., S.M.E., S.C.K., G.R. and R.J.A.G. produced and analysed protein and optimized sample for the EM experiments. C.F.R. and H.E. performed the computational analysis.

Additional information

Accession codes. The cryo-EM map of poly-C9 is available in the Electron Microscopy Data Bank (accession number EMD-3235). The structure coordinate file for the fitted pore model is available in the Protein Data Bank database (accession number 5FMW).

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Dudkina, N. V. *et al.* Structure of the poly-C9 component of the complement membrane attack complex. *Nat. Commun.* **7:**10588 doi: 10.1038/ncomms10588 (2016).
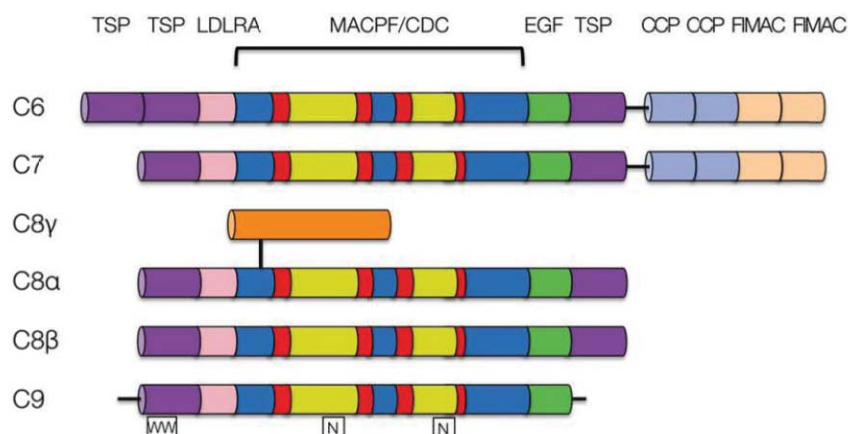
Supplementary Fig. 1. Schematic of domain composition of the Membrane Attack Complex proteins

Schematic of the domain composition of C9 in comparison with C6, C7, C8α, C8β. The schema of C9 includes the sites of C-mannosylation (WW) and N-glycosylation (N) of C9 as determined according to methods and in agreement with published data [1,2]. The two predicted TMH regions are also labelled. Colours are the same as used throughout the domain coloured figures. TSP = Thrombospondin Type 1 domain, LDLRA = Low-Density Lipoprotein Receptor Type A, MACPF = Membrane Attack Complex/Perforin / Cholesterol Dependent Cytolysin, EGF = Epidermal Growth Factor-like, CCP = Complement Control Protein, FIMAC = Factor I / Membrane Attack Complex domain. Colours are the same as used in Figs 1, 2 and 3.
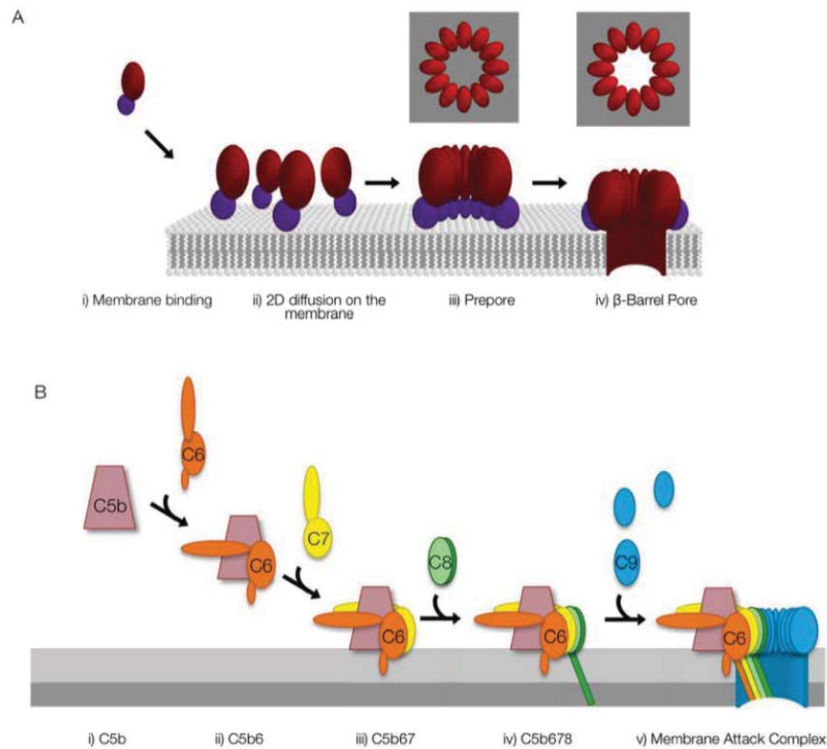
Supplementary Fig. 2 The general mechanism of MACPF/CDC pore forming proteins (adapted from [3]).

a) The CDC pore forming mechanism. i) Membrane recognition and binding, ii) two dimensional diffusion of the monomers on the membrane and iii) oligomerisation into the prepore state. iv) β-barrel pore formation, postulated to occur as a concerted, simultaneous insertion of the two TMH regions. Insets at the top show the state of the membrane in the prepore and pore states.

b) The general mechanism of the Membrane Attack Complex (MAC) formation. After formation of C5b (pink, panel (i)) there is sequential binding of C6 (orange, panel (ii)), C7 (yellow(iii)) and the C8 heterotrimer (green, C8γ not shown, panel (iv)). It is postulated that the C8α component (dark green) of C8 inserts its TMH2 region into the membrane at this stage. In panel (v) the final pore formation is depicted with the sequential addition of ~18 C9 molecules (blue; initial interface proposed to be with C8).
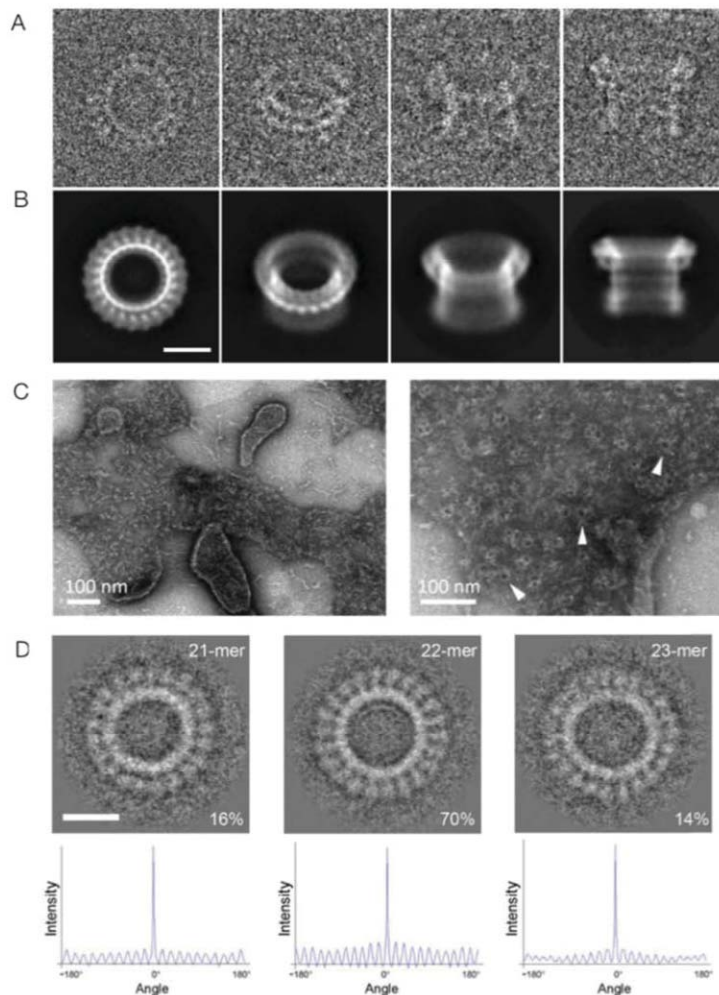
Supplementary Fig. 3 Electron microscopy of polyC9 and MAC.

Representative cryo-EM **(a)** views of four raw images of polyC9 with the **(b)** corresponding averaged views. Scale bar, 10 nm.
(c) Rabbit red blood cell ghosts incubated with C9 depleted serum without the addition of purified C9 (left) and with the addition of purified C9 (right). MAC pores are indicated by white arrows. **(d)** Symmetry of polyC9 pores. Representative averaged views of 21, 22 and 23-fold symmetric pores (upper row) with proportion of particles with each symmetry detected in the data set and corresponding rotational autocorrelations (lower row). The bar is 10 nm.
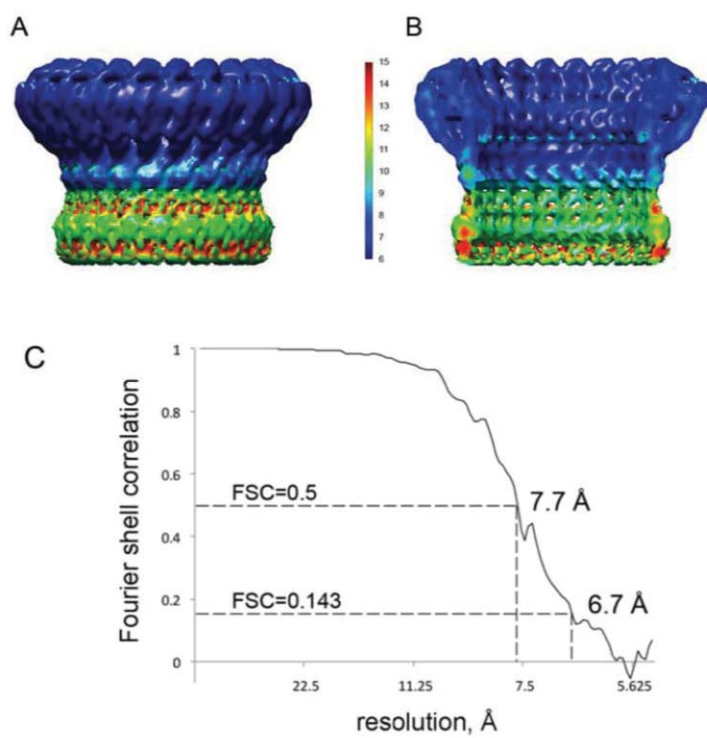
Supplementary Fig. 4 Map quality of poly-C9.
(a) A surface view of the sharpened final map colored according to local resolution. **(b)** As in A, but a cut-through view. **(c)** Fourier-shell correlation curve.
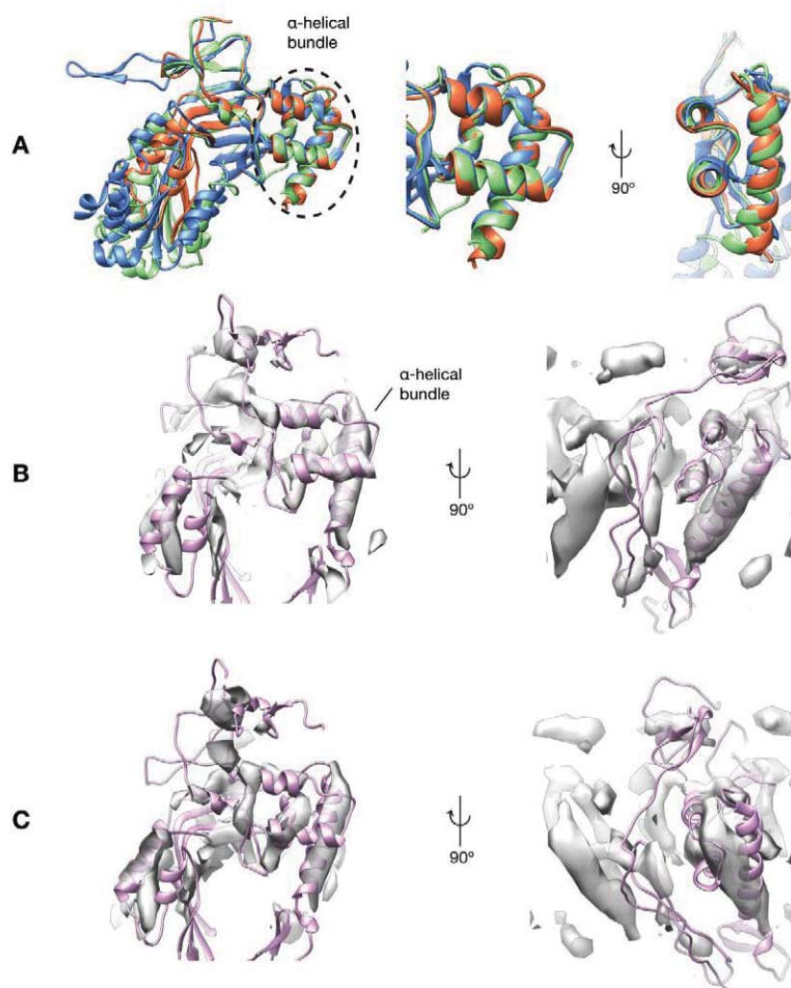
Supplementary Fig. 5 Structural alignment of the MACPF domain of MAC components and determination of handedness.

(a) The alignment identifies two crossed pairs of α-helices (far left, circled) as a highly structurally conserved region of the MACPF domain. The topology and lengths of the individual α-helices are identical in MAC components (middle and right panels). Superpositions of the crystallographic structures of C6 (PDB ID 3T5O, orange), C8α (PDB ID: 2RD7, green) and C8β (PDB ID: 3OJY chain B, blue). (b) The hand of map chosen in this study and (c) the mirrored map. The crystallographic conformation of C6 (PDB ID: 3T5O, pink) is in cartoon representation.

78
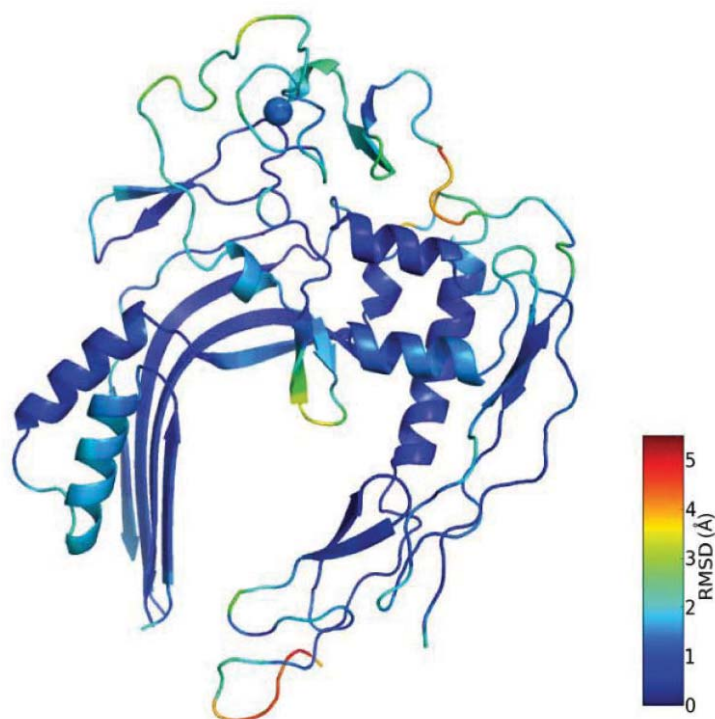
Supplementary Fig. 6 Reproducibility of the flexible fitting step.
Regions displaying a high RMSD are limited to the C-terminus and the loop at positions 74-79.
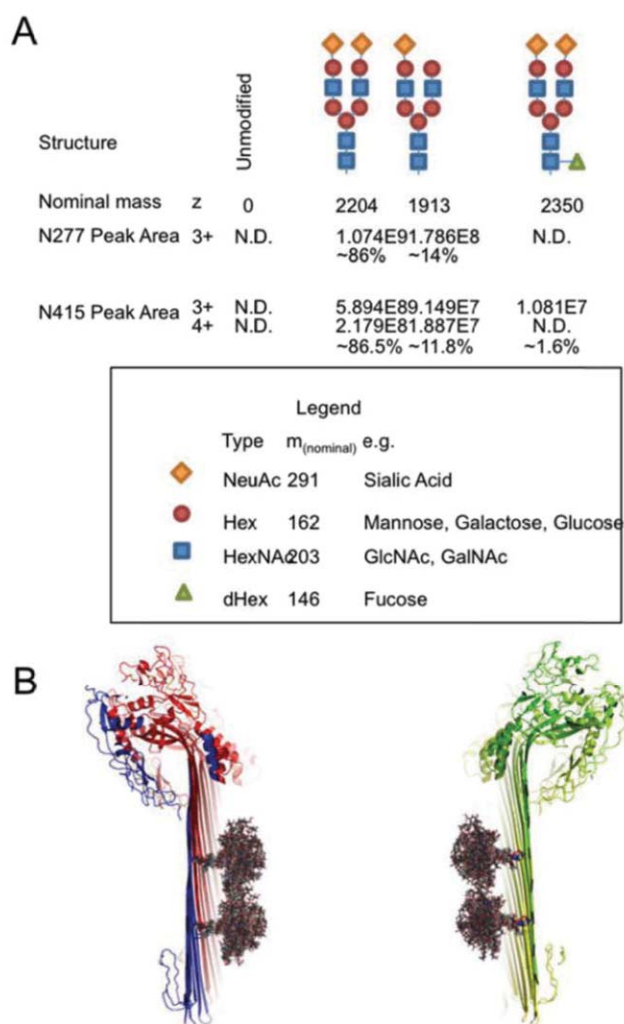
Supplementary Fig. 7 The N-glycosylation of C9

(a) MS-MS characterisation of the two N-glycan sites. Most of the N-glycans have two sialic acid groups each.

(b) Superposition of the glycan models from the NMR structure of human chorionic gonadotropin (PDB ID: 1HD4). These N-glycans lack the sialic groups but show the potential degrees of freedom of each of the 44 glycan groups located in the pore lumen.
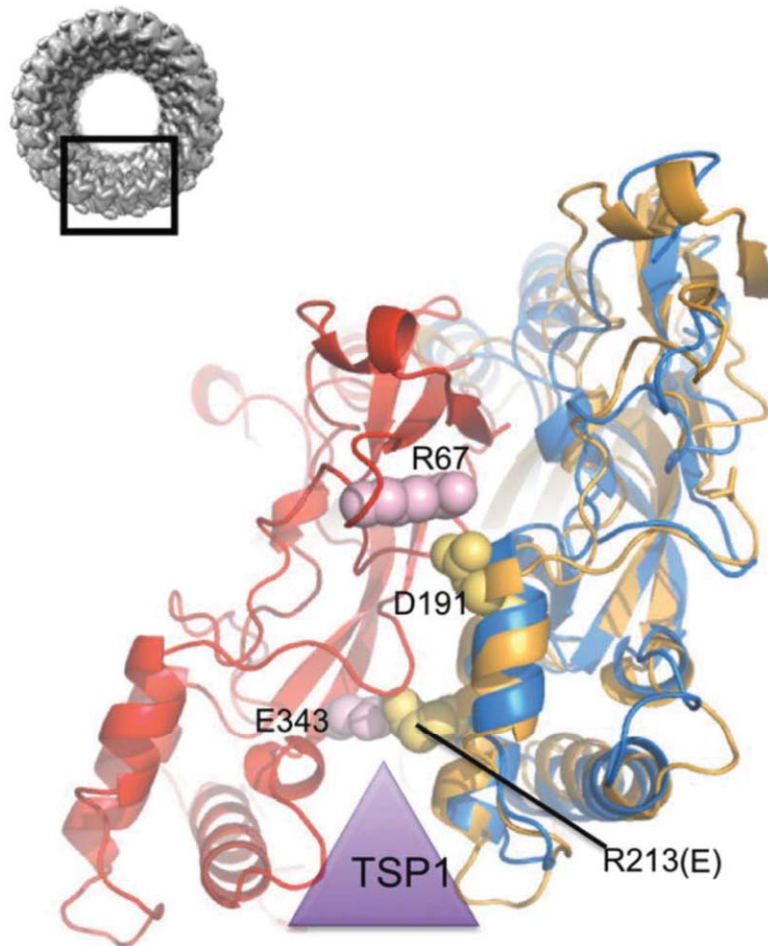
Supplementary Fig. 8 Superposition of perforin on the model of polyC9.
Illustration of the arrangement of two perforin molecules (red and orange) generated by
superposition of the perforin structure (PDB ID: 3NSJ) on the model of poly-C9
(blue). In the superposition R213 (mutated to a Glu in 3NSJ) is in close proximity to
E343 as previously predicted [4]. These data further suggest that D191 which has been
shown to be important for pore formation forms a salt bridge with R67.

Supplementary References

1. Hofsteenge, J., Blommers, M., Hess, D., Furmanek, A. & Miroshnichenko, O. The four terminal components of the complement system are C-mannosylated on multiple tryptophan residues. *J. Biol. Chem*. **274**, 3276-94 (1999).

2. DiScipio, R. G. & Hugli, T. E. The architecture of complement component C9 and poly(C9). *J. Biol. Chem*. **260**, 14802-9 (1985).

3. Kondos, S.C. et al. The structure and function of mammalian membrane-attack complex/perforin-like proteins. *Tissue Antigens* **76**, 341-51 (2010).

4. Baran, K. et al. The molecular basis for perforin oligomerization and transmembrane pore assembly. *Immunity* **30**, 684-95 (2009).

CHAPTER 4: RECOMBINANT EXPRESSION OF C9 AND STRUCTURAL CHARACTERIZATION

4.1 Introduction

The final MAC pore is an integral membrane complex that forms a 12-nm β-barrel in target membranes. In general, it is hypothesised that this transmembrane β-barrel leads to cell death by osmotic flux (120). However, for gram-negative bacteria it is also possible that the MAC delivers lysozyme across the outer membrane into the periplasm (31). The current view of the MAC assembly pathway supports a model where up to 19 soluble C9 monomers bind to the nascent C5b-8 complex. Upon binding, C9 undergoes a conformational change to insert into the membrane to form the final pore. Human C9 is also able to oligomerise into polyC9, an assembly that has the same overall dimensions as the membrane bound MAC (107,108,110).

The use of human plasma-derived C9 has been extensively documented and has been the preferred source of C9 for research (104–106). In contrast, and despite the determination of the human C9 gene sequence over 30 years ago (121,122), there are few published studies utilizing recombinant expression of C9. This could be a consequence of poor protein expression yields. For example, comparative expression analyses revealed that yeast, COS7 and insect cells secrete 5 ng; 111 ng and 700 ng of protein per mL of culture, respectively (123). Another study showed that C9 protein expression from insect cells could be improved to 2.5 µg/mL using Hi5 cells (that are optimised for protein secretion) (124).

Regardless of the problems associated with yield, several studies have utilised recombinant C9 to investigate how glycosylation, truncations and mutations affect MAC assembly (124–129). However, it is notable that most of the published uses of recombinant C9 have used mammalian cell expression media supplemented with FBS, which itself contains complement proteins. The latter contaminations make protein quantification and functional investigation far more challenging. Furthermore, from the

perspective of protein crystallography, the availability of homogenous protein sample is essential (130).

Several studies have utilised C9 purified from plasma to study its oligomerisation within the MAC (including the results presented in Chapter 3) (104–106). These studies have resulted in structures determined at resolutions ranging from 8 Å to 23 Å. Despite these successes, our understanding of MAC assembly has been limited by a paucity of atomic detail. Moreover, the reliance on material purified from serum further meant that it was not possible to study C9 mutants designed to gain additional biological insights.

Collectively, to address these issues, a range of different strategies are reported to produce high quality, pure recombinantly expressed C9 and variants thereof. These experiments were crucial for undertaking the work described in Chapter 5. The methods of expression and purification of C9 reported here includes testing baculovirus-(Hi5) and mammalian (HEK293F)-based expression systems, investigating murine and human C9, codon optimization, glycosylation mutants, truncation mutants and point mutations (**Table 4-1** and **Figure 4-1**).

Table 4-1 Recombinant C9 constructs and their uses for structural biology in this thesis.

This chapter will detail the C9 expression in three parts: first human C9 from insect cells (blue); human C9 from mammalian cells (pink) and mouse C9 from mammalian cells (green). Single ticks represent trials in crystallography or TEM experiments. The triple ticked boxes represent successful use in structural studies for crystallography and TEM discussed in Chapter 5.

Species	C9 Constructs	Expression system	Description	Crystal trials	TEM
Human	C9[bac]	Insect	wild-type	✔	✔
	C9[N-his bac]	Insect	his tagged	✔	✔
	C9[aglyco bac]	Insect	non-glycosylated his tagged	✔	
	C9[HEK]	Mammalian	wild-type	✔	✔✔✔
	C9[aglyco-1] (T258M/T396M)	Mammalian	non-glycosylated	✔	
	C9[aglyco-2] (N256D/N394D)	Mammalian	non-glycosylated	✔	
	C9[ΔN/ΔC] (Δ1-15;527-538)	Mammalian	truncated	✔	
	C9[aglyco ΔN/ΔC]	Mammalian	non-glycosylated truncated	✔	
	C9[SERp] ($_{127}$EESE$_{130}$→AASA)	Mammalian	surface entropy reduction	✔	
Murine	C9[murine]	Mammalian	wild-type	✔	
	C9[aglyco murine] (N28E/N243D/N397D)	Mammalian	non-glycosylated	✔✔✔	
	C9[aglyco ΔN murine] (Δ1-14)	Mammalian	non-glycosylated truncated	✔✔✔	
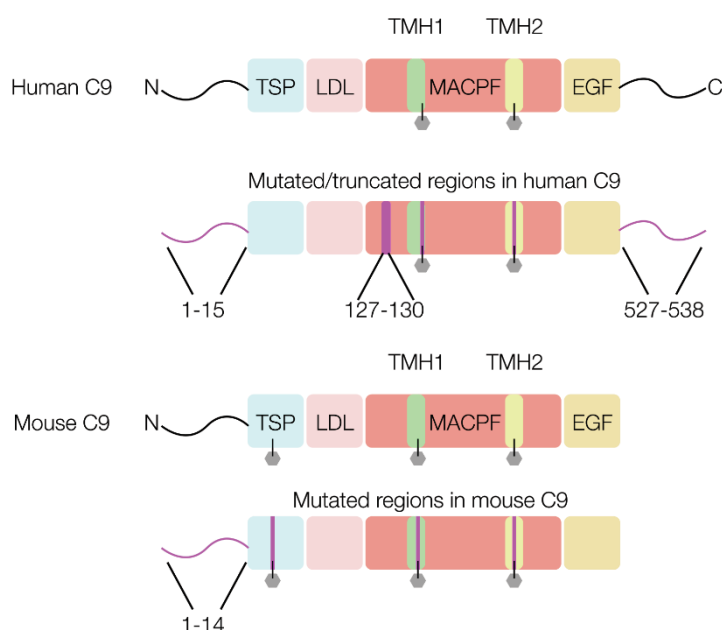
Figure 4-1 C9 topology diagram and regions targeted for mutation or truncation.

The four C9 domains are shown: TSP (blue), LDL (pink), MACPF (red) with corresponding TMH1 (green) and TMH2 (yellow) and the EGF-like domain (orange). N-linked glycans are shown as grey hexagons. The regions that were targeted for mutagenesis or truncation are shown in magenta.

4.2   Methods & Results

The following sections highlight the approach undertaken to develop expression and purification of a recombinant C9 suitable for structural biology and functional studies. These experiments resulted in several key findings, reported in chapter 5, that permitted structural characterisation of C9 in both monomeric (crystal) and polymeric (cryo-EM) form (**Chapter 5**).

4.2.1   Expression and purification of human C9 from insect cells

Historically, C9 expression and purification was initially studied using insect cell expression systems (baculovirus) (125,126,131). This eukaryotic expression system secretes protein containing unbranched N-linked glycans. This contrasts with the mammalian cell based expression system where secreted proteins contain, larger and more complex, branched N-glycans. It was hypothesised that the smaller unbranched N-glycans may be more suited for C9 structural studies since these modifications are often less disruptive with respect to crystallisation (132). To test the secretion of human C9 from insect cells (C9$_{[bac]}$) the Bac-to-Bac expression system (Invitrogen) was deployed using Hi5 cells (See also methods Chapter 2) (**Figure 4-2**). The expression of C9$_{[bac]}$ was performed for 3-days post-infection with baculovirus to minimise the release of non-secreted protein from lytic cells (**Figure 4-2**). The results from this small-scale expression showed a stepwise increase of protein secretion from the Hi5 cells (**Figure 4-2**).
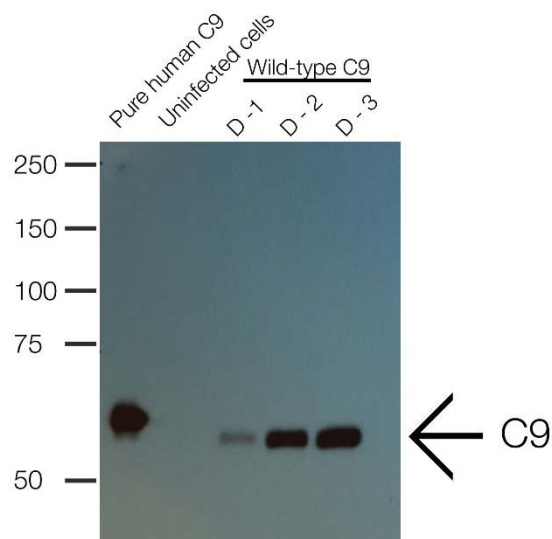
Figure 4-2 Test expression of human C9 from insect cells (Hi5).

Time-course samples of insect media were obtained each day after infection with recombinant baculovirus containing the C9 gene. The samples were resolved by 12% (w/v) non-reducing SDS-PAGE and Western blot analysis performed using goat α-C9 antibody (1:10000 dilution) and rabbit α-goat HRP-conjugated (1:10000 dilution). As shown, the insect cell expressed C9 is slightly smaller than plasma purified C9, due to the smaller N-linked glycans.

Given the success of the small-scale C9 expression, C9[bac] was scaled up for large scale production (typically performed using 1 – 4 litres of cells). The C9[bac] was purified according to protocols reported in previous studies (also described in **Chapter 2**) (131) (**Figure 4-3**). This procedure yielded protein that was an estimated ~70% homogeneous and active in a haemolytic assay (**Figure 4-3**). However, the crystal trials and TEM of this protein were unsuccessful, possibly due to the presence of contaminants.

Figure 4-3 Representative native purification of human C9 from insect cells.

The chromatograms with corresponding 12% (w/v) non-reducing SDS-PAGE shown beneath. Sequential purification by DEAE (top left); CHT (top right) and monoQ (bottom left). The chromatograms show the UV trace (blue) and conductivity (brown). The shaded regions of the chromatographs correspond to the fractions tested by SDS-PAGE; serum purified C9 (sC9), pre-column (PC) and flow through (FT) are also shown. The pooled fractions for the next step are shown with brackets. The final purified C9[bac] contained ~70% homogeneity based on Coomassie stained SDS-PAGE analysis. The haemolytic assay of the C9[bac] was shown to be comparable to serum C9 (bottom right). Control reactions were performed containing no C9-depleted serum.

Next, C9 was expressed with an N-linked His tag (C9[N-his bac]) and was initially affinity purified using Ni-NTA. This approach reduced the amount of contaminants in the purification and resulted in improved purity (~95%) (**Figure 4-4**). However, C9[N-his bac] failed to crystallise, was highly prone to aggregation (**Figure 4-4**) and was unsuitable for TEM studies.
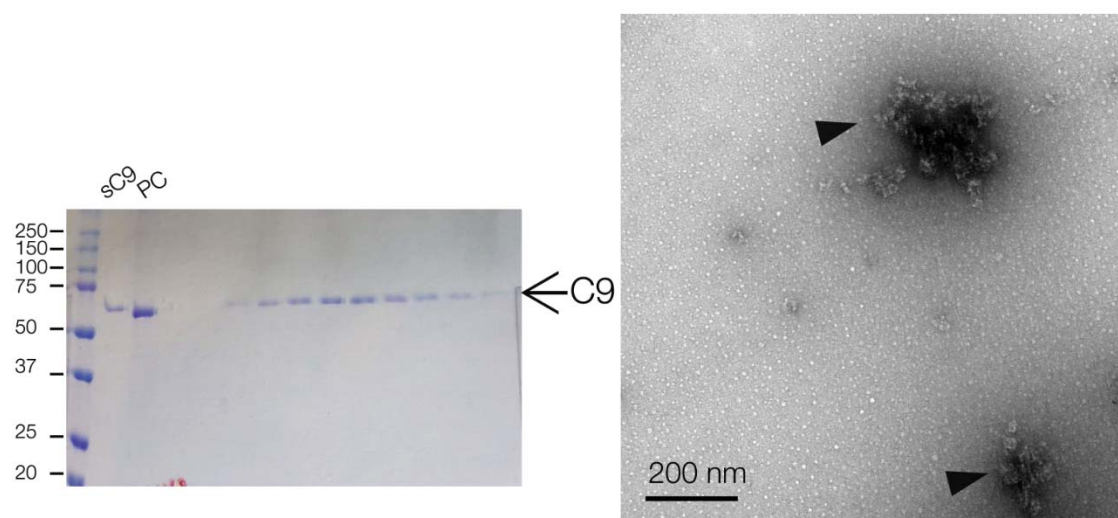
Figure 4-4 Aggregation of N-his polyC9.

SDS-PAGE of purified C9[N-his bac] after the final purification step of superdex 200 (left). The final purified C9[N-his bac] contained ~95% homogeneity. A representative TEM micrograph of C9[N-his bac] His-tag purified C9 protein that was subsequently polymerised to polyC9 rings (right). The polyC9 was induced by incubation at 37° C. Large polyC9 aggregates were typical for the Ni-NTA purified protein (arrows).

It was reasoned that the N-linked glycans in C9 may be adversely affecting the crystal trials. This idea was tested by producing a double mutant T258M/T396M (C9[aglyco]) that had been previously shown to secrete from insect cells (125). These mutations abolish the N-glycan attachment by alteration of the third amino acid of the conserved glycosylation recognition motif (**N-X-T/S**; where **X** can be any amino acid except a proline). The C9[aglyco] was produced with a N-linked His tag for affinity purification, as it was predicted that loss of the N-glycans would affect the processing and transport of

proteins through the Golgi apparatus. Indeed, it was observed that the protein secretion and overall yield was considerably diminished. Additionally, this non-glycosylated form failed to crystallise despite extensive trials.

4.2.2 Expression and purification of human C9 from HEK293F cells

Next, mammalian expression in the HEK293F cell line was tested. This system has the advantage of yielding glycosylated proteins that closely resemble human material. Additionally, numerous advances in cell culture systems have occurred since the first reported mammalian C9 expression system (in adherent COS-7 cells) (123). Newer expression systems include optimised suspension of HEK293F cells (133), which result in higher cell density and can yield milligram quantities of protein in relatively small volumes (~100 mL of culture) (134). Lastly, in the mammalian expression yields can be further increased through improved media components.

The secretion of human C9 (C9$_{[HEK]}$) from suspension cell growth was monitored by obtaining time course samples over five days (**Figure 4-5**). Samples of C9$_{[HEK]}$ expression media were TCA precipitated to assess the total protein amount using Coomassie stained SDS-PAGE (**Figure 4-5**). Secretion of C9$_{[HEK]}$ shown in the time course indicates that C9 is optimally secreted into the expression media (Expi 293 media) between 4-5 days post-transfection. However, the ratio of secreted contaminants was observed to be higher at five days compared to four days (**Figure 4-5**). Therefore, expression protocols were routinely performed for four days. As a general observation, mammalian-based expression also appeared to have improved yields compared to insect cells in the secretion of human C9, with 100 – 400 mL volumes of media yielding sufficient material for structural or functional studies.
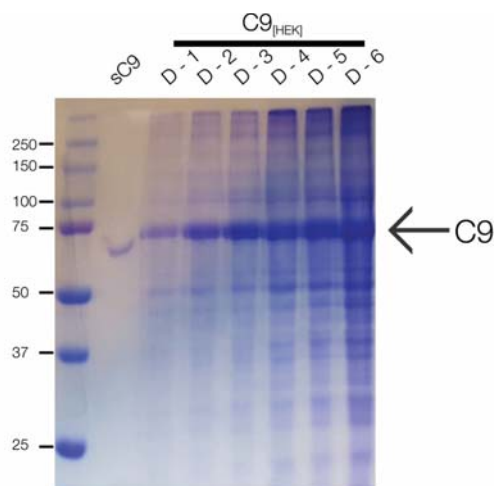
Figure 4-5 Test secretion of human C9 from HEK293F (C9[HEK]) cells.

The HEK293F cells were transfected with C9 plasmids and samples were taken each day of the expression media for TCA precipitation. The precipitant proteins were resolved by 12% (w/v) non-reducing SDS-PAGE which shows expression for human C9 (arrow). The peak secretion was shown to be 4 days which was estimated to contain the highest ratio of secreted C9 to contaminants.

The C9[HEK] was expressed and the protein was purified using the native C9 purification method (**Figure 4-6**). The initial volume of cells used for mammalian protein expression was considerably smaller than that used for insect cell expression. Further, the total number of contaminants that were co-purified was reduced. The resulting purified C9[HEK] was more suitable for TEM experiments (**Figure 4-6** and **Figure 4-7**). The C9[HEK] protein was used for single particle analysis described in **Chapter 5**. Furthermore, the C9[HEK] had a comparable activity to serum C9 in a quick lytic assay as described in methods (section 2.5.7). However, despite the improved yield and purity, the C9[HEK] protein failed to crystallise.

92
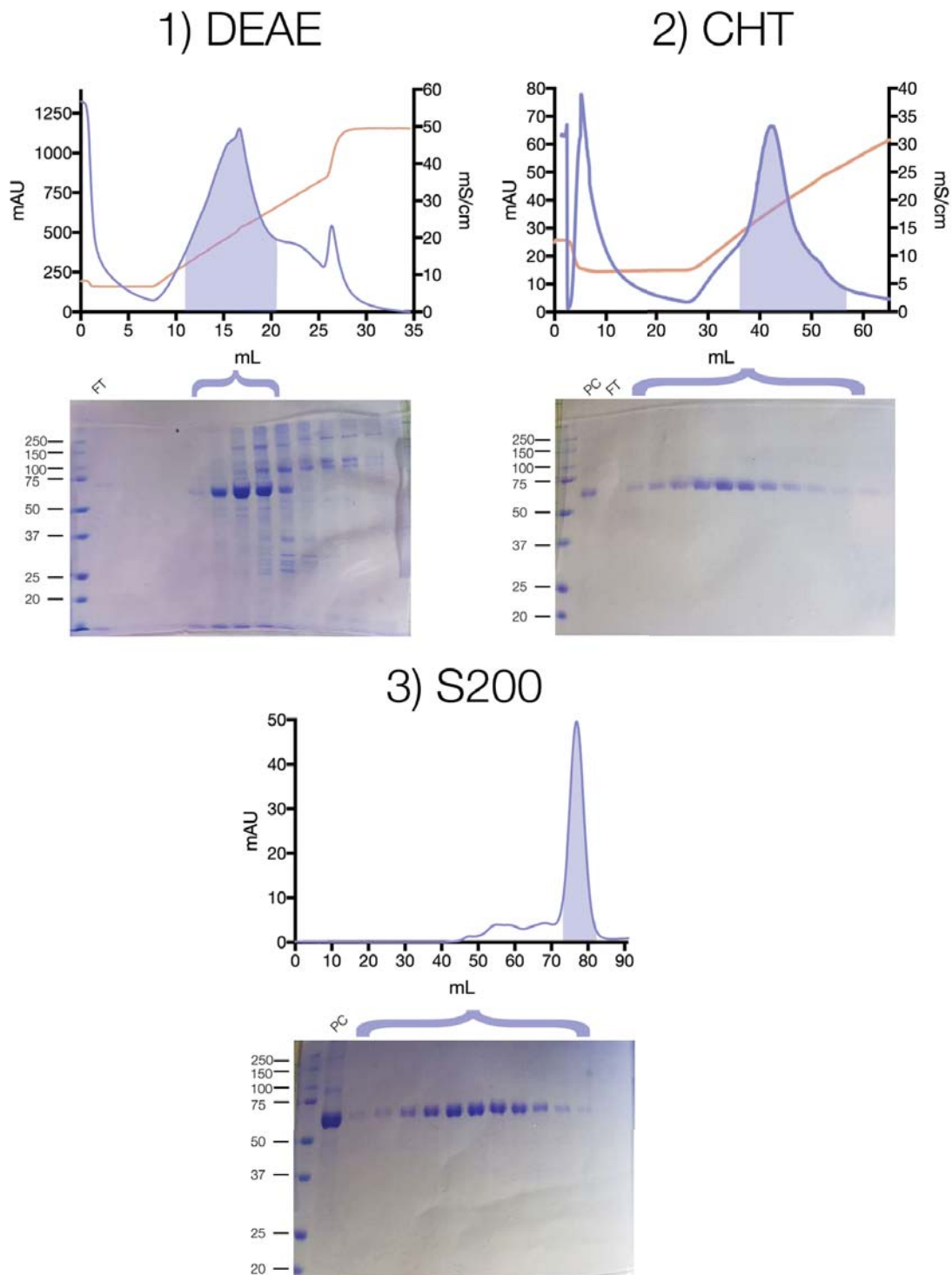
Figure 4-6 Representative native purification of C9[HEK] protein.

The chromatograms with corresponding 12% (w/v) non-reducing SDS-PAGE shown beneath. Sequential purification by DEAE (top left); CHT (top right) and superdex 200 (S200) (bottom). The chromatograms show the UV trace (blue) and conductivity (brown). The shaded regions of the chromatographs correspond to the fractions tested by SDS-PAGE; pre-column (PC) and flow through (FT) are also shown. The pooled fractions for the next step are shown with brackets.
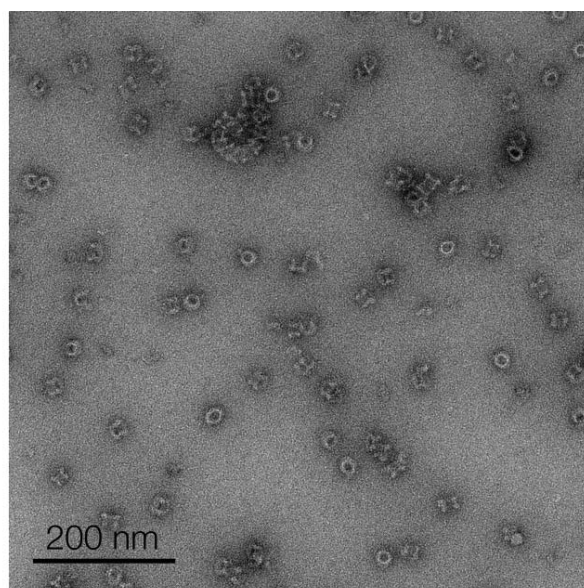
Figure 4-7 PolyC9 rings from wild-type C9[HEK].

*A representative TEM micrograph of C9[HEK] purified by the native purification method. The polyC9 was more suitable compared to the serum C9. This protein was used for single-particle cryo-EM experiments in **Chapter 5**.*

In order to try to overcome the failure of C9 to crystallise, the next approach was to produce a non-glycosylated form of mammalian-derived C9. Two non-glycosylated mutants of C9 were tested for crystallisation, 1) T258M/T396M (C9[aglyco-1] and 2) N256D/N394D (C9[aglyco-2]). As observed in the insect cell expression system, the non-glycosylated C9 exhibited reduced levels of protein secretion, although both mutants could be purified in quantities suitable for crystallography. However, C9[aglyco-1] and C9[aglyco-2] failed to crystallize.

It was reasoned that disordered N-terminal and C-terminal regions in the human C9 may be inhibitory with respect to crystallisation. Accordingly, flexible regions of C9 were identified using limited proteolysis of C9 with the V8 protease *(Staphylococcus aureus)* together with structure guided analysis of sequence alignments of C9 from a range of different species (**Figure 4-8 a** and **b**). The results of limited proteolysis experiments were confirmed by N-terminal sequencing (Edman degradation) (**Figure**

4-8 a). Based on these approaches amino acids 1-15 and 527-538 were identified as being potentially flexible.

Accordingly, a truncated human C9 mutant Δ1-15/Δ527-538 was produced, called C9[ΔN/ΔC]. Moreover, the truncations were also introduced in the non-glycosylated human C9 (T258M/T396M) (C9[aglyco ΔN/ΔC]) backgrounds for crystallisation trials. Both C9 mutants were functionally active as characterised by the lytic assay (section 2.5.7). However, neither C9[ΔN/ΔC] nor C9[aglyco ΔN/ΔC] variants crystallised.
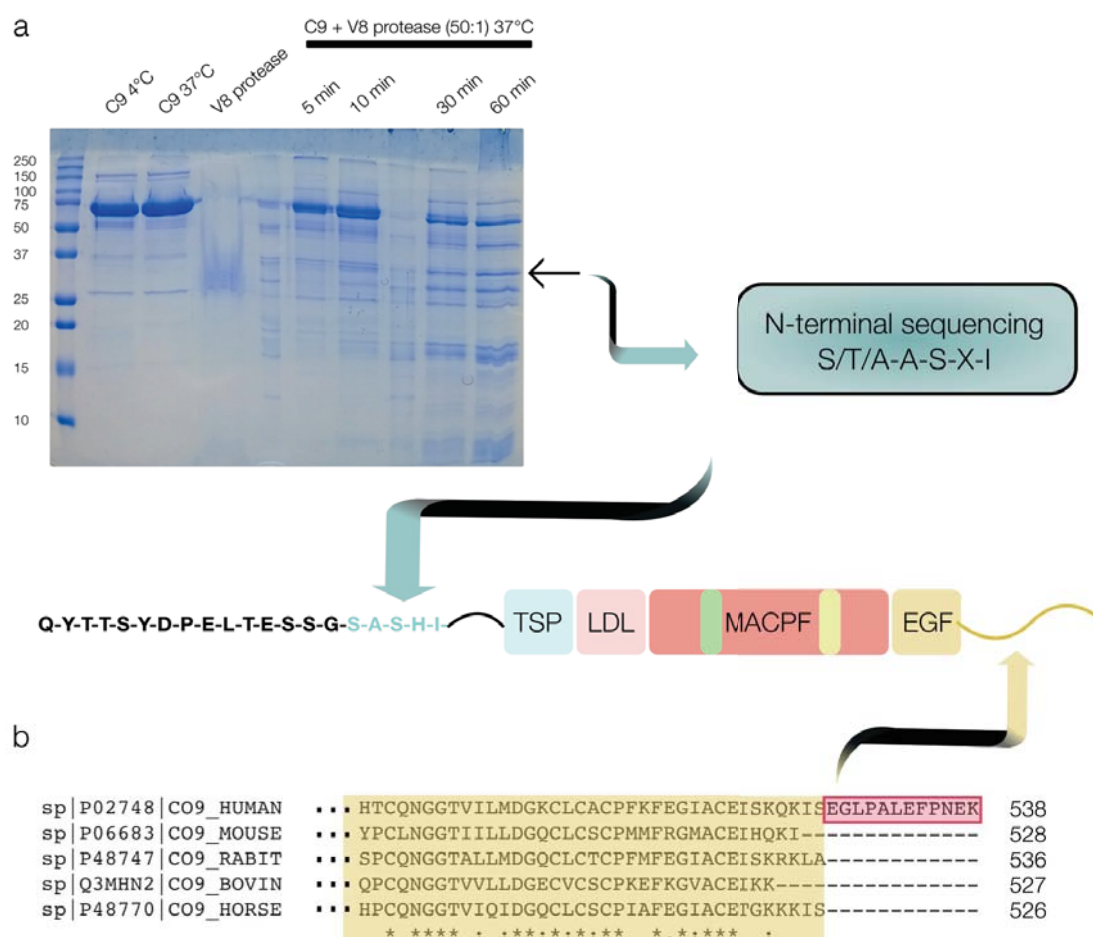
Figure 4-8 Identification of V8 proteolysis fragment of C9 and C-terminal truncation.

a) Plasma purified C9 (1 mg/ml) was treated with V8 protease at a ratio of 50:1 and digested at 37 °C for various time intervals (left). Protein bands for the resultant digests were resolved by 12% (w/v) non-reducing SDS-PAGE, and transferred to PVDF for N-terminal sequencing (right). The corresponding N-terminal sequence from the 35 kDa band was identified as the N-terminus of the secreted C9 sequence. b) sequence alignment of C9 showing the extended C-terminus region of human C9 (red box).

In a final approach to obtain crystals of human C9, point mutations were performed in an area with predicted high surface entropy according to analysis by the surface entropy reduction prediction server (SERp Server) (135). This approach has been used to improve the properties of proteins for crystallisation trials (136). The rationale for this approach is to mutate entropic surface exposed residues (Glu, Lys or Gln) to smaller residues (Ala) thus promoting more favourable crystal contacts. The SERp mutant (C9$_{[SERp]}$) contains three point mutations $_{127}$EESE$_{130}$→$_{127}$AASA$_{130}$. The C9$_{[SERp]}$ variant could be produced and purified, however, it did not crystallise.

4.2.3   Expression and purification of murine C9

As a result of these approaches undertaken to understand the structure of human C9, a closely related homologue, murine C9 (C9$_{[murine]}$), was expressed and purified. C9$_{[murine]}$ was secreted by HEK293F cells, as shown by TCA precipitation, and protein purifications could be achieved from 100 – 400 mL culture volumes (**Figure 4-9**).
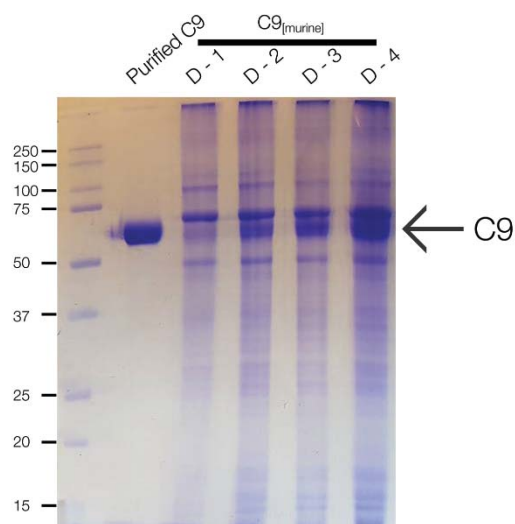
Figure 4-9 Test secretion of C9$_{[murine]}$ in HEK293F cells.

Time course experiment of C9$_{[murine]}$ transfection performed by obtaining the expression media and TCA precipitation resolved on a 12% (w/v) non-reducing Coomassie stained SDS-PAGE. The murine C9 was observed to resolve at a similar molecular weight to human C9 indicating intact protein.

The C9$_{[murine]}$ was purified using the earlier described purification protocol (**Figure 4-10**). This material had similar chromatographic properties to the human C9 form. Similarly, the mouse form of C9 was active with human MAC components in a lytic assay (**Figure 4-10**). However, C9$_{[murine]}$ failed to crystallise.
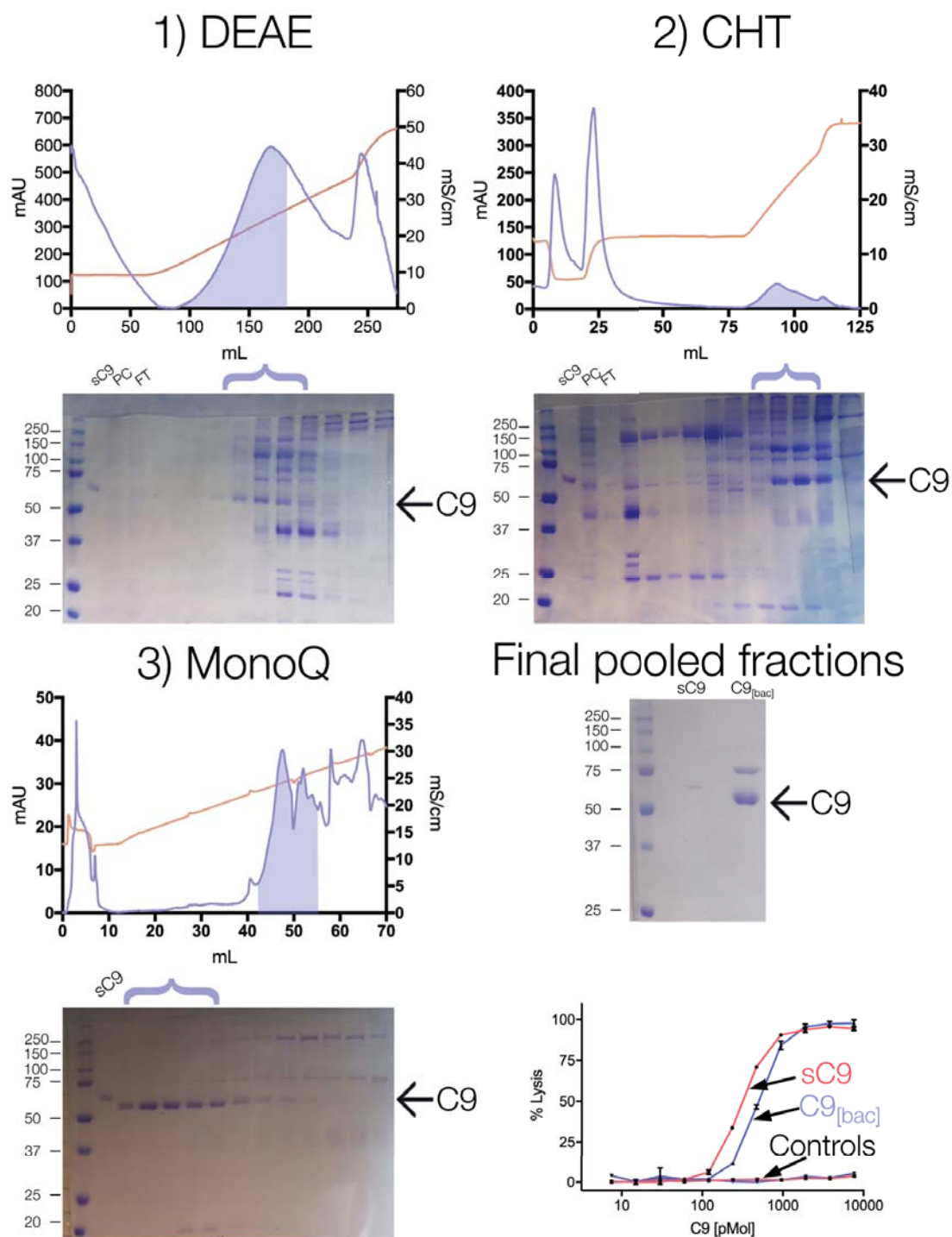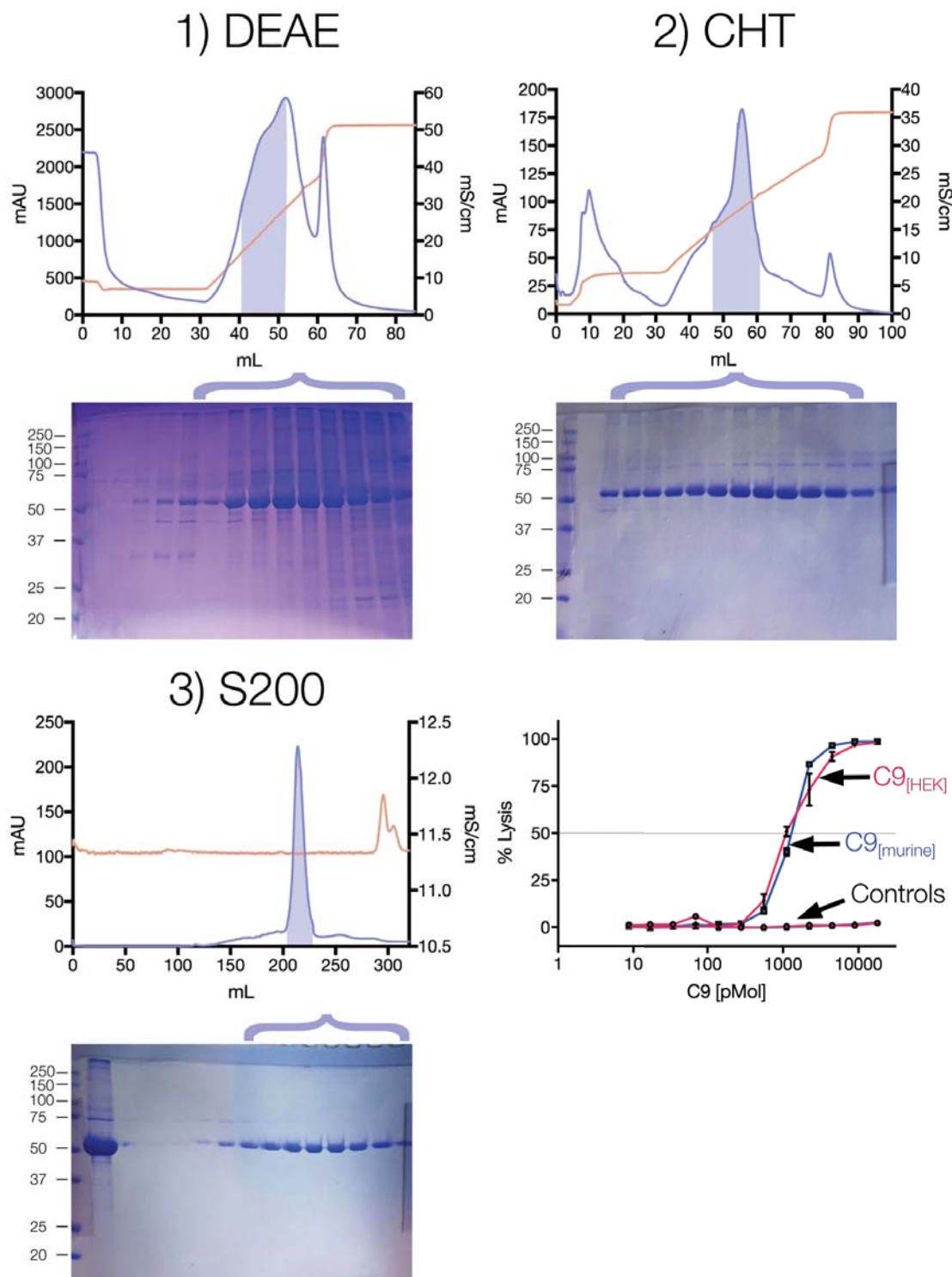
Figure 4-10 Representative native purification of C9[murine]

Chromatograms with corresponding 12% (w/v) non-reducing SDS-PAGE are shown below. The chromatograms show the UV trace (blue) and conductivity (brown). Sequential purification by DEAE (top left); CHT (top right) and superdex 200 (S200) (bottom). The pooled fractions for the next step are shown with brackets. The C9[murine] purification was similar to C9[HEK] in the respective conductivity of elution from both DEAE and CHT. Similarly, C9[murine] elutes from size exclusion at a similar volume to human C9[HEK]. The haemolytic assay of the C9[murine] was determined to be comparable to human C9[HEK] (bottom right).

Next, a non-glycosylated mouse variant (N28E; N243D; N397D) (C9$_{[aglyco\ murine]}$) was expressed and purified to obtain a non-glycosylated form. C9$_{[aglyco\ murine]}$ was secreted by HEK293F cells, albeit in reduced amounts and retained comparable activity to the human form (C9$_{[HEK]}$) (**Figure 4-11**). Excitingly, the C9$_{[aglyco\ murine]}$ material crystallised in several conditions. The crystals were confirmed to contain murine C9 by SDS-PAGE and mass spectrometry (**Figure 4-11**). However, these initial crystals comprised clusters of plates that grew only in 2-dimensions (**Figure 4-11**).
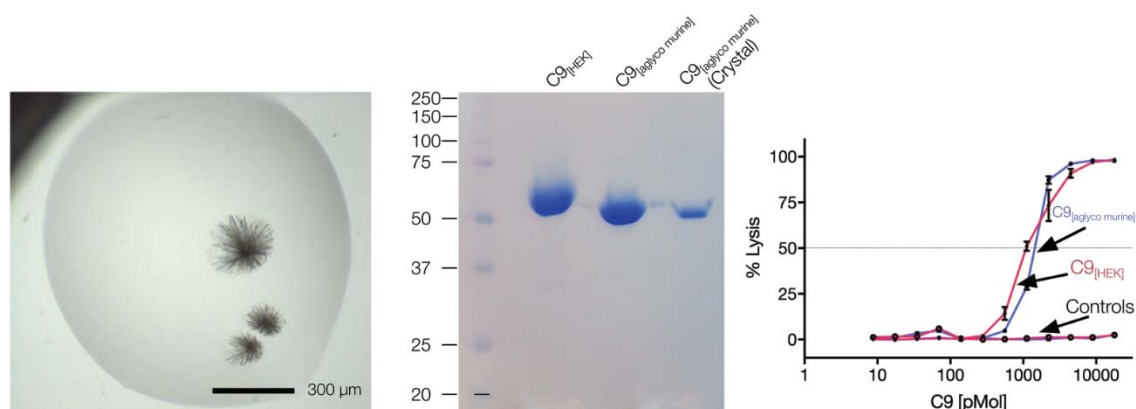
Figure 4-11 Crystals of C9$_{[aglyco\ murine]}$

The C9 crystals were formed using the sitting drop vapour diffusion method, in 20% (w/v) PEG 3350, 0.2 M di-sodium malonate from the PACT crystal screening kit (Molecular Dimensions) (left). A scale bar is shown at the bottom right. Middle, C9$_{[aglyco\ murine]}$ crystals resolved by a 12% (w/v) non-reducing SDS-PAGE. The band from the crystals was extracted for mass spectrometry analysis which confirmed the identity as murine C9 (results in Appendix 1). Right, haemolytic assay of C9$_{[aglyco\ murine]}$ is comparable to human C9$_{[HEK]}$.

In an attempt to improve these crystal forms, an N-terminal truncation (Δ1-14) of C9$_{[aglyco\ murine]}$ was produced (C9$_{[aglyco\ \Delta N\ murine]}$). The truncated form had comparable activity to the human C9 (**Figure 4-12**). This C9$_{[aglyco\ \Delta N\ murine]}$ crystallised in a range of conditions (**Figure 4-12**). These crystal trials yielded small individual crystals, rather than clusters of plates (**Figure 4-12**).
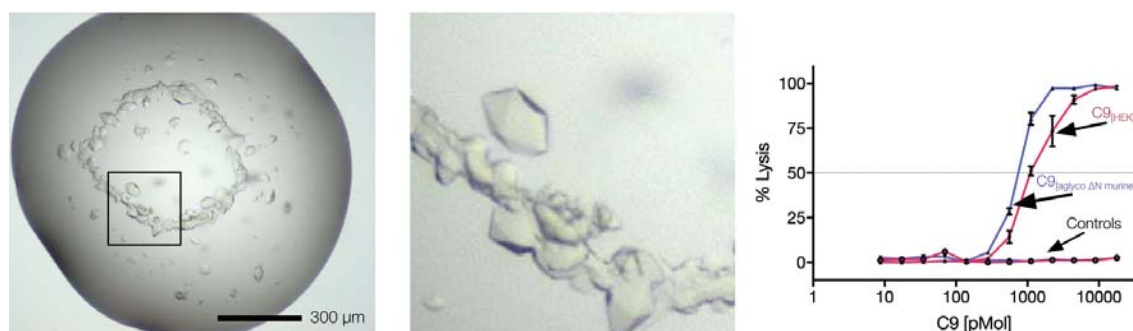
Figure 4-12 Crystals of C9*[aglyco ΔN murine]*.

The C9 crystals were formed using the sitting drop vapour diffusion method, in 10% (w/v) PEG 8000, 0.1 M sodium acetate from the MBClass II Suite (Qiagen). Unlike the non-glycosylated form, the C9[aglyco ΔN murine] formed single small hexagonal crystals. A scale bar is shown at the bottom right. Inset (middle) shows close-up view of single C9 crystals. Right, the C9[aglyco ΔN murine] had comparable activity to human C9[HEK].

4.2.4 Improving the quality of mouse C9 crystals

While the C9[aglyco murine] crystals were promising, it was evident that they were poorly formed (**Figure 4-11**). The C9[aglyco murine] crystals were therefore used to produce a seed stock to nucleate C9 crystal trials (see also **Chapter 5** Materials & Methods). This improved the crystallisation properties of both the C9[aglyco murine] and C9[aglyco ΔN murine]. Furthermore, additive screening identified that $ZnCl_2$ and benzamidine · HCl helped improve the quality of C9[aglyco murine] crystals (**Figure 4-13**).

The final protein crystals used to obtain a 2.2 Å structure were of C9[aglyco murine] (**Figure 4-13**). Notably, the wild-type C9[murine] did not crystallise, even with addition of the C9 crystal seeds, validating the idea that N-linked glycosylation sites hindered crystal formation.
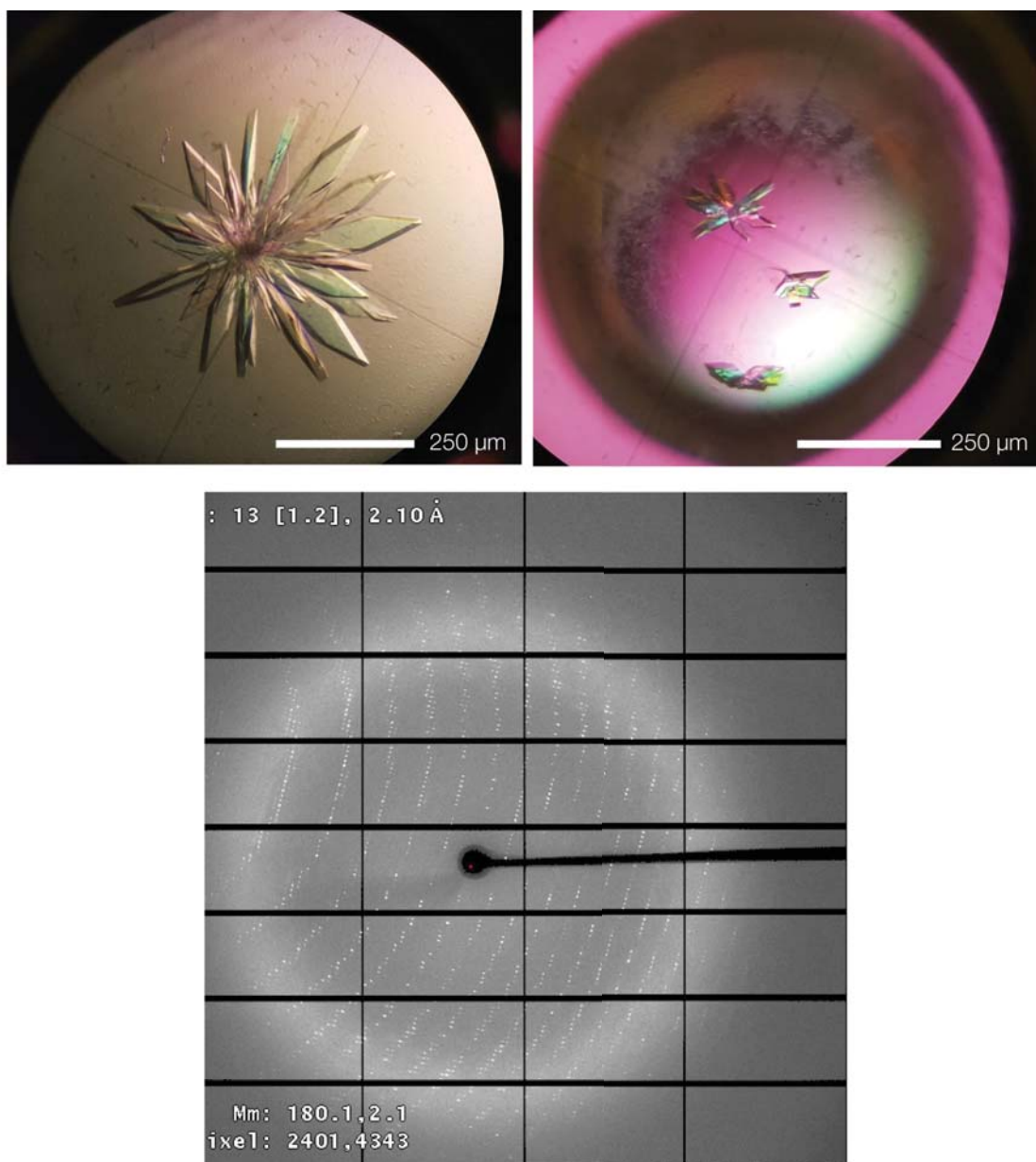
Figure 4-13 Improved C9[aglyco murine] crystals.

The optimised C9 crystals were formed using the hanging drop vapour diffusion method and micro-seeding with the purified C9[aglyco murine]. The crystal condition contained 18% (w/v) PEG 3350, 0.1 M di-sodium malonate with additives: 10 mM ZnCl2 (left) or 10 mM benzamadine · HCl (right). X-ray diffraction of a C9[aglyco murine] crystal (bottom).

4.3 Discussion

The development of a recombinant C9 expression system has been critical for determination of the high-resolution X-ray structure of the precursor form of C9 and the high-resolution structure of polyC9 by cryo-EM (as discussed in Chapter 5). This advance would not have been possible using plasma purified C9, which has undergone

101

extensive crystallisation trial experiments (131). Although we were previously able to obtain a cryo-EM structure of plasma-derived polyC9 with a resolution of 8 Å (106), the use of serum-derived C9 limited both resolution and the ability to test function via mutants. In contrast, the use of expressed recombinant C9 has resulted in crystals that diffract to a resolution of 2.2 Å of the monomeric form of C9 and a 3.9 Å cryo-EM structure of human polyC9 (**Chapter 5**). Critically, these structural data have further yielded key insights with respect to MAC assembly.

Overall, the expression of C9 from HEK293F was found to be higher than insect cells. This resulted in higher homogeneity as smaller starting volumes were required for expression and purification. It was found that both human and mouse C9 proteins could be quickly and reproducibly produced using a purification method without the need for affinity tags (such as 6x histidine tags). This HEK293F cell expression is a serum free media system, which ensures the secreted C9 protein does not contain bovine C9 (derived from fetal bovine serum). These approaches have proven invaluable to the successful determination of multiple important structural forms of C9 and have proven to be an integral component of the studies performed.

In the experiments performed here, only the murine C9 crystallised and only in the non-glycosylated form. Even given this advance, C9$_{[aglyco\ murine]}$ crystals required extensive optimisation through micro-seeding and additive screening. These approaches finally yielded well-diffracting crystals. In addition, the improvements in yield and in purity from HEK293F-expressed C9 were necessary for high resolution cryo-EM studies.

Finally, functional analysis showed that all the recombinant expressed forms of C9 characterised here were active. The mouse C9 had a comparable haemolytic activity to the human C9 form suggesting a common C5b-8 interface is formed in the nascent human MAC that can accommodate oligomerisation of mouse C9 to make a lytic pore. This

observation was initially surprising given that human and mouse C9 proteins share only 58% sequence identity (137,138). However, the areas of lowest sequence identity when comparing mouse and human C9 are in the membrane spanning TMH1 and TMH2 regions. These data suggest that these regions can accommodate evolutionary mutations and retain the lytic activity. Indeed, in these regards, the inhibitor CD59, which targets the TMH2 sequence of C9, is known to be species specific. Collectively, these functional data were of importance in demonstrating that the structure of monomeric murine C9 could be used to cross-compare with and analyse the structure of human polyC9.

CHAPTER 5: THE FIRST TRANSMEMBRANE REGION OF COMPLEMENT COMPONENT-9 ACTS AS A BRAKE ON ITS SELF-ASSEMBLY

The First Transmembrane Region of Complement Component-9 Acts as a Brake on its Self-Assembly

Bradley A. Spicer[1,2], Ruby H.P. Law[1,2], Tom T. Cradoc-Davies[1,3], Sue M. Ekkel[1,2], Charles Bayly-Jones[1,2], Siew-Siew Pang[1,2], Paul J. Conroy[1,2], Georg Ramm[1,2], Mazdak Radjainia[4], Hariprasad Venugopal[1,2], James C. Whisstock[1,2,5,6] & Michelle A. Dunstone[1,2]

[1]Centre for advanced molecular imaging – An ARC centre of Excellence, Monash Node, Clayton, Victoria, Australia, 3800.

[2]Biomedince Discovery Institute, Department of Biochemistry and Molecular Biology, School of Biomedical Sciences, Monash University, Clayton, Victoria, Australia, 3800.

[3]Australian Synchrtron, 800 Blackburn Road, Clayton, Victoria, Australia 3168.

[4]FEI, Achtseweg Noord, Eindhoven, The Netherlands.

[5]EMBL Australia, Monash University, Melbourne, Australia.

[6]South East University-Monash Joint Institute, Institute of Life Sciences, Southeast University, Nanjing, China.

✳To whom correspondence should be addressed: Michelle A. Dunstone, Department of Biochemistry and Molecular Biology, School of Biomedical Sciences, Monash University, Clayton, Victoria, 3800, Australia.

Tel: +61 3 9902 9269; Email: michelle.dunstone@monash.edu

The first transmembrane region of complement component-9 acts as a brake on its self-assembly

Bradley A. Spicer[1,2], Ruby H.P. Law[1,2], Tom T. Caradoc-Davies[1,3], Sue M. Ekkel[1,2], Charles Bayly-Jones[1,2], Siew-Siew Pang[1,2], Paul J. Conroy[1,2], Georg Ramm[1,2], Mazdak Radjainia[4], Hariprasad Venugopal[1,2], James C. Whisstock[1,2,5,6] & Michelle A. Dunstone[1,2]

Complement component 9 (C9) functions as the pore-forming component of the Membrane Attack Complex (MAC). During MAC assembly, multiple copies of C9 are sequentially recruited to membrane associated C5b8 to form a pore. Here we determined the 2.2 Å crystal structure of monomeric murine C9 and the 3.9 Å resolution cryo EM structure of C9 in a polymeric assembly. Comparison with other MAC proteins reveals that the first transmembrane region (TMH1) in monomeric C9 is uniquely positioned and functions to inhibit its self-assembly in the absence of C5b8. We further show that following C9 recruitment to C5b8, a conformational change in TMH1 permits unidirectional and sequential binding of additional C9 monomers to the growing MAC. This mechanism of pore formation contrasts with related proteins, such as perforin and the cholesterol dependent cytolysins, where it is believed that pre-pore assembly occurs prior to the simultaneous release of the transmembrane regions.

[1] ARC Centre of Excellence in Advanced Molecular Imaging, Department of Biochemistry and Molecular Biology, 23 Innovation Walk, Monash University, Victoria, 3800 Australia. [2] Biomedicine Discovery Institute, Department of Biochemistry and Molecular Biology, 23 Innovation Walk, Monash University, Victoria, 3800 Australia. [3] Australian Synchrotron, 800 Blackburn Road, Clayton, Victoria, 3168 Australia. [4] Achtseweg Noord 5, Building, 5651 GG Eindhoven, The Netherlands. [5] EMBL Australia, Monash University, Melbourne, Australia. [6] South East University-Monash Joint Institute, Institute of Life Sciences, Southeast University, 210096 Nanjing, China. These authors contributed equally: Bradley A. Spicer, Ruby H. P. Law. Correspondence and requests for materials should be addressed to J.C.W. (email: James.Whisstock@monash.edu) or to M.A.D. (email: Michelle.Dunstone@monash.edu)

The MAC represents the terminal portion of the complement system, and functions to form large pores in the membrane of target bacteria, enveloped viruses and parasites[1,2]. Currently, it is suggested that the MAC assembles on the target membrane via sequential addition of five different components (C5, C6, C7, C8 [comprising C8α C8β, and C8γ], and C9). The final stage of MAC formation involves addition of multiple copies of the pore-forming component, C9, to the C5b8 complex. Together these proteins form an asymmetric pore (Supplementary Fig. 1)[3,4].

C6, C7, C8α, C8β, and C9 all contain a membrane attack complex/perforin/cholesterol dependent cytolysin (MACPF/CDC) domain (Supplementary Fig. 2). This domain is generally associated with a pore forming function in a wide variety of different toxins and immune defence proteins[5]. Previous work reveals that the mechanism of MACPF/CDC pore formation involves three steps[6–8]. First, soluble monomers are recruited to the membrane. Next, between 10 and 50 membrane associated molecules then laterally migrate and self-assemble into a circular or arc pre-pore form[8]. Finally, a conformational change in two regions (named TMH1 and TMH2 because of the β-hairpin conformation each region finally adopts in the membrane) results in formation of an unusually large, membrane spanning β-barrel pore (Supplementary Fig. 1)[9,10]. Each subunit contributes two membrane spanning β-hairpins. Further, based on AFM studies of CDCs and structural studies on the fungal MACPF toxin pleurotolysin, it is postulated that membrane insertion of all membrane spanning regions occur in a simultaneous fashion[8,11].

Most MACPF/CDC proteins, such as CDCs[8] and perforin[12], involve a single protein that can self-associate, usually only in the context of having bound first to a lipid membrane. In contrast, the MAC is unusual in that it is initiated by a non-MACPF domain protein, C5b, which then allows the sequential binding of single units of the MACPF-domain containing proteins C6, C7 and C8 complex (C8αβγ). This assembly (C5b8) then allows the binding of multiple units of C9 that form the final ring-shaped pore (Supplementary Fig. 1a). C9 can also form a homogenous ring in vitro, called polyC9, that closely resembles the assembly of C9 in the MAC[13].

Previous studies reveal that the MAC assembles via sequential recruitment of each component from the soluble phase onto the growing, membrane associated complex in a unidirectional manner[14]. Each component of the MAC thus contains a binding surface and an elongation surface (Supplementary Fig. 1b). Once an individual component is associated with the nascent MAC, its elongation surface is activated (presumably via a conformational change) such that it can now interact with the binding surface of the next soluble component to join the complex (Supplementary Fig 1). The final component of the MAC, C9, is the only component of the assembly that can self-associate—an event that completes the structure of the pore. The complete MAC contains ~18 C9-monomers in the full assembly.

Currently, given the ability of C9 to self-associate, it is unknown how aberrant oligomerisation is prevented in the solution phase prior to binding the C5b8 complex. To address this question, we determined the 2.2 Å X-ray crystal structure of monomeric murine C9 together with the 3.9 Å resolution cryo EM structure of poly-C9. These data show that TMH1 functions in monomeric C9 to sterically inhibit its self-assembly in the absence of C5b8. Binding of C9 to C5b8 results in a conformational change in TMH1 and entry of this region into the membrane. This transition, which is likely accompanied by release of TMH2 and movement in a conserved helix-turn-helix (HTH) structure, permits sequential and unidirectional recruitment of the next C9 monomer into the growing MAC.
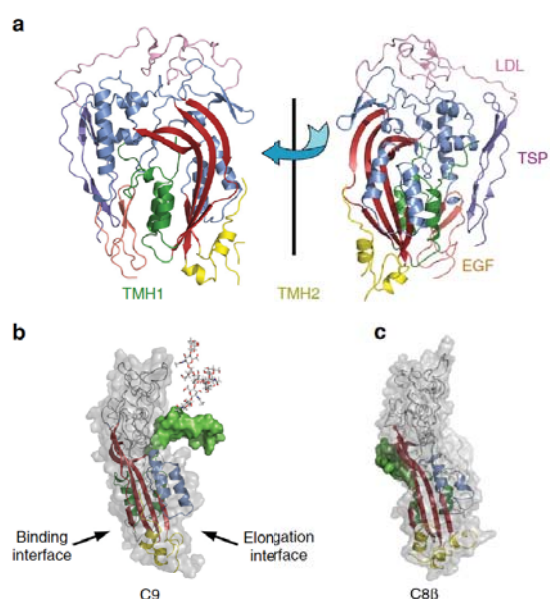
Fig. 1 The X-ray crystal structure of complement component 9. **a** The X-ray structure of C9 shown in cartoon in two orientations, rotated 180° apart. The bent β-sheet of the MACPF domain is shown in red with α-helices in blue, TMH1 (green) and TMH2 (yellow). The ancillary domains: TSP1 (purple), LDLRA (pink) and EGF domain (orange). Domain colours also match the colours used to show the domain features in Supplementary Fig. 2. **b** Cartoon model of C9 with the modelled TMH1 loop (green surface) and N-glycan (PDB ID 1HD4) located on the elongation face of the protein. The key features of the MACPF domain are shown as cartoon and coloured as follows: central β-sheet (red), TMH 1 (green), TMH2 (yellow), HTH (blue). **c** The C8β structure in the same orientation as C9 showing the TMH1 domain on the docking interface (PDB ID 3OJY)

Results

The X-ray crystal structure of monomeric C9. The structure of C9 reveals that the four domains (Thrombospondin type 1 [TSP1], Low density Lipoprotein Receptor Type A [LDLRA], MACPF/CDC and Epidermal growth factor [EGF], Supplementary Fig. 2) are arranged into a globular bundle. Structural comparisons with other MAC proteins (e.g., C8β) reveal that the overall arrangement of domains is similar (Supplementary Fig. 3) except for a striking difference in the position of TMH1 with respect to the core body of the molecule (Fig. 1b, c)[14,15].

In the structure of C6, C8α, and C8β, TMH1 is arranged such that it does not obviously obstruct binding to the next subunit. Indeed, the structure of the complex between C8α and C8β reveals that the TMH1 of the monomeric C8α is buried within the interface[15]. The interface of each of these proteins is relatively flat and currently the precise nature of the conformational changes that take place in order for a new molecule to join the assembly remains to be completely understood. In these regards, we and others, have suggested that a structurally conserved Helix-Turn-Helix motif that sits on top of TMH2 represents the major component that must shift during pre-pore assembly[11,13,16,17].

The monomeric C9 crystal structure reveals that a large proportion of TMH1 (a portion of which is flexible and cannot be resolved in electron density) is located in the centre of the elongation surface where it would be anticipated to block binding
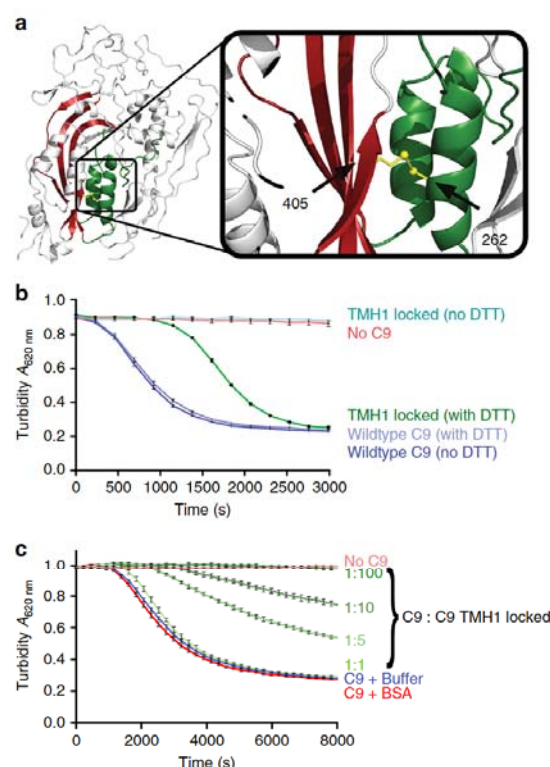
Fig. 2 The C9 TMH1 movement is necessary for pore assembly. **a** Cartoon model of a C9 monomer (left) disulphide locked mutant (also called C9$_{mutant}$ with F262C/V405C mutations) (shown as yellow sticks), that links the TMH1 region to β-strand 4 of the MACPF domain (right). **b** Haemolytic activity of disulphide locked C9 against erythrocytes/antibody/complement 1–8 (EAC1-8). The TMH1 locked (no DTT) alone is inactive; however, activity can be rescued with 1 mM DTT (TMH1 locked (with DTT)). Also shown are control experiments: no C9, and wildtype C9 (with and without DTT). **c** Competition assay of disulphide locked mutant with wildtype C9 showing that the disulphide trapped variant competes for the elongation face with wild-type C9. A range of ratios of wildtype C9 and C9 TMH1 locked mutant used in the assays are as shown and it reveals that the disulphide locked C9 competes for the nascent MAC and stalls assembly in a dose-dependent manner. Also shown are no C9, C9 in buffer and C9 plus BSA controls. The results (**b** and **c**) are presented as the averaged turbidity measurements from three independently prepared samples ($n = 3$) with error reported as the standard error of the mean (SEM). See also Supplementary Fig. 7 for more detail

of another C9 monomer (Fig. 1b). Interestingly, the flexible region of TMH1 is the least conserved region across all vertebrate species and thus may represent a site under significant evolutionary pressure, for example, as a site of MAC inhibition by bacteria[18]. In addition, it is notable that TMH1 contains an N-glycan (found in most species) on residue N243 (human equivalent N256), a modification that would add additional bulk to this region. This finding suggests that TMH1 may function to block self-assembly, and that a key event of the interaction between C9 subunits would be a conformational

change of TMH1 such that it moves to reveal the C9 elongation surface.

Mobility in TMH1 is essential for C9 self-association. To investigate the hypothesis that TMH1 blocks self-assembly, we designed a disulphide trap mutant (F262C/V405C, [C9$_{mutant}$]) that linked TMH1 to β-strand 4 of the MACPF/CDC domain (Fig. 2a). Time resolved haemolytic assays revealed that the disulphide-trapped C9 variant is completely inactive with respect to lytic function and that addition of reducing agent resulted in restoration of lytic function (Fig. 2b). Crucially competition assays further reveal that the C9$_{mutant}$ competes with wild type C9 and thus is competent to bind the C5b8 or C5b89$_n$ complex to form C5b89$_{mutant}$ or C5b89$_{n+mutant}$ respectively (Fig. 2c). Together these data suggest that the sequential addition of C9 molecules to C5b89$_1$ relies on a rearrangement in TMH1.

The cryo-EM structure of poly C9. To further investigate the structural transitions associated with C9 self-assembly, we determined the 3.9 Å resolution cryo EM structure of polyC9 (Fig. 3a), the highest resolution structure to date of any MACPF or CDC protein in the pore form. PolyC9 mimics the form seen in the complete MAC[13] and is formed in vitro following prolonged incubation of concentrated C9 at 37 °C.

The resolution of the polyC9 map ranges from 3.2 to 4.4 Å, with the best resolution present at the top of the β-barrel around the HTH region (Supplementary Figs. 4, 5). As previously reported from analysis of our lower resolution (8 Å) structure[13], the final structure of human polyC9 reported in this study contains 22 monomers. The improved resolution of the maps permitted construction and refinement of a full atomistic model. In this model, we were able to unambiguously assign 460/528 residues of main-chain atoms. Clear electron density was observed for side chains located at the oligomer interface and around the HTH region (described below). Further, and in regards to the remarkable 88 stranded β-barrel itself, our data were of sufficient quality to reveal individual strands within the assembly, providing experimental evidence that the β-barrel adopts the $S = n/2$ architecture as predicted by bioinformatics analysis[19].

Analysis of the polyC9 model revealed charge complementarity between the elongation face and binding face of each subunit (Supplementary Fig. 6). A total of 91 contacts are made at the interface between subunits (Supplementary Table 1). Six of these interactions involve the TSP1 domain, which plays an important role in pore assembly and is intercalated around the outer edge of the ring[13,20].

The molecular transitions that control MAC assembly. A comparison with the polyC9 structure with the monomeric C9 form revealed that TMH1 must move to expose the elongation face for an additional C9 monomer to bind to the growing assembly (Fig. 3b). In addition, these structural comparisons reveal that the HTH region must also be substantially repositioned to permit binding of the next C9 subunit.

In the monomeric structure of C9 the HTH packs against the underlying β-sheet as well as part of TMH2 (Fig. 3c). In polyC9, however, both TMH1 and TMH2 are released and, as a consequence, the HTH region has moved such that it partially occupies the position vacated by TMH2 (Fig. 3c). Analysis of side chain interactions revealed that, in both the monomer and pore structures the HTH is loosely packed against the surrounding structures, and makes mainly hydrophobic bonds (e.g., Fig. 3d, e). These data are consistent with this region being able to readily move in response to conformational change in TMH2. The HTH

108
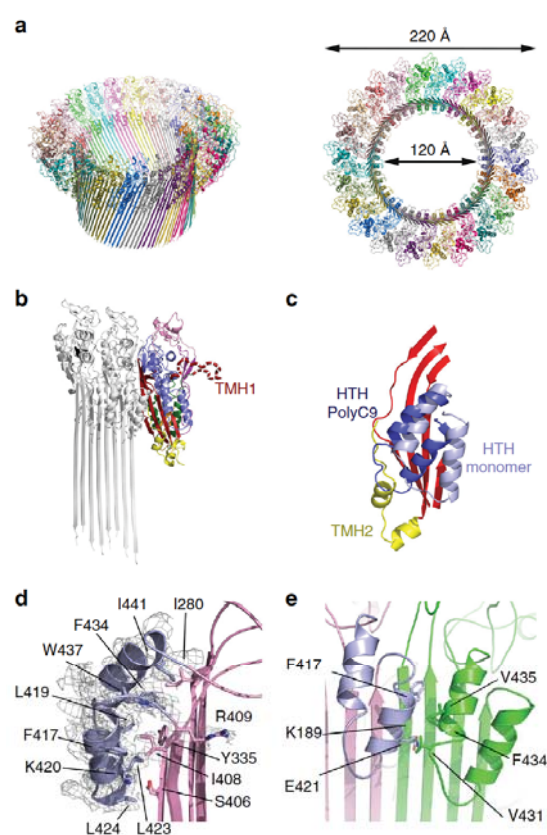
Fig. 3 C9 structure in the monomeric and assembled forms. **a** The cryo EM structure of polyC9 with 22 subunits (different colours) in a circular assembly, two orientations shown, oblique-view and top-down. The resolved strands in the β-barrel conform to the $S = n/2$ architecture[19]. The model excludes the membrane spanning region due to the lower resolution (Supplementary Fig. 5a). **b** Cartoon representation of C9 monomer (colour) docked to a previously unfurled assembled C9 dimer (grey, left). The position of TMH1 as well as the HTH blocks the elongation face. **c** Relative positions of the HTH of polyC9 (dark blue) and monomeric C9 (light blue). The central β-sheet of the MACPF domain is also shown (red). In polyC9, the HTH partially occupies the region vacated by the TMH2 α-helices (yellow). **d** Zoom in view of the HTH domain in the polyC9 EM map, key residues found in the interface between HTH and β-barrel are shown in sticks and labelled. **e** Cartoon representation of two HTH regions, plus key residues in the interface, from neighbouring C9 molecules in the polyC9 structure

further makes new inter-subunit interactions in polyC9 (Fig. 3e) such that it forms a continuous band of α-helices that line the top of the β-barrel lumen (Fig. 3a). Outside of these regions, and in the context of the elongation face of C9, only modest shifts of individual rigid bodies and domains are required to permit MAC polymerisation (Supplementary Table 2), for example, the TSP1 translates with respect to the EGF domain by ~2 Å.

Discussion

Previous data suggest that both TMH1 and TMH2 of C8 fully enter the membrane prior to recruitment of C9[21]. This finding is consistent with the observation that incomplete arc-like structures can form and penetrate the membrane[4,8]. In such a structure (where C8 is fully inserted) the edge strand at the elongation face is the TMH2 of C8α (Supplementary Fig. 1). Our biochemical data reveal that a C9 variant in which TMH1 is disulphide trapped is able to join C5b8, however, further elongation is not possible without the release of TMH1. We hypothesise that upon binding to C5b8 the most likely next step is for the TMH1 of C9 to add to the nascent barrel structure by forming a canonical β-hairpin with the membrane inserted TMH2 of C8α (Fig. 4). Alternatively, it is possible that TMH1 moves sufficiently to permit C9 binding, but without inserting into the membrane. However, we have no evidence for such an intermediate pre-pore like state.

We further suggest that prior to the next C9 subunit joining the assembly, it is highly likely that the TMH2 of C9 is also released to enter the membrane, and that this permits the HTH region to slide across the underlying β-sheet. The removal of TMH1 together with the shift in the HTH region will expose the elongation face of C5b89₁ and permit recruitment of the next C9 subunit into the growing MAC. Taken together these data explain how the MAC has evolved a mechanism of coupling sequential insertion with elongation. The mechanism of C9 pore formation also directly contrasts that of related molecules such as perforin and the CDCs, where it is suggested that the assembly of pre-pores (or pre-pore-like arcs) takes place prior to the simultaneous release of the membrane spanning regions[6–8,12,22].

Methods

109

Fig. 4 Schematic diagram of the unidirectional C9 assembly. We hypothesise that during the assembly, binding to the elongation face of a C9 subunit leads to the release of TMH1 which inserts to form a canonical β-hairpin. Following this, the release of TMH2 and a conformational change in the HTH region uncovers the elongation face of the newly assembled C9, allowing the next C9 subunit to join the assembly

X-ray data collection and model building. The data were merged and processed using XDS[24,25], POINTLESS[26,27] and AIMLESS[28]. Five percent of the datasets were flagged as a validation set for calculation of the R_{free} with neither a sigma, nor a low-resolution cut-off applied to the data. Experimental phases (Supplementary Table 4) were obtained by the MIRAS (multiple isomorphous replacement plus anomalous differences) method. Molecular replacement was attempted using the MACPF domain of C6 (PDB ID 3T5O) and with both MACPF domains of C8 (PDB ID 3OJY). None of the MR experiments were successful. The Ta and U heavy atoms were not ordered and the structure was phased using the anomalous signal of the Zn and Ca ions bound to the protein. Two datasets collected at 10,300 eV (which is above both the K-edge of Zn and L-III edge of Ta), were used as derivative 1 and derivative 2 datasets (Supplementary Table 4). Experimental phasing strategies and dataset combinations were evaluated using HKL2MAP[29] and final phasing was carried out using the CRANK2 pipeline[30]; heavy atom positions were located using SHELXC/SHELXD[31], substructure refinement was done using BP3[32]. The initial FOM (figure of merit) from phasing was 0.26 and after density modification with PARROT[33] this increased to 0.57. Automated model building was carried out using BUCCANEER[34] with the initial model consisting of 944 residues with R/R_{free} of 34.1/40.0%. Two molecules were found per asymmetric unit. Model building was performed using COOT[35] while refinement was performed using PHENIX[36], REFMAC[37], and autoBUSTER[38]. Water molecules were added to the model when the R_{free} reached 30%. Crystallographic and structural analysis was performed using CCP4 suite[39] unless otherwise specified. All Zn atoms were modelled into the omit-map generated using ANODE[40] from a dataset collected at 9674.0 eV (1.28162 Å), above the K-edge of Zn (9659.0 eV), and confirmed by the absence of anomalous signal at the Zn sites in a dataset collected below the Zn K-edge at 9643.9 eV (1.28562 Å) (Supplementary Table 5). Figures 1a, b, and 2a; Supplementary Fig. 3; Supplementary Fig. 4 were generated in part using PYMOL and Chimera[41]. The final model contains two chains: chain A is less flexible with residues 18–226, 248–365, 395–526 modelled into the electron density; chain B residues 18–73, 78–113, 116–205, 214–225, 249–364, and 395–526 modelled. In the final model, the number of residues in the Ramachandran favoured region is 873 residues (out of a total of 874 residues). Structural validation was performed using MolProbity[42]. The MolProbity score is 0.87 which is in the 100th percentile of structures reported at this resolution.

PolyC9 preparation and data collection. Mammalian cell expressed human C9 (with the two native N-glycans) was buffer exchanged by dialysis into 10 mM HEPES pH 7.5, 50 mM NaCl overnight at 4 °C at a concentration between 100 and 250 μg mL⁻¹. Following dialysis, the human C9 was concentrated between 1.1–1.5 mg mL⁻¹ with a 30 kDa MWCO (Amicon) protein spin filter and 1:9 (v/v) of amphipol A8-35 (Anatrace) was added to a final concentration of 0.015–0.02 mg mL⁻¹. Polymerisation reactions were initiated by incubating at 37 °C overnight and stored at 4 °C.

The polyC9 reaction producing the best grid was from an initial C9 concentration of 1.3 mg mL⁻¹ containing 0.02 mg mL⁻¹ A8-35. Plunge-freezing was performed using a Vitrobot Mark IV (FEI/Thermo Fisher Scientific). PolyC9 (2.5 μL) was added to a freshly glow discharged Quantifoil copper grid (R1.2/1.3, 200 mesh). Data was collected on a Titan Krios (FEI/Thermo Fisher Scientific) operated at 300 kV at a magnification of 130 K in microprobe EFTEM mode, resulting in a magnified pixel size of 1.06 Å pixel⁻¹. The movies were collected using a Gatan K2 Summit with a quantum energy filter in super resolution mode (for an effective pixel size of 0.53 Å pixel⁻¹). Each movie consists of 20 sub frames and the exposure time was 8 s which amounted to a total dose of 46.4 e⁻ Å⁻² at a dose rate of 6 e⁻ Å⁻² s⁻¹.

Cryo-EM data processing. Unless stated otherwise all processing was performed with RELION (v2.1b.1)[43]. Movies were down sampled in Fourier space by a factor of 2 and summed after correction of beam-induced motion by MotionCor2[44]. CTF estimation was performed by CTFFIND4.1[45] and micrographs with

ice contamination were discarded by visual inspection of the power spectra. Initially ~1000 particles were manually picked and subjected to reference-free 2D classification to serve as templates for auto picking. A total of ~220,000 particles were extracted from summed micrographs and subjected to multiple rounds of 2D classification. A representative subset of class averages was selected for initial model generation in EMAN2.2[46] using the common line method. The initial model was low pass filtered to 20 Å and particles were subjected to 3D classification, giving rise to two classes of C22 and C21 symmetry. Initial refinement with C22 symmetry led to a 4.2 Å map. These initial refinements were used to create a solvent mask, which was low pass filtered to 15 Å for subsequent refinements. This final subset of 58,000 particles was selected for masked movie refinement and particle polishing with C22 symmetry, where the MTF of the detector was used to determine a b-factor of −180 Å². High resolution features were enhanced by sharpening with this b-factor for the purposes of map visualisation. The global resolution was estimated by the Gold Standard 0.143 criterion when comparing the Fourier shell correlation between two independent half maps[47]. The local variation of resolution was further analysed using blocres using a search box size of 20 voxels and FSC criterion of 0.5[48].

PolyC9 model building and model validation. Model building of the polyC9 was performed in COOT (0.8.8)[49]. The crystal structure of murine C9 was manually positioned into the best density of the cryo-EM map and rigid body fitting of individual domains was performed. The TMH1 and TMH2 regions, which significantly alter conformation, were removed for manual building. Non-conserved amino acids were mutated from murine to human residues and their side chains manually positioned to maximise fit in the map. Following initial model building, C22 symmetry was applied to the single subunit of polyC9 using Chimera (UCSF, USA)[50] and further real space refinement performed in COOT to minimise clashes between subunits and improve the overall geometry.

The final three-dimensional model of polyC9 was refined into the cryo-EM map using the phenix.real_space_refine programme within PHENIX suite to optimise and correct for poor geometry (Supplementary Table 6)[36]. During the refinement, standard restrains for covalent geometry, Ramachandran plot and internal molecular (NCS) symmetry were imposed. In addition, secondary-structure restrains were defined for the β-barrel region of the pore (β-strands 186–216, 251–281, 333–363, 379–409) because the map quality towards the end of the pore is of lower resolution. Protein Interactions Calculator[51] was used to calculate intermolecular contacts between adjacent molecules of polyC9 (Supplementary Table 1).

Haemolytic assay. Turbidity measurements were performed using sheep EAC1-8 prepared in DHB++ pH 7.4 (Dextrose HEPES Buffer; containing 2.5% (w/v) D-glucose, 5 mM HEPES, 71 mM NaCl, 0.15 mM CaCl₂, 0.5 mM MgCl₂). EAC1-8 was produced by sensitising 6.5×10^8 cells mL⁻¹ sRBC with equal volume of anti-sheep antibody (0.75 mg mL⁻¹) (Rockland immunochemicals cat no. C220–0002) at 30 °C. Sensitised cells were washed 2 min at 3220×g by centrifugation, and then C9-depleted serum (Complement Tech) added in batch with 1 μL per 3.75×10^6 cells and incubated at 37 °C for 30 min. The absorbance at 620 nm was continuously measured while incubating at 37 °C with intermittent orbital mixing. For unlocking experiments three independently prepared dilutions of C9 were combined with EAC1–8 (3.75×10^6 cells) and a final concentration of 1 mM DTT in a 96-well plate. Competition assays were prepared by combining different ratios of the TMH1 locked C9 (F262C/V405C, [C9$_{mutant}$]) with a constant amount of wild-type C9 (final concentration 270 ng mL⁻¹) prior to addition of EAC1-8 or BSA which was used as a non-specific binding control. Data are reported as the raw turbidity curves with error reported as the standard error of the mean.

Data availability. Data supporting the findings of this manuscript are available from the corresponding authors upon reasonable request. The datasets generated

References

1. Morgan, B. P., Boyd, C. & Bubeck, D. Molecular cell biology of complement membrane attack. *Semin. Cell Dev. Biol.* **72**, 124–132 (2017).
2. Bayly-Jones, C., Bubeck, D. & Dunstone, M. A. The mystery behind membrane insertion: a review of the complement membrane attack complex. *Philos. Trans. R. Soc. B Biol. Sci.* **327**, https://doi.org/10.1098/rstb.2016.0221 (2017).
3. Serna, M., Giles, J. L., Morgan, B. P. & Bubeck, D. Structural basis of complement membrane attack complex formation. *Nat. Commun.* **7**, 10587 (2016).
4. Sharp, T. H., Koster, A. J. & Gros, P. Heterogeneous MAC Initiator and Pore Structures in a lipid bilayer by phase-plate cryo-electron tomography. *Cell Rep.* **15**, 1–8 https://doi.org/10.1016/j.celrep.2016.03.002 (2016).
5. Rosado, C. J. et al. The MACPF/CDC family of pore-forming toxins. *Cell Microbiol.* **10**, 1765–1774 (2008).
6. Shepard, L. A., Shatursky, O., Johnson, A. E. & Tweten, R. K. The mechanism of pore assembly for a cholesterol-dependent cytolysin: formation of a large prepore complex precedes the insertion of the transmembrane β-hairpins. *Biochemistry* **39**, 10284–10293 (2000).
7. Czajkowsky, D. M., Hotze, E. M., Shao, Z. & Tweten, R. K. Vertical collapse of a cytolysin prepore moves its transmembrane beta-hairpins to the membrane. *EMBO J.* **23**, 3206–3215 (2004).
8. Leung, C. et al. Stepwise visualization of membrane pore formation by suilysin, a bacterial cholesterol-dependent cytolysin. *eLife* **3**, e04247 (2014).
9. Shepard, L. A. et al. Identification of a membrane-spanning domain of the thiol-activated pore-forming toxin *Clostridium perfringens* perfringolysin O: an α-helical to β-sheet transition identified by fluorescence spectroscopy. *Biochemistry* **37**, 14563–14574 (1998).
10. Shatursky, O. et al. The mechanism of membrane insertion for a cholesterol-dependent cytolysin. *Cell* **99**, 293–299 (1999).
11. Lukoyanova, N. et al. Conformational changes during pore formation by the perforin-related protein pleurotolysin. *PLoS Biol.* **13**, e1002049 (2015).
12. Leung, C. et al. Real-time visualization of perforin nanopore assembly. *Nat. Nanotechnol.* **12**, 467–473 (2017).
13. Dudkina, N. V. et al. Structure of the poly-C9 component of the complement membrane attack complex. *Nat. Commun.* **7**, 10588 (2016).
14. Aleshin, A. E. et al. Structure of complement C6 suggests a mechanism for initiation and unidirectional, sequential assembly of membrane attack complex (MAC). *J. Biol. Chem.* **287**, 10210–10222 (2012).
15. Lovelace, L. L., Cooper, C. L., Sodetz, J. M. & Lebioda, L. Structure of human C8 protein provides mechanistic insight into membrane pore formation by complement. *J. Biol. Chem.* **286**, 17585–17592 (2011).
16. Hadders, M. A. et al. Assembly and regulation of the membrane attack complex based on structures of C5b6 and sC5b9. *Cell Rep.* **1**, 200–207 (2012).
17. van Pee, K. et al. CryoEM structures of membrane pore and prepore complex reveal cytolytic mechanism of Pneumolysin. *Elife* **6**, e23644 (2017).
18. Rosado, C. J. et al. A common fold mediates vertebrate defense and bacterial attack. *Science* **317**, 1548–1551 (2007).
19. Reboul, C. F., Mahmood, K., Whisstock, J. C. & Dunstone, M. A. Predicting giant transmembrane β-barrel architecture. *Bioinformatics* **28**, 1299–1302 (2012).
20. Musingarimi, P., Plumb, M. E. & Sodetz, J. M. Interaction between the C8α-γ and C8β subunits of human complement C8: role of the C8β N-terminal thrombospondin type 1 module and membrane attack complex/perforin domain. *Biochemistry* **41**, 11255–11260 (2002).
21. Podack, E. R., Stoffel, W., Esser, A. F. & Müller-Eberhard, H. J. Membrane attack complex of complement: distribution of subunits between the hydrocarbon phase of target membranes and water. *Proc. Natl Acad. Sci. USA* **78**, 4544–4548 (1981).
22. Law, R. H. P. et al. The structural basis for membrane binding and pore formation by lymphocyte perforin. *Nature* **468**, 447–451 (2010).
23. Biesecker, G. & Müller-Eberhard, H. J. The ninth component of human complement: purification and physicochemical characterization. *J. Immunol.* **124**, 1291–1296 (1980).
24. Kabsch, W. XDS. *Acta Crystallogr. Sect. D* **66**, 125–132 (2010).
25. Kabsch, W., IUCr. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. Sect. D* **66**, 133–144 (2010).
26. Evans, P., IUCr. Scaling and assessment of data quality. *Acta Crystallogr. Sect. D* **62**, 72–82 (2006).
27. Evans, P. R., IUCr. An introduction to data reduction: space-groupdetermination, scaling and intensity statistics. *Acta Crystallogr. Sect. D* **67**, 282–292 (2011).
28. Evans, P. R. & Murshudov, G. N., IUCr. How good are my data and what is the resolution? *Acta Crystallogr. Sect. D* **69**, 1204–1214 (2013).
29. Pape, T. & Schneider, T. R., IUCr. HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *J. Appl. Crystallogr.* **37**, 843–844 (2004).
30. Pannu, N. S. et al. Recent advances in the CRANK software suite for experimental phasing. *Acta Crystallogr. Sect. D* **67**, 331–337 (2011).
31. Sheldrick, G. M., IUCr. Experimental phasing with SHELXC / D / E: combining chain tracing with density modification. *Acta Crystallogr. Sect. D* **66**, 479–485 (2010).
32. Pannu, N. S. & Read, R. J., IUCr. The application of multivariate statistical techniques improves single-wavelength anomalous diffraction phasing. *Acta Crystallogr. Sect. D* **60**, 22–27 (2004).
33. Cowtan, K., IUCr. Recent developments in classical density modification. *Acta Crystallogr. Sect. D* **66**, 470–478 (2010).
34. Cowtan, K., IUCr. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. Sect. D* **62**, 1002–1011 (2006).
35. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K., IUCr. Features and development of Coot. *Acta Crystallogr. Sect. D* **66**, 486–501 (2010).
36. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D* **66**, 213–221 (2010).
37. Murshudov, G. N., Vagin, A. A. & Dodson, E. J., IUCr. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **53**, 240–255 (1997).
38. Bricogne, G. et al. BUSTER, version BUSTER (Global Phasing Ltd, Cambridge, 2017).
39. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D* **67**, 235–242 (2011).
40. Thorn, A. & Sheldrick, G. M., IUCr. ANODE: anomalous and heavy-atom density calculation. *J. Appl. Crystallogr.* **44**, 1285–1287 (2011).
41. DeLano, W. L. The PyMOL Molecular Graphics System v1.8 (Schrödinger LLC, 2014). https://doi.org/10.1038/hr.2014.17
42. Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
43. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
44. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
45. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
46. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
47. Scheres, S. H. W. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* **9**, 853–854 (2012).
48. Cardone, G., Heymann, J. B. & Steven, A. C. One number does not fit all: mapping local variations in resolution in cryo-EM reconstructions. *J. Struct. Biol.* **184**, 226–236 (2013).
49. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D* **60**, 2126–2132 (2004).
50. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
51. Tina, K. G., Bhadra, R. & Srinivasan, N. PIC: protein interactions calculator. *Nucleic Acids Res.* **35**, W473–W476 (2007).

Author contributions

M.A.D. and J.C.W. conceived the study, co-led the work and co-wrote the paper. B.A.S., R.H.P.L., T.T.C.D. collected data, determined the structures and co-wrote the paper. B.A.S.,

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-05717-0.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

112

The first transmembrane region of Complement component-9 acts as a brake on its self-assembly

Spicer *et al.*

SUPPLEMENTARY INFORMATION

Supplementary Table 1 Intermolecular contacts between adjacent molecules of polyC9.
Chain A and B are the elongating and binding molecules, respectively.

Intermolecular hydrophobic interactions (< 5 Å)					
Residue No	Residue	Chain	Residue No	Residue	Chain
60	ALA	A	454	PRO	B
60	ALA	A	456	TYR	B
146	PRO	A	286	VAL	B
147	LEU	A	182	ILE	B
379	PHE	A	214	PHE	B
379	PHE	A	216	ALA	B
392	ALA	A	203	PHE	B
393	VAL	A	200	ILE	B
393	VAL	A	202	ALA	B
395	ILE	A	200	ILE	B
404	VAL	A	191	PHE	B
417	PHE	A	280	ILE	B
417	PHE	A	434	PHE	B
417	PHE	A	435	VAL	B
417	PHE	A	438	ALA	B
Intermolecular Cation-Pi interactions (< 6 Å)					
107	PHE	A	122	ARG	B
133	ARG	A	171	TYR	B
522	LYS	A	196	TYR	B
420	LYS	A	434	PHE	B
Inter main-chain hydrogen bonds							

Residue No	Residue	Chain	Atom	Residue No	Residue	Chain	Atom
380	A	ASN	N	215	B	ASN	O
382	A	ASP	N	213	B	ASN	O
384	A	CYS	N	211	B	THR	O
386	A	LYS	N	209	B	GLU	O
388	A	GLY	N	207	B	ILE	O
390	A	GLY	N	205	B	SER	O
392	A	ALA	N	203	B	PHE	O
394	A	ASN	N	201	B	GLU	O
396	A	THR	N	199	B	GLN	O
398	A	GLU	N	197	B	GLU	O
400	A	LEU	N	195	B	HIS	O
402	A	ASP	N	193	B	THR	O
404	A	VAL	N	191	B	PHE	O
406	A	SER	N	189	B	LYS	O
408	A	ILE	N	187	B	GLY	O
187	B	GLY	N	408	A	ILE	O
189	B	LYS	N	406	A	SER	O
191	B	PHE	N	404	A	VAL	O

193	B	THR	N	402	A	ASP	O
195	B	HIS	N	400	A	LEU	O
197	B	GLU	N	398	A	GLU	O
199	B	GLN	N	396	A	THR	O
201	B	GLU	N	394	A	ASN	O
203	B	PHE	N	392	A	ALA	O
205	B	SER	N	390	A	GLY	O
207	B	ILE	N	388	A	GLY	O
209	B	GLU	N	386	A	LYS	O
211	B	THR	N	384	A	CYS	O
213	B	ASN	N	382	A	ASP	O
215	B	ASN	N	380	A	ASN	O
Main-chain and side-chain hydrogen bonds							
59	A	ASP	OD2	452	B	LEU	O
59	A	ASP	OD2	452	B	LEU	O
66	A	GLN	OE1	461	B	VAL	O
66	A	GLN	OE1	461	B	VAL	O
147	A	LEU	N	180	B	SER	OG
382	A	ASP	OD1	213	B	ASN	O
382	A	ASP	OD1	213	B	ASN	O
384	A	CYS	N	211	B	THR	OG1
399	A	ASN	OD1	195	B	HIS	O
399	A	ASN	OD1	195	B	HIS	O
403	A	ASP	OD1	191	B	PHE	O
403	A	ASP	OD1	191	B	PHE	O
425	A	ARG	NH2	429	B	ILE	O
425	A	ARG	NH2	429	B	ILE	O
164	B	ARG	NH1	443	A	ASP	O
164	B	ARG	NH1	443	A	ASP	O
185	B	THR	OG1	411	A	GLY	O
190	B	ASN	OD1	404	A	VAL	O
190	B	ASN	OD1	404	A	VAL	O
211	B	THR	OG1	384	A	CYS	O
289	B	ASN	ND2	316	A	ALA	O
289	B	ASN	ND2	316	A	ALA	O
487	B	ARG	NH2	53	A	ASN	O
487	B	ARG	NH2	53	A	ASN	O
Inter side-chain hydrogen bonds							
320	A	THR	OG1	289	B	ASN	ND2
387	A	ARG	NH1	208	B	GLN	NE2
387	A	ARG	NH1	208	B	GLN	NE2
394	A	ASN	ND2	201	B	GLU	OE1
394	A	ASN	ND2	201	B	GLU	OE1
189	B	LYS	NZ	421	A	GLU	OE1
189	B	LYS	NZ	421	A	GLU	OE2
201	B	GLU	OE1	394	A	ASN	ND2
201	B	GLU	OE1	394	A	ASN	ND2
288	B	ARG	NH2	301	A	ASP	OD2

288	B	ARG	NH2	301	A	ASP	OD2
289	B	ASN	ND2	320	A	THR	OG1
289	B	ASN	ND2	320	A	THR	OG1
290	B	ARG	NE	313	A	GLU	OE1
487	B	ARG	NH1	19	A	HIS	NE2
487	B	ARG	NH1	19	A	HIS	NE2
487	B	ARG	NH2	19	A	HIS	NE2
487	B	ARG	NH2	19	A	HIS	NE2

Supplementary Table 2 Substructure movement of poly C9 in comparison with monomeric C9*.

Fragment	Residue range[3]	Structure components	rmsd Å	Translation Å	Number of residues	Number conserved
RB1[1]	286-329 452-472	MACPF	0.917	-	65	52
RB2[1]	80-110 119-134	LDL	1.082	1.770	73	60
	151-166 172-180	MACPF				
TSP[1]	18-67	TSP1	0.849	2.551	50	37
Helix-EGF[1]	474-520	EGF	0.864	3.013	47	30
TSP-EGF[1]	18-24 47-58	TSP1	0.864	2.533	66	43
	474-520	EGF				
HTH-2[2]	408-413 431-447	HTH helix 2	0.524	-	23	16
HTH-1[2]	415-424	HTH helix 1	0.334	5.205	10	6

*Analysis was performed on human poly C9 and murine monomer C9 and their respective movements from soluble to pore transition.
[1]The alignment of RB1 of monomer C9 (see residue range in Table) and poly C9 was used as a probe to determine the translation to the RB2, TSP, Helix-EGF and TSP-EGF substructures.
[2]The alignment of the HTH-2 regions was used as a probe to determine the translation of the HTH-1 helix.
[3]Amino acid numbering of the rigid body fragments is based on the sequence of human C9.

Supplementary Table 3 Oligonucleotide primers for cloning and site-directed mutagenesis

PRIMER	SEQUENCE
oBS86	5'-GAGGCGGCCCAGCCGGCCCAGTACACGACCAGTTATGACCC-3'
oBS101	5'-ACTTACCAACTATGTTTGTCATATTCTTCAAAG-3'
oBS102	5'-GAAGAATATGACAAACATAGTTGGTAAGTTTC-3'
oBS105	5'-CATAGATGATGTTTGTTCACTCATAAGAGGTGG-3'
oBS106	5'-CTCTTATGAGTGAACAAACATCATCTATGAG-3'
BGH	5'-TAGAAGGCACAGTCGAGG-3'

Supplementary Table 4 Data collection, for MIRAS phasing and refinement statistics*

	Native	Crystal 1 (Tantalum bromide)	Crystal 2 (Uranyl format)
Data collection			
Space group	P 2_1 2 2_1	P 2_1 2 2_1	P 2_1 2 2_1
Cell dimensions			
a, b, c (Å)	52.91 148.87 165.78	52.91 148.29 165.43	52.59 149.21 165.83
α, β, γ (°)	90.00 90.00 90.00	90.00 90.00 90.00	90.00 90.00 90.00
Wavelength (eV)	10300	10300	10300
Resolution (Å)	2.2	2.3	2.5
R_{merge}	0.075 (1.359)	0.09 (2.182)	0.1 (1.341)
R_{pim}	0.023 (0.514)	0.02 (0.586)	0.023 (0.323)
$I / \sigma I$	16.8 (1.4)	1 (0.59)	20 (2)
Completeness (%)	98.7 (97.3)	99.1 (95.5)	99.8 (98.5)
Redundancy	12.1 (8.6)	20.2 (15.1)	21.8 (18.5)
Refinement			
Resolution (Å)		2.20	
No. reflections (work/free)		66058/3369	
R_{work} / R_{free}		20.91/25.02	
No. atoms			
Protein (Chain A/B)		3632/3528	
Ligand/ion (Chain X/Y/Z)		29/29/1	
Water		335	
B-factors			
Protein (Chain A/B)		72.43/62.66	
Ligand/ion (Chain X/Y/Z)		91.84/75.47/99.17	
Water		63.15	
R.m.s deviations			
Bond lengths (Å)		0.008	
Bond angles (°)		0.98	

*Values in parentheses are for highest-resolution shell.

Supplementary Table 5 Data collection for Zinc Verification *

	Low Energy	High Energy
Data collection		
Space group	P 2_1 2 2_1	P 2_1 2 2_1
Cell dimensions		
a, b, c (Å)	53.46 148.86 165.03	53.56 149.06 165.37
α, β, γ (°)	90.00 90.00 90.00	90.00 90.00 90.00
Wavelength (eV)	9643.9	9674.0
Resolution (Å)	2.48	2.6
R_{merge}	0.116 (1.849)	0.171 (3.125)
R_{pim}	0.02 (0.453)	0.042 (0.764)
$I / \sigma I$	17.9 (1.7)	13.7 (0.9)
Completeness (%)	99.8 (98.4)	100 (100)
Redundancy	18.0 (18.1)	18 (18.5)

*Values in parentheses are for highest-resolution shell.

Supplementary Table 6 Cryo-EM data collection, refinement and validation statistics

	PolyC9 (EMDB-7773) (PDB id 6DLW)
Data collection and processing	
Magnification	75,000
Voltage (kV)	300
Electron exposure (e–/Å^2)	46.4
Defocus range (μm)	- 0.5 to -3.0
Pixel size (Å)	1.06
Symmetry imposed	C22
Initial particle images (no.)	220,000
Final particle images (no.)	52,000
Map resolution (Å)	3.9
FSC threshold	0.143
Map resolution range (Å)	3.2 – 4.5
Refinement	
Initial model used (PDB code)	6CXO
Model resolution (Å)	3.9
FSC threshold	0.143
Model resolution range (Å)	3.2 – 4.5
Map sharpening B factor (Å^2)	-180
Model composition	
Non-hydrogen atoms	3,630
Protein residues	460
Ligands	25
B factors (Å^2)	
Protein	-180
Ligand	-180
Validation	
MolProbity score	1.80 (85th percentile)
Clashscore	4.87 (94th percentile)
Poor rotamers (%)	0.25%
Ramachandran plot	
Favored (%)	90.17
Allowed (%)	9.83
Disallowed (%)	0.0
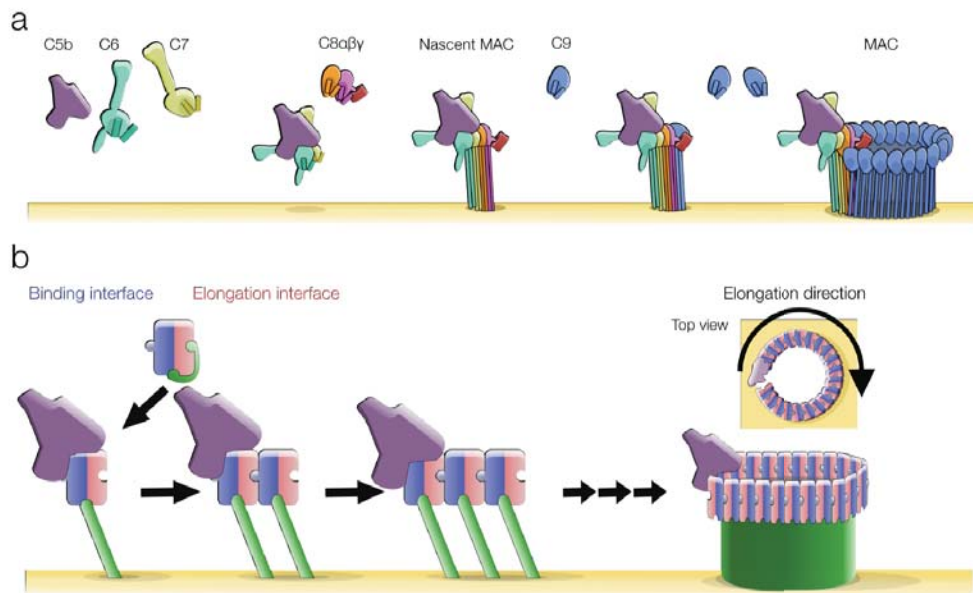
Supplementary Figure 1 a) Schematic diagram of MAC assembly. Soluble C5b forms a metastable complex with C6 and C7 (to C5b7) that superficially anchors to membranes. The C5b7 recruits C8 (αβγ) to form the nascent MAC C5b8. Seventeen to nineteen molecules of the C9 component then oligomerize with the nascent MAC forming a large membrane spanning β-barrel. The C6, C7, C8α and C8β and C9 components each contribute two β-hairpins to the barrel shown in the soluble components as cylinders and the membrane bound form as long sticks. **b)** Schematic illustration of unidirectional assembly of C9. Each component of the MAC contains a binding surface and an elongation surface. Once an individual component is associated with the nascent MAC, its elongation surface is activated (presumably *via* a conformational change) such that it can now interact with the binding surface of the next soluble component to join the complex.

122
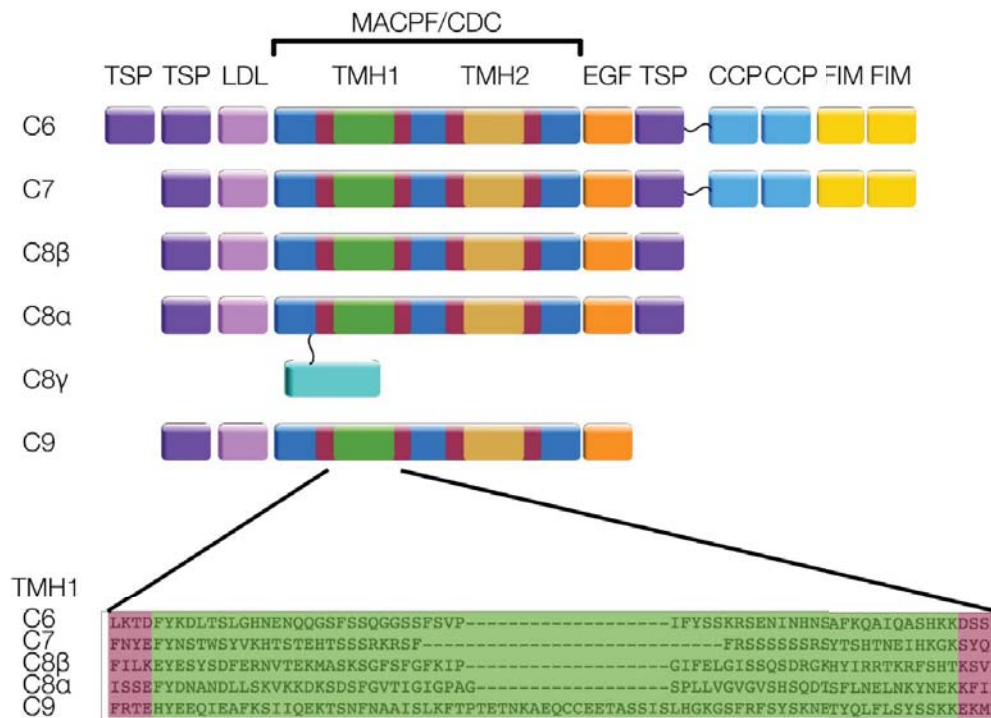
Supplementary Figure 2 Domain layout of the MAC components C6, C7, C8 and C9. These components contain a MACPF/CDC domain (blue and red) each with two transmembrane beta hairpins (TMH1, green, and TMH2 gold). The components also contain several ancillary domains, including thrombospondin-like domains (purple); low density lipoprotein receptor class A (LDL, pink); and EGF-like domain (orange); complement control protein (CCP, light blue) and factor I-membrane attack complex (FIM, bright yellow). The C8α component contains a disulphide link to C8γ (cyan). The sequence alignment of the TMH1 regions of all human protein sequences shows that this region is highly variable.
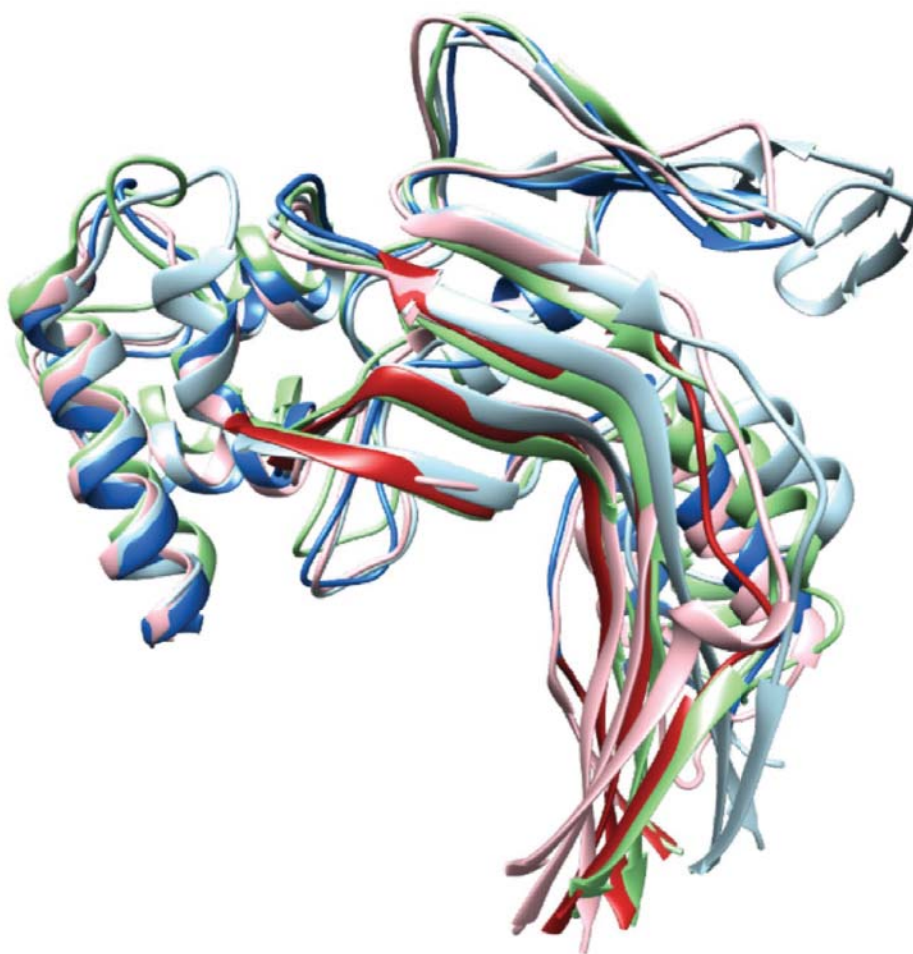
Supplementary Figure 3 Structural alignment of complement proteins. C6, pink (PDB ID 3T5O); C8α, light blue and C8β, green (PDB ID 3OJY); and C9, red and dark blue.
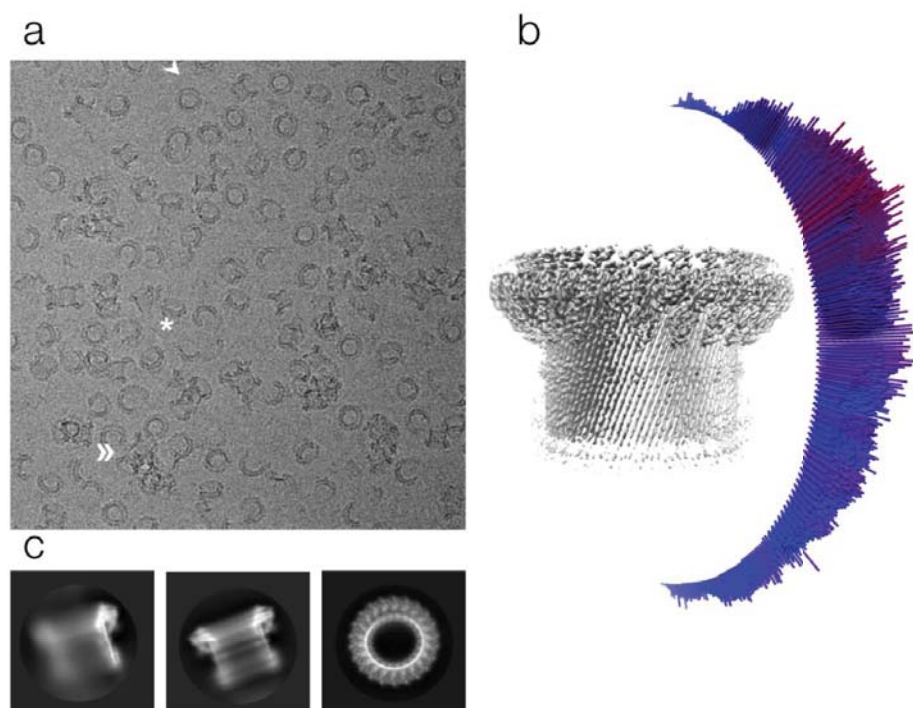
Supplementary Figure 4   Electron microscopy data of polyC9 a) Representative micrograph of vitrified recombinant human polyC9 imaged at 300 kV. Examples of a well-formed pore (arrow), an arc (asterisk) and aggregation (double-headed arrows) are shown. **b)** Angular distribution of views used in the final reconstruction. **c)** Representative 2D class averages of polyC9.
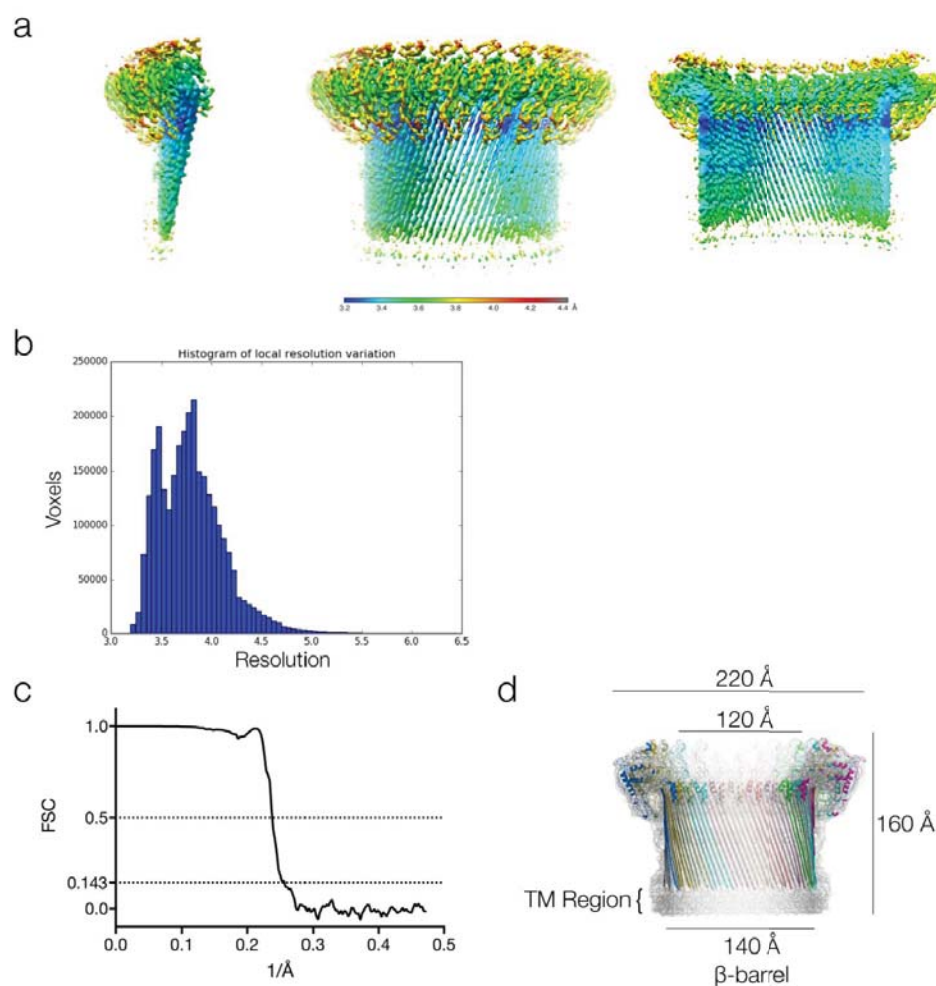
Supplemental Figure 5 Resolution data of polyC9 reconstruction a) Local variation of resolution in polyC9 reconstruction, colored according to the scale bar below, highlights local resolution features within the upper region of the β-barrel. **b)** Histogram of local resolution values throughout the reconstruction. **c)** Fourier shell correlation plot of the final polyC9 half maps. **d)** Unsharpened map with the final atomic model excluding the TM region which were not modelled due to lower resolution. The dimensions are also shown (220 Å denotes the outer-most dimension and 120 Å denotes the inner most dimension of the pore).

126
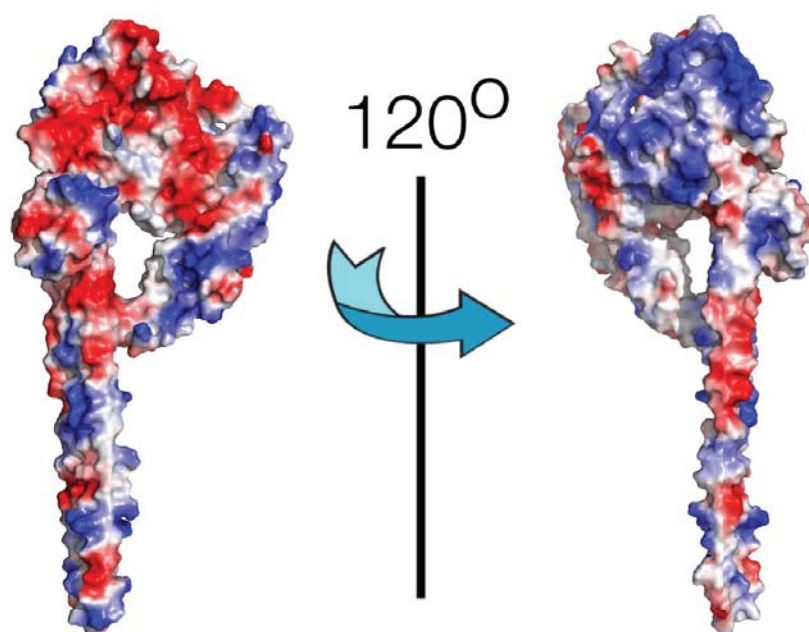
Supplementary Figure 6 Charge complementarity of the oligomer interfaces. The elongation face (left) and the binding face (right) of two neighbouring subunits of C9, rotated 120° along the y-axis. The electrostatic potential surfaces are coloured; blue, basic; red, acidic.
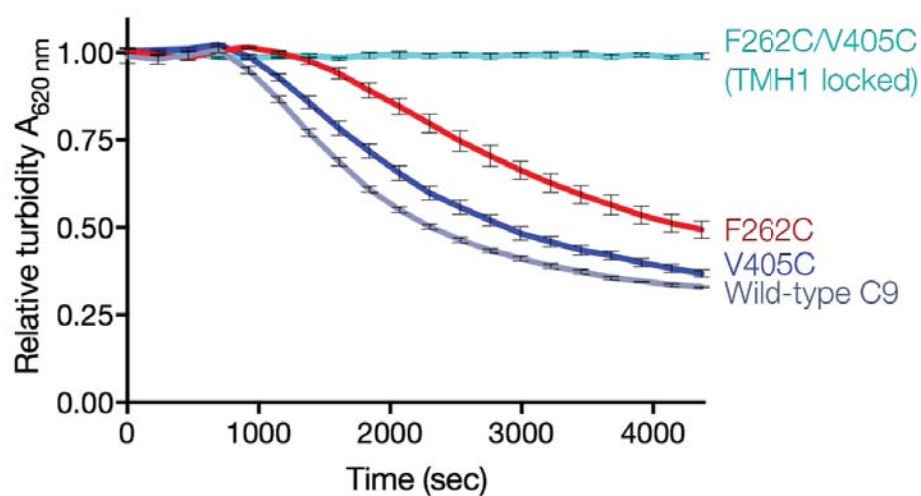
Supplementary Figure 7 Haemolytic activity of double and single cysteine mutants. The TMH1 disulphide locked C9 mutant (F262C/V405C, $C9_{mutant}$) shows a synergistic loss in haemolytic activity, as determined by the turbidity assay, against EAC1-8 compared to the single mutants (F262C and V405C); this loss in activity can be rescued by addition of a reducing agent (1 mM DDT, **Fig. 2b**). Together it suggests that the loss in activity is due to a disulphide lock. Results are reported as the average turbidity curve from three independently prepared samples (error is shown as the standard error of the mean, SEM).

6.1   General discussion

Members of the MACPF/CDC superfamily perform a vast diversity of functions, as bacterial toxins, venoms and immune effectors (139). To perform these roles MACPF/CDCs assemble into pores via a multi-step process. This includes initial membrane binding, lateral diffusion and self-association of 12-50 subunits to form a pre-pore. Finally a concerted conformational change culminates in pore insertion (91,98,99,140) (**Figure 6-1**). A feature common to each of these systems is the presence of dedicated membrane binding ancillary domains – for example, an immunoglobulin-like domain (domain 4) in CDCs, the calcium binding C2 domain in perforin and an actinoporin-like lipid binding domain in pleurotolysin.
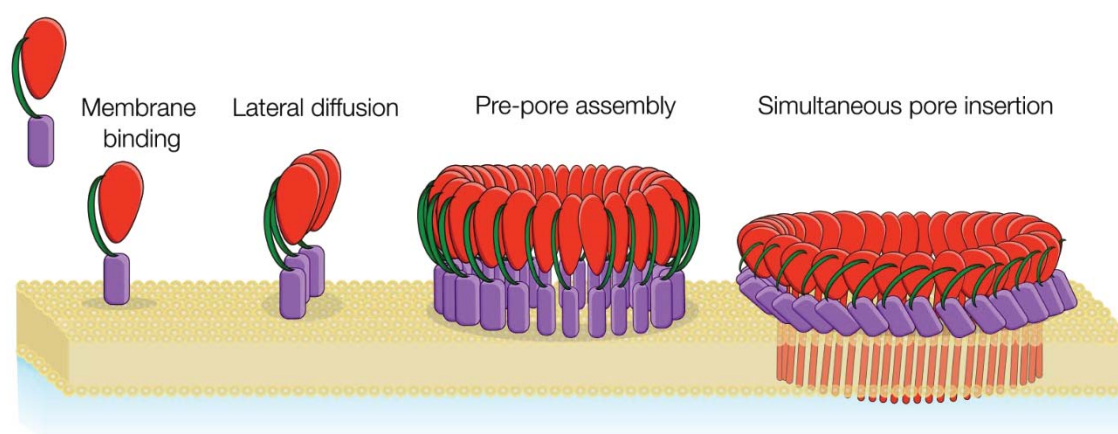
Figure 6-1 Pore formation of CDCs.

Pore assembly model for CDCs suggest that the numerous protein subunits bind to the membrane and diffuse along the membrane to form a pre-pore that undergoes simultaneous pore insertion.

The membrane attack complex (MAC) is a pore forming effector of the vertebrate innate immune system that has evolved to assemble on a variety of target membranes. The MAC forms ~12 nm pores in the membranes of pathogens such as gram-negative bacteria and parasites (31). A recent report also suggests that MAC can assemble on gram-positive bacteria (141).

Unlike other MACPF/CDC proteins characterised to date, assembly of the MAC is initiated by a non-MACPF protein, complement C5b. The latter protein sequentially recruits C6, C7, C8 and finally 17-19 C9 monomers (**Figure 6-1**). C6-C9 all lack obvious membrane binding domains; instead the MAC appears to assemble above the membrane, rather than oligomerising on the membrane surface (**Figure 6-2**). Currently, the molecular details of this unique membrane-free assembly pathway are not well understood. Additionally, while several structures of the MAC precursors have been determined prior to this study, less was known about the MAC assembly and how monomeric C9 transitioned into the final MAC pore (77–80,83) (**Table 1-2** and **Table 1-3**).
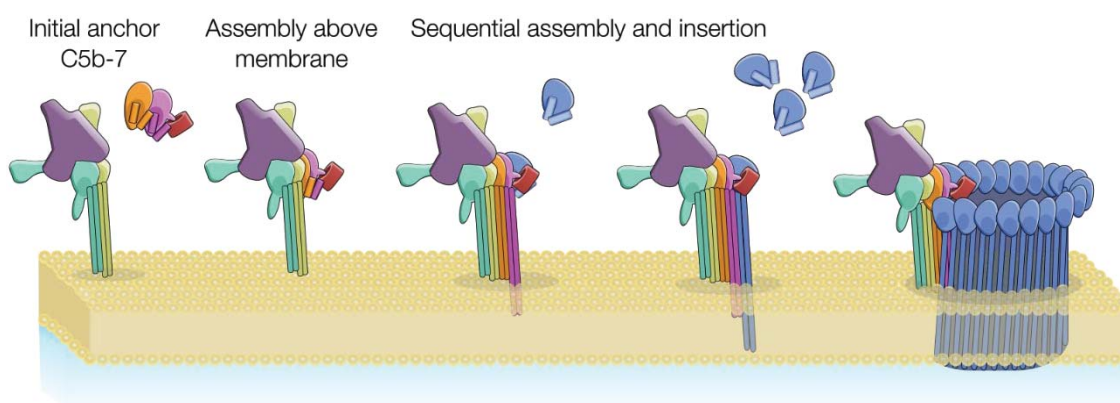
Figure 6-2 Sequential assembly of the MAC.

Assembly of the MAC which forms a pore by sequential recruitment and insertion of the MAC subunits. In this MAC model, the initial interaction of the subunits occurs above the membrane (no membrane binding regions).

To address these problems, in this thesis the structures of monomeric and polymeric C9 are presented. In **Chapter 3**, the sub-nanometre resolution cryo-EM structure of plasma-derived polyC9 revealed the importance of the C9 TSP domain in mediating C9 self-association. Next, in order to obtain higher resolution structures, in **Chapter 4** a recombinant C9 expression system was developed. The advantage of using such recombinantly expressed material, was the ability to engineer the protein to improve both

its crystallisation properties and behaviour under EM. In this way, high resolution structures of both C9 and polyC9 were obtained (**Chapter 5**). These recombinant systems were also important for generating mutations that were of utility in the investigation of the mechanism of MAC assembly (**Chapter 5**).

Together, the research presented in this thesis has contributed to an improved understanding of the MAC assembly. This information may, in the future, be useful for engineering novel inhibitors of MAC assembly to prevent a range of immune-driven diseases.

6.2 *Investigating how the MAC assembles in the absence of a membrane*

The MAC assembles on a remarkable variety of bacterial and eukaryotic membrane surfaces (142). While some preference for negatively charged membranes has been identified (66), it is known that the components of the MAC lack affinity for the membrane in their soluble precursor form (59,60). Further, it is also notable that the MAC has been reported to decorate the surface of gram-positive bacteria (141). The latter organisms are thought to be resistant to MAC pore formation owing to their thick peptidoglycan cell wall. Currently, it is unclear how membrane-independent MAC assembly takes place (59,60).

The work presented in **Chapter 3** reveals how oligomerisation and pore formation occur in the MAC without a dedicated membrane binding domain. The 8.0 Å resolution structure of polyC9, presented in **Chapter 3**, revealed the unexpected finding that a region ancillary to the C9 MACPF/CDC domain, the N-terminal TSP domain (106), performed a key role in mediating C9 self-association. The TSP domain of C9 is positioned at the top of the pore and contributes to the outer rim structure of the cylinder. Positioned in this way, each TSP domain is effectively wedged between two subunits of

C9 (one *in cis-* and one *in trans-*), and makes interactions with both the MACPF/CDC and EGF domains. Based on these observations, it was suggested that the TSP interactions drive the solution-based oligomerisation of C9 in the absence of a membranous template. This idea is supported by the observation that C8 complex formation assembly is markedly reduced in the absence of the TSP domain (143).

In the broader context, the model of assembly where MAC components are recruited from the soluble phase and bind to a nascent complex is advantageous for assembling on a wide variety of target membranes. In an *in vivo* context, the activation pathways can be initiated by a range of distinctly different membrane surfaces that culminate in the deposition of C5 convertases on target membranes. For this reason, the MAC has evolved such that it has overcome the need to interact with a defined lipid chemistry or specific protein receptor, thereby allowing the assembly to proceed on a variety of surfaces wherever C5 convertase is present.

While the events of MAC oligomerisation are now better understood, it remains unclear how the C5b-7 complex becomes membrane bound. One hypothesis recently presented is that C5b remains tethered to a membrane bound C5 convertase and is displaced from this interaction upon binding C6 and C7, presumably after insertion into the membrane (104). Indeed, interactions of C5 with the C-terminal CCPs and FIM domains of C6 and C7 suggest a structural means of displacing the C5 convertase. However this hypothesis has yet to be experimentally tested (78,79,86,87).

6.3 High resolution structures of monomeric and polymeric C9 reveal the mechanism of controlled MAC assembly

The 2.2 Å resolution structure of monomeric C9 reveals the overall arrangement of the core domains is similar to C6 and C8 (77,80,83). Notably, however, the position of the TMH1 loop is different in comparison to the other MAC components (**Figure 6-3**),

such that it appears to block the elongation interface (**Figure 6-3**). To test this idea, a TMH1 disulphide lock mutant C9$_{F262C/V405C}$ was produced. C9$_{F262C/V405C}$ can be incorporated into a nascent MAC, as evidenced by competition assays with wild-type C9. However, this variant remains inactive unless reduced with DTT. Collectively these data support the idea that TMH1 functions as a brake on C9 self-assembly.

Next, the cryo-EM structure of polyC9 (from C9$_{[HEK]}$) was determined at 3.9 Å resolution. This allowed a near complete atomistic model of the 22-subunit structure to be built. Structural comparison of the monomeric C9 with polyC9 supports the hypothesis that rearrangement of the TMH1 loop is required for polymerisation. In addition, these data further suggested that the HTH region, which sits on top of TMH2, must move to permit an additional C9 molecule to join the assembly (**Figure 6-3**).
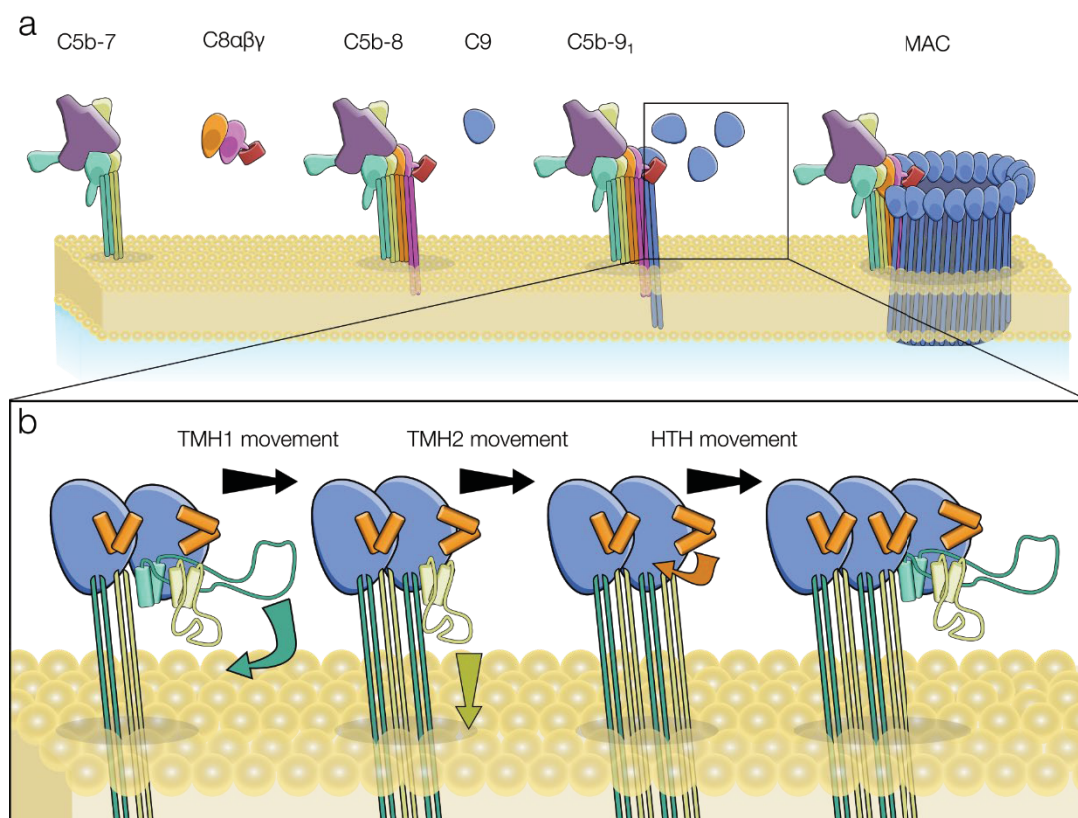
Figure 6-3 Cartoon schematic of MAC assembly and C9 oligomerisation.

a) Representation of MAC assembly by sequential recruitment of complement components. b) Improved model of C9 oligomerisation and auto-inhibition. In this model, the TMH1 of C9 maintains an inhibited state by sterically blocking the addition of further C9 subunits. The movement of the C9 TMH1 (green), TMH2 (yellow) and HTH (orange) are required for the next subunit to be added to the MAC.

6.4 Implications of the work for understanding the MAC inhibitor CD59

The structures may provide insight into how the MAC assembly is inhibited by CD59. This inhibitor acts on C8 and C9 by blocking both membrane insertion (via binding TMH2) and recruitment of additional C9 monomers (section 1.2.1) (43,144). Considering the structural insights presented here, it seems plausible that CD59 can achieve both inhibitory functions simply by restricting the movement of TMH2 (**Figure 6-4**). Such an interaction would prevent insertion into the membrane and may also inhibit the movement of the HTH region that forms part of the elongation surface.
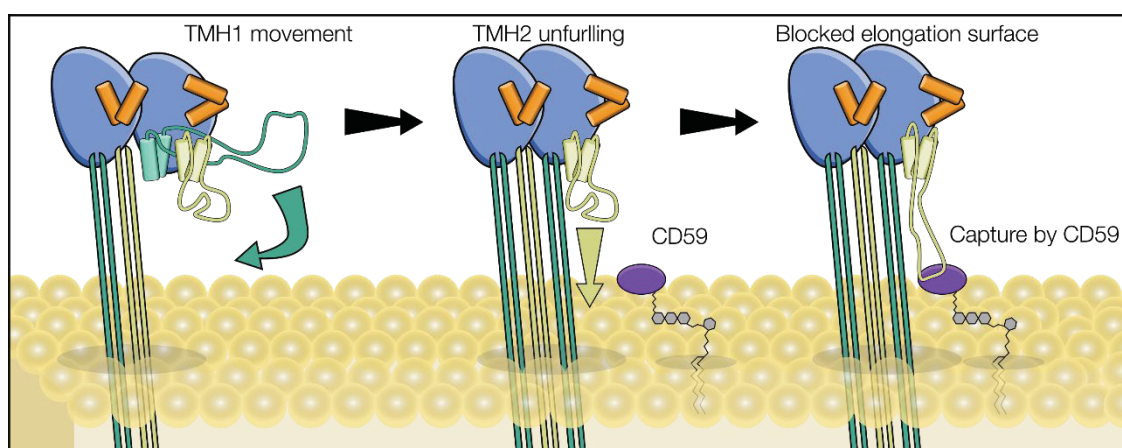
Figure 6-4 Possible inhibitory mechanism of MAC assembly by CD59.

This model suggests that CD59 can bind to a partially unfurled TMH2 in a nascent MAC. This binding may prevent important conformational change such as the HTH rearrangement. Overall, this may block subsequent C9 monomers from binding to a growing ensemble.

6.5 Conclusions

This thesis has presented an improved structural understanding of how the soluble C9 component in the plasma transitions into the MAC pore.

Future questions include understanding how damage to the outer membrane of gram-negative bacteria results in cell death. One idea is that MAC assembly on the outer membrane precedes pore formation on the inner membrane (145). This may provide a mechanistic role for lysozyme to remove the obstructive peptidoglycan and provide

access for MAC components to the inner membrane (31). Here, a single MAC could uncouple the proton motive force that is necessary for cellular processes similar to bacterial colicins (146,147). This presents opportunities for further investigation, to determine if MAC assembly can disrupt the bacteria inner membrane.

In contrast, the function of MAC formation in the context of gram-positive bacteria is not known (141). These organisms contain a thick cell wall thought to be protective from the MAC. However, a range of organisms have been shown to be coated by these components (141). Notably, this does not affect the cell viability *in vitro* (141). As such, it is currently unclear whether the MAC plays a non-lytic role on these cells. This may include inducing cellular responses such as phagocytosis. Support for this idea includes the observation that *Streptococcus pyogenes* produces an inhibitor that prevents C5b-7 from binding membranes (148).

Finally, the MAC has been implicated in several inflammatory conditions. Examples include rheumatoid arthritis and ischemia-reperfusion injury (149–151). The mechanistic insights gained from the studies presented in this thesis may pave the way for the development of novel MAC inhibitors that may be of therapeutic utility (41).

1.   Kaufmann SHE. Immunology's foundation: the 100-year anniversary of the Nobel Prize to Paul Ehrlich and Elie Metchnikoff. Nat Immunol. Nature Publishing Group; 2008 Jul;9(7):705–12.

2.   Nesargikar P, Spiller B, Chavez R. The complement system: History, pathways, cascade and inhibitors. Eur J Microbiol Immunol. Akadémiai Kiadó ; 2012 Jun 13;2(2):103–11.

3.   Morgan BP. Complement: Clinical aspects and relevance to disease. Acad Press UK. Elsevier; 1991 Jul 1;12(7):250–1.

4.   Kindt T, Goldsby R, Osborne B, Kuby J. Kuby immunology. J Exp Med. 2007;206(13):2925–35.

5.   Dunkelberger JR, Song W-C. Complement and its role in innate and adaptive immune responses. Cell Res. Nature Publishing Group; 2010 Jan 15;20(1):34–50.

6.   Howard M, Farrar CA, Sacks SH. Structural and functional diversity of collectins and ficolins and their relationship to disease. Semin Immunopathol. Springer Berlin Heidelberg; 2018 Jan 11;40(1):75–85.

7.   Bexborn F, Andersson PO, Chen H, Nilsson B, Ekdahl KN. The tick-over theory revisited: Formation and regulation of the soluble alternative complement C3 convertase (C3(H2O)Bb). Mol Immunol. 2008;45:2370–9.

8.   Hughes-Jones NC, Gardner B. Reaction between the isolated globular sub-units of the complement component Clq and IgG-complexes. Mol Immunol. Pergamon; 1979 Sep 1;16(9):697–701.

9.   Ugurlar D, Howes SC, de Kreuk B-J, Koning RI, de Jong RN, Beurskens FJ, Schuurman J, Koster AJ, Sharp TH, Parren PWHI, Gros P. Structures of C1-IgG1 provide insights into how danger pattern recognition activates complement. Science. American Association for the Advancement of Science; 2018 Feb 16;359(6377):794–7.

10.   Medzhitov R, Janeway CA. Decoding the patterns of self and nonself by the innate immune system. Science. American Association for the Advancement of Science; 2002 Apr 12;296(5566):298–300.

11.   Fujita T, Matsushita M, Endo Y. The lectin-complement pathway - Its role in innate immunity and evolution. Immunol Rev. 2004;198:185–202.

12.   Sørensen R, Thiel S, Jensenius JC. Mannan-binding-lectin-associated serine proteases, characteristics and disease associations. Springer Seminars in Immunopathology. 2005. p. 299–319.

13.   Kjaer TR, Le LTM, Pedersen JS, Sander B, Golas MM, Jensenius JC, Andersen GR, Thiel S. Structural Insights into the Initiating Complex of the Lectin Pathway of Complement Activation. Structure. Cell Press; 2015 Feb 3;23(2):342–51.

14. Sato T, Endo Y, Matsushita M, Fujita T. Molecular characterization of a novel serine protease involved in activation of the complement system by mannose-binding protein. Int Immunol. Oxford University Press; 1994 Apr 1;6(4):665–9.

15. Takahashi M, Iwaki D, Kanno K, Ishida Y, Xiong J, Matsushita M, Endo Y, Miura S, Ishii N, Sugamura K, Fujita T. Mannose-binding lectin (MBL)-associated serine protease (MASP)-1 contributes to activation of the lectin complement pathway. J Immunol. American Association of Immunologists; 2008 May 1;180(9):6132–8.

16. Dobó J, Szakács D, Oroszlán G, Kortvely E, Kiss B, Boros E, Szász R, Závodszky P, Gál P, Pál G. MASP-3 is the exclusive pro-factor D activator in resting blood: the lectin and the alternative complement pathways are fundamentally linked. Sci Rep. Nature Publishing Group; 2016 Oct 18;6(1):31877.

17. Dahl MR, Thiel S, Matsushita M, Fujita T, Willis AC, Christensen T, Vorup-Jensen T, Jensenius JC. MASP-3 and Its Association with Distinct Complexes of the Mannan-Binding Lectin Complement Activation Pathway. Immunity. Cell Press; 2001 Jul 1;15(1):127–35.

18. Götze O, Müller-Eberhard HJ. The C3-activator system: An alternate pathway of complement activation. J Exp Med. 1971;134(3).

19. Zhu Y, Thangamani S, Ho B, Ding JL. The ancient origin of the complement system. EMBO J. Wiley-Blackwell; 2005 Jan 26;24(2):382–94.

20. Forneris F, Wu J, Gros P. The modular serine proteases of the complement cascade. Curr Opin Struct Biol. 2012 Jun;22(3):333–41.

21. Gros P, Milder FJ, Janssen BJC. Complement driven by conformational changes. Nat Rev Immunol. Nature Publishing Group; 2008 Jan;8(1):48–58.

22. Pangburn MK, Schreiber RD, Müller-Eberhard HJ. C3b deposition during activation of the alternative complement pathway and the effect of deposition on the activating surface. J Immunol. 1983;131(4):1930–5.

23. Gerard C, Gerard NP. C5A anaphylatoxin and its seven transmembrane-segment receptor. Annu Rev Immunol. 1994;12:775–808.

24. MK P, N R. Structure and function of complement C5 convertase enzymes. Biochem Soc Trans. Portland Press Ltd.; 2002 Nov 1;30(6):1006–10.

25. Podack ER, Tschopp J. Membrane attack by complement. Mol Immunol. 1984;21(7):589–603.

26. Muller-Eberhard HJ. The killer molecule of complement. Journal of Investigative Dermatology. 1985.

27. Borsos T, Dourmashkin RR, Humphrey JH. Lesions in Erythrocyte Membranes Caused by Immune Hæmolysis. Nature. 1964 Apr 18;202(4929):251–2.

28.   Lachmann PJ, Thompson RA. 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. 1970;131:643–57.

29.   Thompson RA, Lachmann PJ. Reactive lysis: the complement-mediated lysis of unsensitized cells. I. The characterization of the indicator factor and its identification as C7. J Exp Med. 1970;131:629–41.

30.   Schreiber RD, Morrison DC, Podack ER, Müller-Eberhard HJ. Bactericidal activity of the alternative complement pathway generated from 11 isolated plasma proteins. J Exp Med. 1979;149(4).

31.   Martinez RJ, Carroll SF. Sequential metabolic expressions of the lethal process in human serum-treated Escherichia coli: Role of lysozyme. Infect Immun. 1980;28:735–45.

32.   Miajlovic H, Smith SG. Bacterial self-defence: how Escherichia coli evades serum killing. FEMS Microbiol Lett. 2014 May;354(1):1–9.

33.   Ross SC, Densen P. Complement deficiency states and infection: epidemiology, pathogenesis and consequences of neisserial and other infections in an immune deficiency. Medicine (Baltimore). 1984 Sep;63(5):243–73.

34.   Nagata M, Hara T, Aoki T, Mizuno Y, Akeda H, Inaba S, Tsumoto K, Ueda K. Inherited deficiency of ninth component of complement: An increased risk of meningococcal meningitis. J Pediatr. 1989;114(2):260–3.

35.   Inai S, Akagaki Y, Moriyama T, Fukumori Y, Yoshimura K, Ohnoki S, Yamaguchi H. Inherited deficiencies of the late-acting complement components other than C9 found among healthy blood donors. Int Arch Allergy Immunol. 1989;90(3):274–9.

36.   Harriman GR, Podack ER, Braude AI, Corbeil LC, Esser AF, Curd JG. Activation of complement by serum-resistant Neisseria gonorrhoeae. Assembly of the membrane attack complex without subsequent cell death. J Exp Med. 1982 Oct 1;156(4):1235–49.

37.   Hammerschmidt C, Klevenhaus Y, Koenigs A, Hallström T, Fingerle V, Skerka C, Pos KM, Zipfel PF, Wallich R, Kraiczy P. BGA66 and BGA71 facilitate complement resistance of Borrelia bavariensis by inhibiting assembly of the membrane attack complex. Mol Microbiol. 2016 Jan;99(2):407–24.

38.   Rosenfeld SI, Jenkins Jr DE, Leddy JP. Enhanced reactive lysis of paroxysmal nocturnal hemoglobinuria erythrocytes. Studies on C9 binding and incorporation into high molecular weight complexes. J Exp Med. 1986;164(4):981–97.

39.   Kim SH, Carney DF, Hammer CH, Shin ML. Nucleated cell killing by complement: effects of C5b-9 channel size and extracellular Ca2+ on the lytic process. J Immunol. 1987;138(5):1530–6.

40.   Niculescu F, Rus H, van Biesen T, Shin ML. Activation of Ras and mitogen-activated protein kinase pathway by terminal complement complexes is G

protein dependent. J Immunol. 1997;158(9):4405–12.

41. Morgan BP. The membrane attack complex as an inflammatory trigger. Immunobiology. 2015 Apr 30;

42. Hillmen P, Richards SJ. Implications of recent insights into the pathophysiology of paroxysmal nocturanal haemoglobinuria. Br J Haematol. 2000 Mar;108(3):470–9.

43. Huang Y, Qiao F, Abagyan R, Hazard S, Tomlinson S. Defining the CD59-C9 binding interaction. J Biol Chem. 2006 Sep 15;281(37):27398–404.

44. Meri S, Morgan BP, Davies A, Daniels RH, Olavesen MG, Waldmann H, Lachmann PJ. Human protectin (CD59), an 18,000-20,000 MW complement lysis restricting factor, inhibits C5b-8 catalysed insertion of C9 into lipid bilayers. Immunology. Wiley-Blackwell; 1990 Sep;71(1):1–9.

45. Chatterjee S, Mayor S. The GPI-anchor and protein sorting. Cell Mol Life Sci. 2001;58(14):1969–87.

46. Takeda J, Miyata T, Kawagoe K, Iida Y, Endo Y, Fujita T, Takahashi M, Kitani T, Kinoshita T. Deficiency of the GPI anchor caused by a somatic mutation of the PIG-A gene in paroxysmal nocturnal hemoglobinuria. Cell. Cell Press; 1993 May 21;73(4):703–11.

47. Bessler M, Mason PJ, Hillmen P, Miyata T, Yamada N, Takeda J, Luzzatto L, Kinoshita T. Paroxysmal nocturnal haemoglobinuria (PNH) is caused by somatic mutations in the PIG-A gene. EMBO J. Wiley-Blackwell; 1994 Jan 1;13(1):110–7.

48. Zachée P, Henckens M, Van Damme B, Boogaerts MA, Rigauts H, Verberckmoes RK. Chronic renal failure due to renal hemosiderosis in a patient with paroxysmal nocturnal hemoglobinuria. Clin Nephrol. 1993 Jan;39(1):28–31.

49. Clark D, Butler S, Braren V, Hartmann R, Jenkins DJ. The kidneys in paroxysmal nocturnal hemoglobinuria. Blood. 1981;57(1).

50. Socié G, Mary J-Y, de Gramont A, Rio B, Leporrier M, Rose C, Heudier P, Rochant H, Cahn J-Y, Gluckman E. Paroxysmal nocturnal haemoglobinuria: long-term follow-up and prognostic factors. Lancet. 1996;348(9027):573–7.

51. Antin J, Ginsburg D, Smith B, Nathan D, Orkin S, Rappeport J. Bone marrow transplantation for paroxysmal nocturnal hemoglobinuria: eradication of the PNH clone and documentation of complete lymphohematopoietic engraftment. Blood. 1985;66(6).

52. Schrezenmeier H, Hóchsmann B. Eculizumab opens a new era of treatment for paroxysmal nocturnal hemoglobinuria. Expert Review of Hematology. 2009. p. 7–16.

53. Podack ER. Molecular composition of the tubular structure of the membrane

attack complex of complement. J Biol Chem. 1984;259(13):8641–7.

54.   Tschopp J, Podack ER, Muller Eberhard HJ. Ultrastructure of the membrane attack complex of complement: Detection of the tetramolecular C9-polymerizing complex C5b-8. Proc Natl Acad Sci U S A. 1982;79(23 I):7474–8.

55.   Podack E. Molecular organization of C9 within the membrane attack complex of complement. Induction of circular C9 polymerization by the C5b-8 assembly. J Exp Med. 1982 Jul 1;156(1):268–82.

56.   Tschopp J. Ultrastructure of the membrane attack complex of complement. Heterogeneity of the complex caused by different degree of C9 polymerization. J Biol Chem. 1984;259(12):7857–63.

57.   Biesecker G, Podack ER, Halverson C a, Müller-Eberhard HJ. C5b-9 dimer: isolation from complement lysed cells and ultrastructural identification with complement-dependent membrane lesions. J Exp Med. 1979;149(2):448–58.

58.   Podack ER, Tschoop J, Muller-Eberhard HJ. Molecular organization of C9 within the membrane attack complex of complement. Induction of circular C9 polymerization by the C5b-8 assembly. J Exp Med. 1982 Jul 1;156(1):268–82.

59.   Podack ER, Biesecker G, Müller-Eberhard HJ. Membrane attack complex of complement: generation of high-affinity phospholipid binding sites by fusion of five hydrophilic plasma proteins. Proc Natl Acad Sci U S A. 1979;76:897–901.

60.   Podack ER, Muller-Eberhard HJ. Binding of desoxycholate, phosphatidylcholine vesicles, lipoprotein and of the S-protein to complexes of terminal complement components. J Immunol. 1978;121(3):1025–30.

61.   Kolb WP, Haxby JA, Arroyave CM, Müller-Eberhard HJ. The membrane attack mechanism of complement. Reversible interactions among the five native components in free solution. J Exp Med. Rockefeller University Press; 1973 Aug 1;138(2):428–37.

62.   Podack ER, Biesecker G, Kolb WP, Müller-Eberhard HJ. The C5b-6 Complex: Reaction with C7, C8, C9. J Immunol. 1978;121(2):484–90.

63.   Preissner KT, Podack ER, Muller-Eberhard HJ. The membrane attack complex of complement: Relation of C7 to the metastable membrane binding site of the intermediate complex C5b-7. J Immunol. 1985;135(1):445–51.

64.   Preissner KT, Podack ER, Muller-Eberhard HJ. Self-association of the seventh component of human complement (C7): Dimerization and polymerization. J Immunol. 1985;135(1):452–8.

65.   Yamamoto KI. Proteolysis of the C5b-7 complex: Cleavage of the C5b and C6 subunits and its effect on the interaction of the complex with phospholipid bilayers. J Immunol. 1980;125(4):1745–50.

66.   Silversmith RE, Nelsestuen GL. Interaction of complement proteins C5b-6 and C5b-7 with phospholipid vesicles: Effects of phospholipid structural features.

142

Biochemistry. American Chemical Society; 1986 Nov 18;25(23):7717–25.

67. Ramm LE, Whitlow MB, Mayer MM. Size of the transmembrane channels produced by complement proteins C5b-8. J Immunol. 1982;129(3):1143–6.

68. Ng SC, Gururaj Rao A, Zack Howard OM, Sodetz JM. The eighth component of human complement: Evidence that it is an oligomeric serum protein assembled from products of three different genes. Biochemistry. 1987;26(17):5229–33.

69. Sodetz JM. Structure and Function of C8 in the Membrane Attack Sequence of Complement. In Springer, Berlin, Heidelberg; 1988. p. 19–31.

70. Monahan JB, Sodetz JM. Binding of the eighth component of human complement to the soluble cytolytic complex is mediated by its β subunit. J Biol Chem. 1980;255(22):10579–82.

71. Stewart JL, Sodetz JM. Analysis of the specific association of the eighth and ninth components of human complement: identification of a direct role for the .alpha. subunit of C8. Biochemistry. American Chemical Society; 1985 Aug;24(17):4598–602.

72. Hu VW, Esser AF, Podack ER, Wisnieski BJ. The membrane attack mechanism of complement: photolabeling reveals insertion of terminal proteins into target membrane. J Immunol. 1981;127(1):380–6.

73. Esser AF, Kolb WP, Podack ER, Müller-Eberhard HJ. Molecular reorganization of lipid bilayers by complement: a possible mechanism for membranolysis. Proc Natl Acad Sci U S A. 1979;76:1410–4.

74. Podack ER, Stoffel W, Esser AF, Müller-Eberhard HJ. Membrane attack complex of complement: distribution of subunits between the hydrocarbon phase of target membranes and water. Proc Natl Acad Sci U S A. 1981;78:4544–8.

75. Fredslund F, Laursen NS, Roversi P, Jenner L, Oliveira CLP, Pedersen JS, Nunn MA, Lea SM, Discipio R, Sottrup-Jensen L, Andersen GR. Structure of and influence of a tick complement inhibitor on human complement component 5. Nat Immunol. 2008 Jul;9(7):753–60.

76. Laursen NS, Andersen KR, Braren I, Spillner E, Sottrup-Jensen L, Andersen GR. Substrate recognition by complement convertases revealed in the C5-cobra venom factor complex. EMBO J. 2011;30:606–16.

77. Aleshin AE, Schraufstatter IU, Stec B, Bankston LA, Liddington RC, DiScipio RG. Structure of complement C6 suggests a mechanism for initiation and unidirectional, sequential assembly of membrane attack complex (MAC). J Biol Chem. 2012 Mar 23;287(13):10210–22.

78. Aleshin AE, DiScipio RG, Stec B, Liddington RC. Crystal structure of C5b-6 suggests structural basis for priming assembly of the membrane attack complex. J Biol Chem. 2012 Jul 1;287(23):19642–52.

79. Hadders MA, Bubeck D, Roversi P, Hakobyan S, Forneris F, Morgan BP,

Pangburn MK, Llorca O, Lea SM, Gros P. Assembly and regulation of the membrane attack complex based on structures of C5b6 and sC5b9. Cell Rep. 2012 Mar 29;1(3):200–7.

80. Hadders MA, Beringer DX, Gros P. Structure of C8alpha-MACPF reveals mechanism of membrane attack in complement immune defense. Science. 2007 Sep 14;317(5844):1552–4.

81. Slade DJ, Lovelace LL, Chruszcz M, Minor W, Lebioda L, Sodetz JM. Crystal structure of the MACPF domain of human complement protein C8 alpha in complex with the C8 gamma subunit. J Mol Biol. 2008 May 30;379(2):331–42.

82. Bubeck D, Roversi P, Donev R, Morgan BP, Llorca O, Lea SM. Structure of human complement C8, a precursor to membrane attack. J Mol Biol. 2011 Jan 14;405(2):325–30.

83. Lovelace LL, Cooper CL, Sodetz JM, Lebioda L. Structure of human C8 protein provides mechanistic insight into membrane pore formation by complement. J Biol Chem. 2011 May 20;286(20):17585–92.

84. Janssen BJC, Huizinga EG, Raaijmakers HCA, Roos A, Daha MR, Nilsson-Ekdahl K, Nilsson B, Gros P. Structures of complement component C3 provide insights into the function and evolution of immunity. Nature. Nature Publishing Group; 2005 Sep 22;437(7058):505–11.

85. Kidmose RT, Laursen NS, Dobó J, Kjaer TR, Sirotkina S, Yatime L, Sottrup-Jensen L, Thiel S, Gál P, Andersen GR, Dobo J, Kjaer TR, Sirotkina S, Yatime L, Sottrup-Jensen L, Thiel S, Gal P, Andersen GR. Structural basis for activation of the complement system by component C4 cleavage. Proc Natl Acad Sci U S A. 2012 Sep 18;109(38):15425–30.

86. Thai CT, Ogata RT. Recombinant C345C and factor I modules of complement components C5 and C7 inhibit C7 incorporation into the complement membrane attack complex. J Immunol. 2005;174(10):6227–32.

87. Thai CT, Ogata RT. Complement components C5 and C7: Recombinant factor I modules of C7 bind to the C345C domain of C5. J Immunol. 2004;173(7):4547–52.

88. Phelan MM, Thai C-T, Soares DC, Ogata RT, Barlow PN, Bramham J. Solution structure of factor I-like modules from complement C7 reveals a pair of follistatin domains in compact pseudosymmetric arrangement. J Biol Chem. 2009 Jul 17;284(29):19637–49.

89. Rosado CJ, Kondos S, Bull TE, Kuiper MJ, Law RHP, Buckle AM, Voskoboinik I, Bird PI, Trapani JA, Whisstock JC, Dunstone MA. The MACPF/CDC family of pore-forming toxins. Cell Microbiol. 2008 Sep;10(9):1765–74.

90. Heuck AP, Hotze EM, Tweten RK, Johnson AE. Mechanism of Membrane Insertion of a Multimeric β-Barrel Protein: Perfringolysin O Creates a Pore Using Ordered and Coupled Conformational Changes. Mol Cell. Cell Press;

144

2000 Nov 1;6(5):1233–42.

91. Shatursky O, Heuck AP, Shepard LA, Rossjohn J, Parker MW, Johnson AE, Tweten RK. The Mechanism of Membrane Insertion for a Cholesterol-Dependent Cytolysin. Cell. 1999 Oct;99(3):293–9.

92. Rossjohn J, Feil SC, McKinstry WJ, Tweten RK, Parker MW. Structure of a cholesterol-binding, thiol-activated cytolysin and a model of its membrane form. Cell. 1997;89(5):685–92.

93. Rosado CJ, Buckle AM, Law RHP, Butcher RE, Kan W-TW-T, Bird CH, Ung K, Browne KA, Baran K, Bashtannyk-Puhalovich TA, Faux NG, Wong W, Porter CJ, Pike RN, Ellisdon AM, Pearce MC, Bottomley SP, Emsley J, Smith AI, Rossjohn J, Hartland EL, Voskoboinik I, Trapani JA, Bird PI, Dunstone MA, Whisstock JC. A common fold mediates vertebrate defense and bacterial attack. Science. 2007 Sep 14;317(5844):1548–51.

94. Hotze EM, Heuck AP, Czajkowsky DM, Shao Z, Johnson AE, Tweten RK. Monomer-monomer interactions drive the prepore to pore conversion of a β-barrel-forming cholesterol-dependent cytolysin. J Biol Chem. 2002;277:11597–605.

95. Tilley SJ, Orlova E V., Gilbert RJC, Andrew PW, Saibil HR. Structural basis of pore formation by the bacterial toxin pneumolysin. Cell. 2005;121:247–56.

96. Czajkowsky DM, Hotze EM, Shao Z, Tweten RK. Vertical collapse of a cytolysin prepore moves its transmembrane beta-hairpins to the membrane. EMBO J. 2004;23:3206–15.

97. Reboul CF, Whisstock JC, Dunstone MA. A new model for pore formation by cholesterol-dependent cytolysins. PLoS Comput Biol. 2014 Aug 21;10(8):e1003791.

98. Lukoyanova N, Kondos SC, Farabella I, Law RHP, Reboul CF, Caradoc-Davies TT, Spicer BA, Kleifeld O, Traore DAK, Ekkel SM, Voskoboinik I, Trapani JA, Hatfaludi T, Oliver K, Hotze EM, Tweten RK, Whisstock JC, Topf M, Saibil HR, Dunstone MA. Conformational Changes during Pore Formation by the Perforin-Related Protein Pleurotolysin. PLoS Biol. 2015 Feb 5;13(2):e1002049.

99. Law RHP, Lukoyanova N, Voskoboinik I, Caradoc-Davies TT, Baran K, Dunstone MA, D'Angelo ME, Orlova E V, Coulibaly F, Verschoor S, Browne KA, Ciccone A, Kuiper MJ, Bird PI, Trapani JA, Saibil HR, Whisstock JC. The structural basis for membrane binding and pore formation by lymphocyte perforin. Nature. 2010;468:447–51.

100. Ellisdon AM, Reboul CF, Panjikar S, Huynh K, Oellig CA, Winter KL, Dunstone MA, Hodgson WC, Seymour J, Dearden PK, Tweten RK, Whisstock JC, McGowan S. Stonefish toxin defines an ancient branch of the perforin-like superfamily. Proc Natl Acad Sci. National Academy of Sciences; 2015 Dec 15;112(50):15360–5.

101. Morgan BP, Walters D, Serna M, Bubeck D. Terminal complexes of the

complement system: new structural insights and their relevance to function. Immunological Reviews. 2016.

102. Bayly-Jones C, Bubeck D, Dunstone MA. The mystery behind membrane insertion: a review of the complement membrane attack complex. Philos Trans R Soc B Biol Sci. 2017;372:10.1098/rstb.2016.0221.

103. Morgan BP, Boyd C, Bubeck D. Molecular cell biology of complement membrane attack. Semin Cell Dev Biol. Academic Press; 2017 Dec 1;72:124–32.

104. Serna M, Giles JL, Morgan BP, Bubeck D. Structural basis of complement membrane attack complex formation. Nat Commun. Nature Publishing Group; 2016 Jan 4;7:10587.

105. Sharp TH, Koster AJ, Gros P. Heterogeneous MAC Initiator and Pore Structures in a Lipid Bilayer by Phase-Plate Cryo-electron Tomography. Cell Rep. Elsevier; 2016 Mar;15:10.1016/j.celrep.2016.03.002.

106. Dudkina N V, Spicer BA, Reboul CF, Conroy PJ, Lukoyanova N, Elmlund H, Law RHP, Ekkel SM, Kondos SC, Goode RJA, Ramm G, Whisstock JC, Saibil HR, Dunstone MA. Structure of the poly-C9 component of the complement membrane attack complex. Nat Commun. Nature Publishing Group; 2016 Jan 4;7:10588.

107. Tschopp J, Muller-Eberhard HJ, Podack ER. Formation of transmembrane tubules by spontaneous polymerization of the hydrophilic complement protein C9. Nature. 1982;298(5874):534–8.

108. Podack ER, Tschopp J. Polymerization of the ninth component of complement (C9): Formation of poly(C9) with a tubular ultrastructure resembling the membrane attack complex of complement. Proc Natl Acad Sci U S A. 1982;79(2 I):574–8.

109. Podacks ER, Tschopps J. Circular Polymerization of the Ninth Component of Complement. J Biol Chem. 1982;257(24):15204–12.

110. Tschopp J. Circular polymerization of the membranolytic ninth component of complement. Dependence on metal ions. J Biol Chem. 1984;259(16):10569–73.

111. Dankert JR, Shiver JW, Esser AF. Ninth component of complement: self-aggregation and interaction with lipids. Biochemistry. 1985;24(11):2754–62.

112. Falk RJ, Dalmasso AP, Kim Y, Lam S, Michael A. Radioimmunoassay of the Attack Complex of Complement in Serum from Patients with Systemic Lupus Erythematosus. N Engl J Med. 1985 Jun 20;312(25):1594–9.

113. Hanahan D, Jessee J, Bloom FR. [4] Plasmid transformation of Escherichia coli and other bacteria. Methods Enzymol. Academic Press; 1991 Jan 1;204:63–113.

114. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF. Protein identification and analysis tools in the ExPASy server.

Methods Mol Biol. 1999;112:531–52.

115. Bradford M. A rapid and sensitive method for the quantification of microgram quantities of protein using the principle of protein-dye binding. Anal Biochem. 1976;72:248–54.

116. Biesecker G, Müller-Eberhard HJ. The ninth component of human complement: purification and physicochemical characterization. J Immunol. American Association of Immunologists; 1980 Mar 1;124(3):1291–6.

117. Biesecker G, Lachmann P, Henderson R. Structure of complement poly-C9 determined in projection by cryo-electron microscopy and single particle analysis. Mol Immunol. 1993 Oct;30(15):1369–82.

118. DeLano WL. The PyMOL Molecular Graphics System, Version 1.8. Schrödinger LLC. 2014;

119. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. J Comput Chem. 2004;25:1605–12.

120. Morgan BP. Mechanisms of tissue damage by the membrane attack complex of complement. Complement Inflamm. 1989;6(2):104–11.

121. DiScipio RG, Gehring MR, Podack ER, Kan CC, Hugli TE, Fey GH. Nucleotide sequence of cDNA and derived amino acid sequence of human complement component C9. Proc Natl Acad Sci. 1984 Dec 1;81(23):7298–302.

122. Esser AF. The membrane attack complex of complement. Assembly, structure and cytotoxic activity. Toxicology. 1994 Feb;87(1–3):229–47.

123. Tomlinson S, Ueda E, Maruniak JE, Garciacanedo A, Bjes ES, Esser AF. The Expression of Hemolytically Active Human Complement Protein C9 in Mammalian, Insect, and Yeast Cells. Protein Expr Purif. 1993;4(2):141–8.

124. Taylor KM, Morgan BP, Campbell AK. Altered glycosylation and selected mutation in recombinant human complement component C9: effects on haemolytic activity. Immunology. 1994 Dec;83(3):501–6.

125. Rossi V, Wang Y, Esser AF. Topology of the membrane-bound form of complement protein C9 probed by glycosylation mapping, anti-peptide antibody binding, and disulfide modification. Mol Immunol. 2010 May;47(7–8):1553–60.

126. Taylor KM, Trimby AR, Campbell AK. Mutation of recombinant complement component C9 reveals the significance of the N-terminal region for polymerization. Immunology. 1997;91(1):20–7.

127. Dupuis M, Peitsch MC, Hamann U, Stanley KK, Tschopp J. Mutations in the putative lipid-interaction domain of complement C9 result in defective secretion of the functional protein. Mol Immunol. 1993 Jan;30(1):95–100.

128. Hüsler T, Lockert DH, Kaufman KM, Sodetz JM, Sims PJ. Chimeras of Human

Complement C9 Reveal the Site Recognized by Complement Regulatory Protein CD59. J Biol Chem. American Society for Biochemistry and Molecular Biology; 1995 Feb 24;270(8):3483–6.

129. Tomlinson S, Wang Y, Ueda E, Esser AF. Chimeric horse/human recombinant C9 proteins identify the amino acid sequence in horse C9 responsible for restriction of hemolysis. J Immunol. 1995;155(1):436–44.

130. Rupp B. Biomolecular crystallography: Principles, practice, and application to structural biology. Garland Science. 2009. 800 p.

131. Dunstone MA. A Structural Investigation of Human Serum Proteins Ph.D Thesis. St. Vincent's Hospital, The University of Melbourne; 2004.

132. Shi X, Jarvis D. Protein N-Glycosylation in the Baculovirus-Insect Cell System. Curr Drug Targets. 2007 Oct 1;8(10):1116–25.

133. Graham FL, Smiley J, Russell WC, Nairn R. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. J Gen Virol. 1977;36(1):59–74.

134. Portolano N, Watson PJ, Fairall L, Millard CJ, Milano CP, Song Y, Cowley SM, Schwabe JWR. Recombinant protein expression for structural biology in HEK 293F suspension cells: a novel and accessible approach. J Vis Exp. MyJoVE Corporation; 2014 Oct 16;(92):e51897.

135. Goldschmidt L, Cooper DR, Derewenda ZS, Eisenberg D. Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. Protein Sci. 2007 Aug;16(8):1569–76.

136. Cooper DR, Surendranath Y, Devedjiev Y, Bielnicki J, Derewenda ZS, IUCr. Structure of the *Bacillus subtilis* OhrB hydroperoxide-resistance protein in a fully oxidized state. Acta Crystallogr Sect D Biol Crystallogr. International Union of Crystallography; 2007 Dec 1;63(12):1269–73.

137. Stanley KK, Herz J. Topological mapping of complement component C9 by recombinant DNA techniques suggests a novel mechanism for its insertion into target membranes. EMBO J. 1987 Jul;6(7):1951–7.

138. Mondragón-Palomino M, Piñero D, Nicholson-Weller A, Laclette JP. Phylogenetic analysis of the homologous proteins of the terminal complement complex supports the emergence of C6 and C7 followed by C8 and C9. J Mol Evol. 1999;49(2).

139. Dunstone MA, Tweten RK. Packing a punch: The mechanism of pore formation by cholesterol dependent cytolysins and membrane attack complex/perforin-like proteins. Current Opinion in Structural Biology. 2012. p. 342–9.

140. Shepard LA, Shatursky O, Johnson AE, Tweten RK. The mechanism of pore assembly for a cholesterol-dependent cytolysin: Formation of a large prepore complex precedes the insertion of the transmembrane β-hairpins. Biochemistry. 2000 Aug;39(33):10284–93.

148

141. Berends ETM, Dekkers JF, Nijland R, Kuipers A, Soppe JA, van Strijp JAG, Rooijakkers SHM. Distinct localization of the complement C5b-9 complex on Gram-positive bacteria. Cell Microbiol. 2013;15(12):1955–68.

142. Walport MJ. Advances in immunology: Complement (first of two parts). N Engl J Med. 2001;344(14):1058–66.

143. Scibek JJ, Plumb ME, Sodetz JM. Binding of Human Complement C8 to C9: Role of the N-Terminal Modules in the C8α Subunit †. Biochemistry. American Chemical Society; 2002 Dec 14;41(49):14546–51.

144. Bodian DL, Davis SJ, Morgan BP, Rushmere NK. Mutational Analysis of the Active Site and Antibody Epitopes of the Complement-inhibitory Glycoprotein, CD59. J Exp Med. 1997 Feb 3;185(3):507–16.

145. Berends ETM, Kuipers A, Ravesloot MM, Urbanus RT, Rooijakkers SHM. Bacteria under stress by complement and coagulation. FEMS Microbiol Rev. 2014 Nov;38(6):1146–71.

146. Cascales E, Buchanan SK, Duche D, Kleanthous C, Lloubes R, Postle K, Riley M, Slatin S, Cavard D. Colicin Biology. Microbiol Mol Biol Rev. 2007;71(1):158–229.

147. Kell DB, Clarke DJ, Morris JG. On proton-coupled information transfer along the surface of biological membranes and the mode of action of certain colicins. FEMS Microbiol Lett. No longer published by Elsevier; 1981 May 1;11(1):1–11.

148. Fernie-King BA, Seilly DJ, Willers C, Wurzner R, Davies A, Lachmann PJ. Streptococcal inhibitor of complement (SIC) inhibits the membrane attack complex by preventing uptake of C567 onto cell membranes. Immunology. Wiley/Blackwell (10.1111); 2001 Jul;103(3):390–8.

149. Chen M, Daha MR, Kallenberg CGM. The complement system in systemic autoimmune disease. J Autoimmun. 2010;34(3):1–11.

150. Park JL, Lucchesi BR. Mechanisms of myocardial reperfusion injury. ATS. 1999;68(5):1905–12.

151. Morgan BP, Daniels RH, Williams BD. Measurement of terminal complement complexes in rheumatoid arthritis. Clin Exp Immunol. Wiley-Blackwell; 1988 Sep;73(3):473–8.

Appendix 1: The following pages contain the mass spectroscopy results from murine C9 crystals as referenced in Figure 4-11.

Monash Biomedical Proteomics facility

Report: nanoLC ESI MS/MS analysis

Facility Reference: 2017_031

Prepared by: Dr David Steer

Sample details

Name: Bradley Spicer
Date sample processed: 7/03/2017

Sample Preparation

For LCMS analysis, the gel pieces were washed and the protein was reduced and alkylated with DTT\Iodoacetamide and digested overnight with trypsin in a total of 50ul of 20mM ammonium Bicarbonate buffer. No Further clean up prior to analysis.

Acquisition

Tryptic digests were analysed by LC-MS/MS by separation over a 30 minute gradient on a Thermo RSLC pepmap100, 50cm reversed phase nano column at a flow rate of 300nl/minute. The eluant is nebulised and ionised using the Thermo nano electrospray source. Peptides are selected for MSMS analysis in Full MS/dd-MS^2 (TopN) mode with the following parameter settings: TopN 10, resolution 17500, MSMS AGC target 1e5, 60ms Max IT, NCE 27 and 3 m/z isolation window. Underfill ratio was at 10% and dynamic exclusion was set to 15 seconds.

Instrumentation

LCMS

MS Instrument:	**QExactive** mass spectrometer, (ThermoFisher scientific)
LC instrument:	Ultimate 3000 nano RSLC (ThermoFisher scientific)

Data processing

Data from LCMSMS run was exported to Mascot generic file format (*.mgf) using proteowizard open source software and searched against an the swissprot database using the MASCOT search engine (version 2.4, Matrix Science Inc., London, UK) with all taxonomy selected. The following search parameters were used: missed cleavages, 1; peptide mass tolerance, ± 20 ppm Da; peptide fragment tolerance, ± 20mmu; peptide charge, 2+, 3+ and 4+; fixed modifications, carbamidomethyl; Variable modification, oxidation (Met).

Results

{MATRIX SCIENCE} Mascot Search Results

```
User                      :
Email                     :
Search title              : Trypsin Digests
MS data file              : C:\Users\mascot\Desktop\david mgf's\2017_031\P120170308_DSteer_031_BS52kDa.mgf
Database                  : SwissProt 2016_11 (553231 sequences; 197953409 residues)
Taxonomy                  : Mus musculus (house mouse) (16846 sequences)
Timestamp                 : 9 Mar 2017 at 05:29:38 GMT
Warning                   : Max number of ions is 10000. Ignoring ms-ms set starting at line 4684172
Warning                   : Max number of ions is 10000. Ignoring ms-ms set starting at line 4779533
Warning                   : Max number of ions is 10000. Ignoring ms-ms set starting at line 4847768
Warning                   : Max number of ions is 10000. Ignoring ms-ms set starting at line 4909433
Warning                   : Max number of ions is 10000. Ignoring ms-ms set starting at line 4966797
Enzyme                    : Trypsin
Fixed modifications       : Carbamidomethyl (C)
Variable modifications    : Oxidation (M)
Mass values               : Monoisotopic
Protein Mass              : Unrestricted
Peptide Mass Tolerance    : ± 20 ppm (# ¹³C = 2)
Fragment Mass Tolerance:  : ± 20 mmu
Max Missed Cleavages      : 2
Instrument type           : ESI-QUAD
Number of queries         : 9424
Protein hits              : CO9_MOUSE   Complement component C9 OS=Mus musculus GN=C9 PE=1 SV=2
                            K1C10 MOUSE  Keratin, type I cytoskeletal 10 OS=Mus musculus GN=Krt10 PE=1 SV=3
                            ENOG MOUSE   Gamma-enolase OS=Mus musculus GN=Eno2 PE=1 SV=2
                            K2C73 MOUSE  Keratin, type II cytoskeletal 73 OS=Mus musculus GN=Krt73 PE=1 SV=1
                            CHTOP MOUSE  Chromatin target of PRMT1 protein OS=Mus musculus GN=Chtop PE=1 SV=2
                            H2B1B MOUSE  Histone H2B type 1-B OS=Mus musculus GN=Hist1h2bb PE=1 SV=3
                            K2C1 MOUSE   Keratin, type II cytoskeletal 1 OS=Mus musculus GN=Krt1 PE=1 SV=4
                            H2B2E MOUSE  Histone H2B type 2-E OS=Mus musculus GN=Hist2h2be PE=1 SV=3
                            TBA1A MOUSE  Tubulin alpha-1A chain OS=Mus musculus GN=Tuba1a PE=1 SV=1
                            K2C5 MOUSE   Keratin, type II cytoskeletal 5 OS=Mus musculus GN=Krt5 PE=1 SV=1
                            ACTB MOUSE   Actin, cytoplasmic 1 OS=Mus musculus GN=Actb PE=1 SV=1
                            K1C27 MOUSE  Keratin, type I cytoskeletal 27 OS=Mus musculus GN=Krt27 PE=1 SV=1
                            H2A1F MOUSE  Histone H2A type 1-F OS=Mus musculus GN=Hist1h2af PE=1 SV=3
                            H4 MOUSE     Histone H4 OS=Mus musculus GN=Hist1h4a PE=1 SV=2
                            LDHB MOUSE   L-lactate dehydrogenase B chain OS=Mus musculus GN=Ldhb PE=1 SV=2
                            HNRPK MOUSE  Heterogeneous nuclear ribonucleoprotein K OS=Mus musculus GN=Hnrnpk PE=1 SV=1
                            K1C13 MOUSE  Keratin, type I cytoskeletal 13 OS=Mus musculus GN=Krt13 PE=1 SV=2
                            K2C79 MOUSE  Keratin, type II cytoskeletal 79 OS=Mus musculus GN=Krt79 PE=1 SV=2
                            KPYM MOUSE   Pyruvate kinase PKM OS=Mus musculus GN=Pkm PE=1 SV=4
                            PPIA MOUSE   Peptidyl-prolyl cis-trans isomerase A OS=Mus musculus GN=Ppia PE=1 SV=2
                            K22E MOUSE   Keratin, type II cytoskeletal 2 epidermal OS=Mus musculus GN=Krt2 PE=1 SV=1
                            EF1A1 MOUSE  Elongation factor 1-alpha 1 OS=Mus musculus GN=Eef1a1 PE=1 SV=3
                            RS2 MOUSE    40S ribosomal protein S2 OS=Mus musculus GN=Rps2 PE=1 SV=3
                            HSP7C MOUSE  Heat shock cognate 71 kDa protein OS=Mus musculus GN=Hspa8 PE=1 SV=1
                            IP4A1 MOUSE  Eukaryotic initiation factor 4A-I OS=Mus musculus GN=Eif4a1 PE=1 SV=1
                            ACTA MOUSE   Actin, aortic smooth muscle OS=Mus musculus GN=Acta2 PE=1 SV=1
                            HB22 MOUSE   H-2 class II histocompatibility antigen, E-D beta chain OS=Mus musculus PE=3 SV=1
                            AAAT MOUSE   Neutral amino acid transporter B(0) OS=Mus musculus GN=Slc1a5 PE=1 SV=2
                            ANXA2 MOUSE  Annexin A2 OS=Mus musculus GN=Anxa2 PE=1 SV=2
                            H12 MOUSE    Histone H1.2 OS=Mus musculus GN=Hist1h1c PE=1 SV=2
                            RL10A MOUSE  60S ribosomal protein L10a OS=Mus musculus GN=Rpl10a PE=1 SV=3
                            EWS MOUSE    RNA-binding protein EWS OS=Mus musculus GN=Ewsr1 PE=1 SV=2
                            K22O MOUSE   Keratin, type II cytoskeletal 2 oral OS=Mus musculus GN=Krt76 PE=1 SV=1
                            K1C42 MOUSE  Keratin, type I cytoskeletal 42 OS=Mus musculus GN=Krt42 PE=1 SV=1
                            HYEP MOUSE   Epoxide hydrolase 1 OS=Mus musculus GN=Ephx1 PE=1 SV=2
                            HS90A MOUSE  Heat shock protein HSP 90-alpha OS=Mus musculus GN=Hsp90aa1 PE=1 SV=4
                            TBB2A MOUSE  Tubulin beta-2A chain OS=Mus musculus GN=Tubb2a PE=1 SV=1
                            RL11 MOUSE   60S ribosomal protein L11 OS=Mus musculus GN=Rpl11 PE=1 SV=4
                            TEX2 MOUSE   Testis-expressed sequence 2 protein OS=Mus musculus GN=Tex2 PE=1 SV=2
                            ZN488 MOUSE  Zinc finger protein 488 OS=Mus musculus GN=Znf488 PE=2 SV=2
                            CALM MOUSE   Calmodulin OS=Mus musculus GN=Calm1 PE=1 SV=2
                            TMPSD MOUSE  Transmembrane protease serine 13 OS=Mus musculus GN=Tmprss13 PE=2 SV=2
                            G3P MOUSE    Glyceraldehyde-3-phosphate dehydrogenase OS=Mus musculus GN=Gapdh PE=1 SV=2
                            RS27A MOUSE  Ubiquitin-40S ribosomal protein S27a OS=Mus musculus GN=Rps27a PE=1 SV=2
                            ROA2 MOUSE   Heterogeneous nuclear ribonucleoproteins A2/B1 OS=Mus musculus GN=Hnrnpa2b1 PE=1 SV=2
```

Select Summary Report

Format As [Select Summary (protein hits) ▼]   Help

Significance threshold p< [0.05]   Max. number of hits [AUTO]

Standard scoring ○ MudPIT scoring ● Ions score or expect cut-off [0]   Show sub-sets [0]

Show pop-ups ● Suppress pop-ups ○   Require bold red ☑

[Re-Search]   ● All queries   ○ Unassigned   ○ Below homology threshold   ○ Below identity threshold

```
1.  CO9 MOUSE    Mass: 63217    Score: 36815   Matches: 1379(899)   Sequences: 49(43)   emPAI: 104.60
    Complement component C9 OS=Mus musculus GN=C9 PE=1 SV=2
    Query    Observed    Mr(expt)    Mr(calc)    ppm    Miss  Score    Expect  Rank  Unique   Peptide
```

83	401.7221	801.4296	801.4320	-3.00	0	(32)	0.033	1	U	K.MFVHLR.G 68 69 70 71 73 74 75 76 77 78 79 80 81 82 84 85 86 8
131	407.7540	813.4934	813.4960	-3.20	0	43	0.0015	1	U	R.EQAILLK.E 123 124 125 126 127 128 129 130 132 133 134 135 1
157	409.7190	817.4235	817.4269	-4.10	0	32	0.036	1	U	K.MFVHLR.G 150 151 152 153 154 155 156 158 159 160 161 162 16
262	425.7537	849.4929	849.4960	-3.67	1	31	0.029	1	U	K.IKDAYIK.K 251 253 254 255 256 257 258 259 260 261 263 264 2
286	430.2284	858.4422	858.4447	-2.86	0	32	0.022	1	U	K.GVDLNDVK.H 283 284 285 287 288 289 290
338	436.2565	870.4985	870.5035	-5.84	1	36	0.0075	1	U	R.NRDVVLR.S 339 340 341 342 343 344 345 346 347 348 349
407	440.2070	878.3994	878.3997	-0.29	0	40	0.003	1	U	K.FSYFMGK.N 378 379 380 381 382 383 384 385 386 387 388 389 3
465	448.2026	894.3906	894.3946	-4.41	0	(23)	0.08	1	U	K.FSYFMGK.N 464 466 467 468 469 470 471 472 473 474 475 476 4
527	454.7383	907.4620	907.4651	-3.37	0	33	0.027	1	U	K.ALPTSYEK.G 519 520 521 522 523 524 525 526 528 529 530 531
567	462.7370	923.4595	923.4600	-0.55	0	36	0.011	1	U	R.STFLDDVK.A 559 560 561 562 563 564 565 566 568 569 570
592	465.7691	929.5237	929.5269	-3.44	1	(53)	0.00026	1	U	K.KMFVHLR.G 586 587 588 589 590 591 593
644	471.2682	940.5218	940.5229	-1.23	0	47	0.00051	1	U	R.IPLQDDLK.D 625 626 627 628 629 630 631 632 633 634 635 636
679	473.7671	945.5197	945.5218	-2.29	1	54	0.00031	1	U	K.KMFVHLR.G 672 673 674 675 676 677 678 680 681
783	489.8009	977.5873	977.5909	-3.72	2	34	0.0082	1	U	K.IKDAYIKK.Q 784 785
888	504.2542	1006.4938	1006.4971	-3.27	0	51	0.00036	1	U	K.FSATEVPEK.G 881 882 883 884 885 886 887 889 890 891 892 89
945	510.7419	1019.4692	1019.4706	-1.38	0	58	3.8e-005	1	U	K.SCVDVLGDR.Q 935 936 937 938 939 940 941 942 943 944 946
1041	523.7596	1045.5047	1045.5080	-3.16	0	47	0.00073	1	U	R.LSSYFSQSK.K 1035 1036 1037 1038 1039 1040 1042 1043 1044
1171	537.2455	1072.4764	1072.4794	-2.78	0	52	9.9e-005	1	U	R.GMACEIHQK.I 1169 1170 1172
1233	545.2436	1088.4727	1088.4743	-1.47	0	(45)	0.00037	1	U	R.GMACEIHQK.I 1225 1226 1227 1228 1229 1230 1231 1232
1335	558.7971	1115.5797	1115.5822	-2.26	1	67	1.1e-005	1	U	K.EKGVDLNDVK.H 1330 1331 1332 1333 1334 1335 1336 1337
1449	577.7890	1153.5635	1153.5630	0.41	1	15	1.3	1	U	K.FKFSYFMGK.N
1502	587.8100	1173.6054	1173.6030	2.11	1	20	0.62	1	U	R.LSSYFSQSKK.M
1519	591.3180	1180.6215	1180.6241	-2.18	0	52	0.00032	1	U	R.SILAFGQFNGK.S 1517 1518 1520 1523 1524 1525 1526 1527 152
1522	394.5481	1180.6224	1180.6241	-1.40	0	(41)	0.0039	1	U	R.SILAFGQFNGK.S 1521
1576	601.8102	1201.6058	1201.6091	-2.76	1	34	0.017	1	U	R.RLSSYFSQSK.K 1568 1570 1571 1572 1573 1574 1575
1604	402.2100	1203.6082	1201.6091	1664	1	(22)	0.31	1	U	R.RLSSYFSQSK.K 1569
1691	409.8706	1226.5900	1226.5931	-2.59	0	(44)	0.0012	1	U	K.TSNFNADFALK.F 1693
1700	614.3033	1226.5921	1226.5931	-0.89	0	72	2.3e-006	1	U	K.TSNFNADFALK.F 1680 1681 1683 1684 1685 1686 1687 1688 168
1921	421.5301	1261.5684	1261.5696	-0.92	0	(34)	0.0067	1	U	K.HCLGFNMDLR.I 1911 1912 1913 1916 1937
1925	631.7925	1261.5704	1261.5696	0.63	0	54	7e-005	1	U	K.HCLGFNMDLR.I 1909 1910 1914 1915 1917 1918 1919 1920 1922
1990	426.8599	1277.5578	1277.5645	-5.26	0	(22)	0.075	1	U	K.HCLGFNMDLR.I 1991 1992 1993 2000 2005
1996	639.7870	1277.5595	1277.5645	-3.90	0	(51)	0.0001	1	U	K.HCLGFNMDLR.I 1994 1995 1997 1998 1999 2002 2003 2004 2006
2058	430.2483	1287.7230	1287.7260	-2.36	0	(57)	9.4e-005	1	U	R.MSPIYNLIPLK.I 2062 2071
2074	644.8703	1287.7261	1287.7260	0.05	0	58	6.2e-005	1	U	R.MSPIYNLIPLK.I 2055 2056 2057 2059 2060 2061 2063 2064 206
2193	435.5801	1303.7184	1303.7210	-1.96	0	(25)	0.16	1	U	R.MSPIYNLIPLK.I 2195 2196 2197
2206	652.8693	1303.7241	1303.7210	2.41	0	(57)	8.7e-005	1	U	R.MSPIYNLIPLK.I 2190 2191 2194 2198 2199 2200 2201 2202 220
2421	472.8878	1415.6417	1415.6456	-2.77	0	(56)	5.1e-005	1	U	K.AVEDYIDEFSTK.R 2422 2429 2431
2452	708.8307	1415.6468	1415.6456	0.85	0	92	1.4e-008	1	U	K.AVEDYIDEFSTK.R 2420 2423 2424 2425 2426 2427 2428 2430 24
2514	473.9219	1418.7439	1418.7405	2.38	0	(73)	3.2e-006	1	U	R.DNIIDDVISFIR.G 2489 2505 2506 2507 2512 2515 2525
2526	710.3826	1418.7506	1418.7405	7.11	0	81	4.6e-007	1	U	R.DNIIDDVISFIR.G 2474 2475 2476 2477 2478 2479 2480 2481 24
2715	742.8710	1483.7274	1483.7307	-2.22	1	92	2.5e-008	1	U	R.EKTSNFNADFALK.F 2714 2722 2723 2724
2716	495.5832	1483.7278	1483.7307	-1.97	1	(41)	0.003	1	U	R.EKTSNFNADFALK.F 2713 2719 2721
2764	754.3550	1506.6955	1506.6984	-1.97	0	104	1.2e-009	1	U	K.DASVTASVNADGCIK.T 2759 2760 2761 2762 2763 2765 2766 276
2816	762.3527	1522.6908	1522.6940	-2.05	0	66	6.7e-006	1	U	R.TENYDEHLEVFK.A 2796 2797 2800 2801 2802 2805 2807 2808 28
2817	508.5709	1522.6909	1522.6940	-2.02	0	(45)	0.00074	1	U	R.TENYDEHLEVFK.A 2789 2794 2795 2798 2799 2803 2804 2809 28
2992	786.8805	1571.7465	1571.7467	-0.16	1	77	6.9e-007	1	U	K.AVEDYIDEFSTKR.C 2977 2978 2982 2984 2985 2989 2991 2994 3
2999	524.9233	1571.7481	1571.7467	0.85	1	(59)	4.8e-005	1	U	K.AVEDYIDEFSTKR.C 2976 2979 2980 2981 2983 2986 2987 2990 2
3026	394.9733	1575.8642	1575.8661	-1.17	0	(54)	0.00019	1	U	R.KPWNVVSLIYETK.A
3067	526.2961	1575.8664	1575.8661	0.20	0	(55)	0.00012	1	U	R.KPWNVVSLIYETK.A 3010 3013 3015 3016 3018 3020 3022 3027 3
3104	788.9421	1575.8696	1575.8661	2.23	0	103	1.8e-009	1	U	R.KPWNVVSLIYETK.A 3009 3011 3012 3014 3017 3019 3021 3023 3
3555	874.3710	1746.7274	1746.7308	-1.92	0	100	6.1e-010	1	U	R.TPFDNEFYNGLCDR.V 3545 3546 3547 3548 3549 3552 3553 3556
3559	583.2504	1746.7294	1746.7308	-0.82	0	(53)	3.2e-005	1	U	R.TPFDNEFYNGLCDR.V 3551 3554 3557 3576 3577 3581 3589 3593
3733	907.4089	1812.8033	1812.8101	-3.73	0	87	3.8e-008	1	U	R.EEQEQHYPIPIDCR.M 3734 3735 3736 3741 3744 3745 3746 3751
3752	605.2766	1812.8081	1812.8101	-1.10	0	(47)	0.00038	1	U	R.EEQEQHYPIPIDCR.M 3732 3737 3738 3739 3740 3742 3743 3747
3762	605.3104	1812.9095	1812.9145	-2.80	1	(17)	1.1	1	U	R.STFLDDVKALPTSYEK.G
3763	907.4658	1812.9170	1812.9145	1.34	1	57	0.0001	1	U	R.STFLDDVKALPTSYEK.G
3875	631.0143	1890.0210	1890.0251	-2.15	1	14	1.6	1	U	R.KPWNVVSLIYETKADK.S
4090	984.4818	1966.9490	1965.9756	495	0	24	0.16	1	U	K.GAGEVSPAEHSSKPTNISAK.F 4087 4094 4096 4105 4107 4111 411
4195	495.2464	1976.9566	1976.9592	-1.32	1	(40)	0.0043	1	U	R.TENYDEHLEVFKAINR.E
4196	659.9942	1976.9608	1976.9592	0.85	1	45	0.0015	1	U	R.TENYDEHLEVFKAINR.E
4197	989.4884	1976.9622	1976.9592	1.53	1	(38)	0.0069	1	U	R.TENYDEHLEVFKAINR.E
4379	676.9970	2027.9691	2027.9687	0.16	1	55	0.00013	1	U	K.QNLEKAVEDYIDEFSTK.R 4380
4544	702.3803	2104.1190	2103.1324	469	0	28	0.06	1	U	K.TVNITRDNIIDDVISFIR.G 4540 4541 4542
4545	1053.0692	2104.1238	2103.1324	471	1	(2)	23	1	U	K.TVNITRDNIIDDVISFIR.G
4621	534.0200	2132.0509	2132.0650	-6.63	0	(58)	7.4e-005	1	U	K.TDFANWASSLANAPALISQR.M
4623	1067.0373	2132.0601	2132.0650	-2.32	0	140	5.2e-013	1	U	K.TDFANWASSLANAPALISQR.M 4625 4628 4630 4634 4635 4636 463
4633	711.6955	2132.0648	2132.0650	-0.11	0	(86)	1.3e-007	1	U	K.TDFANWASSLANAPALISQR.M 4620 4622 4624 4626 4627 4629 463
5184	810.7432	2429.2078	2429.2108	-1.23	1	105	1.3e-009	1	U	R.IPLQDDLKDASVTASVNADGCIK.T 5183 5186 5187 5188 5189 5191
5185	1215.6112	2429.2079	2429.2108	-1.21	1	(76)	1.2e-006	1	U	R.IPLQDDLKDASVTASVNADGCIK.T 5190 5194
5325	633.8312	2531.2956	2531.2941	0.60	0	(51)	0.00037	1	U	R.VAEESELGLTAGYGINILGMEPLR.T 5313 5319 5374
5331	844.7730	2531.2972	2531.2941	1.24	0	(90)	4.6e-008	1	U	R.VAEESELGLTAGYGINILGMEPLR.T 5314 5315 5316 5318 5320 532
5357	1266.6590	2531.3035	2531.2941	3.71	0	137	8.4e-013	1	U	R.VAEESELGLTAGYGINILGMEPLR.T 5309 5310 5311 5312 5317 532
5419	637.8286	2547.2851	2547.2890	-1.52	0	(44)	0.0017	1	U	R.VAEESELGLTAGYGINILGMEPLR.T 5424 5446 5463
5420	1274.6500	2547.2854	2547.2890	-1.40	0	(116)	1.1e-010	1	U	R.VAEESELGLTAGYGINILGMEPLR.T 5411 5412 5415 5418 5421 542
5429	850.1030	2547.2873	2547.2890	-0.68	0	(109)	5.6e-010	1	U	R.VAEESELGLTAGYGINILGMEPLR.T 5410 5413 5414 5416 5417 542
5720	656.8311	2623.2953	2623.3030	-2.94	1	(36)	0.011	1	U	K.TFDKTDFANWASSLANAPALISQR.M 5719 5721 5759 5760
5726	1312.6561	2623.2976	2623.3030	-2.06	1	128	7.3e-012	1	U	K.TFDKTDFANWASSLANAPALISQR.M 5718 5725 5731 5736 5743 576
5735	875.4413	2623.3021	2623.3030	-0.36	1	(80)	5e-007	1	U	K.TFDKTDFANWASSLANAPALISQR.M 5717 5722 5723 5724 5727 572
5978	911.9901	2732.9484	2732.9497	-0.48	0	(97)	2.2e-008	1	U	R.LLCNGDNDCGDYSDENDCDDDPR.T 5970 5971 5974 5971 5987 5988
5989	1367.9800	2733.9454	2732.9497	364	0	146	2.5e-015	1	U	R.LLCNGDNDCGDYSDENDCDDDPR.T 5972 5973 5975 5976 5986
6127	935.1473	2802.4201	2802.4221	-0.73	1	96	1.2e-008	1	U	R.DRVAEESELGLTAGYGINILGMEPLR.T 6128 6129 6132
6131	701.8600	2803.4109	2802.4221	353	1	(19)	0.56	1	U	R.DRVAEESELGLTAGYGINILGMEPLR.T

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
6153	940.4812	2818.4219	2818.4171	1.71	1	(68)	6.2e-006	1	U	R.DRVAEESELGLTAGYGINILGMEPLR.T 6149 6150 6151 6152 6154 6
6246	964.0217	2889.0433	2889.0508	-2.62	1	94	4e-010	1	U	R.RLLCNGDNDCGDYSDENDCDDDPR.T 6244
6248	1446.0300	2890.0454	2889.0508	344	1	(18)	0.015	1	U	R.RLLCNGDNDCGDYSDENDCDDDPR.T
6903	1059.0731	3174.1976	3174.2019	-1.38	0	106	2.6e-011	1	U	R.QGCEPTQECEEIQENCGNDFQCETGR.C 6900 6901 6902 6905 6911
6904	1588.1061	3174.1977	3174.2019	-1.34	0	(64)	4.1e-007	1	U	R.QGCEPTQECEEIQENCGNDFQCETGR.C 6910
7753	1194.2306	3579.6699	3579.6722	-0.65	0	76	5.6e-007	1	U	K.GEYFGFLETYGTHYSTSGSLGGQYEIVYVLDK.A 7736 7737 7738 7740
7762	895.9257	3579.6739	3579.6722	0.47	0	(65)	7.3e-006	1	U	K.GEYFGFLETYGTHYSTSGSLGGQYEIVYVLDK.A 7732 7733 7734 7735
7822	717.1400	3580.6636	3579.6722	277	0	(32)	0.013	1	U	K.GEYFGFLETYGTHYSTSGSLGGQYEIVYVLDK.A
7823	1791.3400	3580.6654	3579.6722	277	0	(56)	5.4e-005	1	U	K.GEYFGFLETYGTHYSTSGSLGGQYEIVYVLDK.A
8323	1000.2270	3996.8787	3996.8768	0.49	1	21	0.18	1	U	K.GEYFGFLETYGTHYSTSGSLGGQYEIVYVLDKASMK.E
8487	1118.2920	4469.1390	4469.1267	2.74	1	37	0.0039	1	U	K.ALPTSYEKGEYFGFLETYGTHYSTSGSLGGQYEIVYVLDK.A 8488 8489

2. K1C10 MOUSE   Mass: 57906   Score: 276   Matches: 12(6)   Sequences: 10(5)   emPAI: 0.47
Keratin, type I cytoskeletal 10 OS=Mus musculus GN=Krt10 PE=1 SV=3

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
107	404.2010	806.3874	806.3923	-6.08	0	8	6.5	1		R.LAADDFR.L
243	424.2281	846.4417	846.4447	-3.52	0	15	2.1	1	U	K.SEITELR.R
846	498.2618	994.5090	994.5123	-3.37	1	25	0.14	1	U	K.IKEWYEK.H
871	502.2787	1002.5429	1002.5458	-2.87	1	7	14	1	U	K.SEITELRR.T
1239	545.7700	1089.5254	1089.5237	1.62	0	37	0.0072	1		R.VTMQNLNDR.L 1238
1475	583.2942	1164.5738	1164.5775	-3.18	0	43	0.0027	1	U	R.LENEIQTYR.S
1564	601.3110	1200.6074	1200.6098	-2.05	0	54	0.00018	1		R.QSVEADINGLR.R
2346	679.3600	1356.7054	1356.7110	-4.06	1	9	5.9	1		R.QSVEADINGLRR.V
2373	691.3200	1380.6254	1380.6408	-11.15	0	87	4.9e-008	1	U	R.ALEESNYELEGK.I 2374
2388	695.8398	1389.6650	1389.6736	-6.17	0	75	1.1e-006	1	U	K.QSLEASLAETEGR.Y

3. ENOG MOUSE   Mass: 47609   Score: 266   Matches: 6(5)   Sequences: 5(4)   emPAI: 0.40
Gamma-enolase OS=Mus musculus GN=Eno2 PE=1 SV=2

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
765	487.2676	972.5207	972.5240	-3.33	1	1	37	5	U	R.IERAVEEK.A
2412	703.8628	1405.7110	1405.7089	1.48	0	69	7.1e-006	1	U	R.GNPTVEVDLYTAK.G
2774	760.4208	1518.8270	1518.8228	2.77	0	78	7e-007	1	U	K.FGANAILGVSLAVCK.A
3723	902.9757	1803.9369	1803.9366	0.14	0	48	0.00077	1	U	R.AAVPSGASTGIYEALELR.D
5096	785.0597	2352.1574	2352.1519	2.33	0	(44)	0.0017	1	U	R.SGETEDTFIADLVVGLCTGQIK.T
5097	1177.0861	2352.1577	2352.1519	2.44	0	93	2.5e-008	1	U	R.SGETEDTFIADLVVGLCTGQIK.T

4. K2C73 MOUSE   Mass: 59502   Score: 183   Matches: 5(4)   Sequences: 3(2)   emPAI: 0.18
Keratin, type II cytoskeletal 73 OS=Mus musculus GN=Krt73 PE=1 SV=1

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
184	414.2200	826.4254	826.4225	3.57	0	15	0.87	1		K.FASFIDK.V
1988	639.3598	1276.7051	1276.7027	1.94	0	58	7.8e-005	1	U	K.LALDIEIATYR.K 1987
2702	738.3922	1474.7698	1474.7780	-5.52	0	69	7.3e-006	1		R.FLEQQNQVLQTK.W 2703

5. CHTOP MOUSE   Mass: 26568   Score: 161   Matches: 3(2)   Sequences: 3(2)   emPAI: 0.43
Chromatin target of PRMT1 protein OS=Mus musculus GN=Chtop PE=1 SV=2

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
817	493.7893	985.5641	985.5668	-2.84	1	21	0.33	1	U	R.LGKSNIQAR.L
2676	731.8543	1461.6940	1461.6994	-3.73	0	93	2e-008	1	U	R.ASMQQQQQLASAR.N
2973	785.8543	1569.6941	1569.6981	-2.54	0	88	2.1e-008	1	U	K.EQLDNQLDAYMSK.T

6. H2B1B MOUSE   Mass: 13944   Score: 159   Matches: 7(4)   Sequences: 4(2)   emPAI: 1.98
Histone H2B type 1-B OS=Mus musculus GN=Hist1h2bb PE=1 SV=3

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
705	477.3047	952.5949	952.5957	-0.86	0	7	1.8	2		R.LLLPGELAK.H 704
1400	569.2761	1136.5377	1136.5390	-1.11	0	29	0.051	1	U	K.ESYSVYVYK.V
1484	584.8004	1167.5863	1167.5884	-1.79	0	28	0.076	1		K.QVHPDTGISSK.A
3537	872.4107	1742.8069	1742.8120	-2.95	0	74	1e-006	1		K.AMGIMNSFVNDIFER.I
3616	880.4100	1758.8054	1758.8069	-0.83	0	(38)	0.004	1		K.AMGIMNSFVNDIFER.I
3653	888.4097	1774.8048	1774.8018	1.64	0	(54)	9.8e-005	1		K.AMGIMNSFVNDIFER.I

Proteins matching the same set of peptides:
H2B1C MOUSE   Mass: 13898   Score: 159   Matches: 7(4)   Sequences: 4(2)
Histone H2B type 1-C/E/G OS=Mus musculus GN=Hist1h2bc PE=1 SV=3
H2B1F MOUSE   Mass: 13928   Score: 159   Matches: 7(4)   Sequences: 4(2)
Histone H2B type 1-F/J/L OS=Mus musculus GN=Hist1h2bf PE=1 SV=2
H2B1H MOUSE   Mass: 13912   Score: 159   Matches: 7(4)   Sequences: 4(2)
Histone H2B type 1-H OS=Mus musculus GN=Hist1h2bh PE=1 SV=3
H2B1K MOUSE   Mass: 13912   Score: 159   Matches: 7(4)   Sequences: 4(2)
Histone H2B type 1-K OS=Mus musculus GN=Hist1h2bk PE=1 SV=3
H2B1M MOUSE   Mass: 13928   Score: 159   Matches: 7(4)   Sequences: 4(2)
Histone H2B type 1-M OS=Mus musculus GN=Hist1h2bm PE=1 SV=2
H2B1P MOUSE   Mass: 13984   Score: 159   Matches: 7(4)   Sequences: 4(2)
Histone H2B type 1-P OS=Mus musculus GN=Hist1h2bp PE=1 SV=3
H2B2B MOUSE   Mass: 13912   Score: 159   Matches: 7(4)   Sequences: 4(2)
Histone H2B type 2-B OS=Mus musculus GN=Hist2h2bb PE=1 SV=3

7. K2C1 MOUSE   Mass: 66079   Score: 157   Matches: 4(3)   Sequences: 3(2)   emPAI: 0.16

Keratin, type II cytoskeletal 1 OS=Mus musculus GN=Krt1 PE=1 SV=4

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
184	414.2200	826.4254	826.4225	3.57	0	15	1.8	2		K.FASFIDK.V
1949	633.3184	1264.6223	1264.6299	-6.05	0	58	7.1e-005	1	U	R.TNAENEFVTIK.K
2702	738.3922	1474.7698	1474.7780	-5.52	0	69	7.3e-006	1		R.FLEQQNQVLQTK.W 2703

---

8. **H2B2E MOUSE**   **Mass:** 13985   **Score:** 153   **Matches:** 7(3)   **Sequences:** 4(1)   **emPAI:** 1.98

Histone H2B type 2-E OS=Mus musculus GN=Hist2h2be PE=1 SV=3

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
705	477.3047	952.5949	952.5957	-0.86	0	7	1.8	2		R.LLLPGELAK.H 704
1440	576.2840	1150.5535	1150.5546	-0.94	0	23	0.2	1	U	K.ESYSIYVYK.V
1484	584.8004	1167.5863	1167.5884	-1.79	0	28	0.076	1		K.QVHPDTGISSK.A
3537	872.4107	1742.8069	1742.8120	-2.95	0	74	1e-006	1		K.AMGIMNSFVNDIFER.I
3616	880.4100	1758.8054	1758.8069	-0.83	0	(38)	0.004	1		K.AMGIMNSFVNDIFER.I
3653	888.4097	1774.8048	1774.8018	1.64	0	(54)	9.8e-005	1		K.AMGIMNSFVNDIFER.I

     Proteins matching the same set of peptides:
     H2B3A MOUSE   **Mass:** 13986   **Score:** 153   **Matches:** 7(3)   **Sequences:** 4(1)
     Histone H2B type 3-A OS=Mus musculus GN=Hist3h2ba PE=1 SV=3
     H2B3B MOUSE   **Mass:** 13900   **Score:** 153   **Matches:** 7(3)   **Sequences:** 4(1)
     Histone H2B type 3-B OS=Mus musculus GN=Hist3h2bb PE=1 SV=3

---

9. **TBA1A MOUSE**   **Mass:** 50788   **Score:** 131   **Matches:** 5(4)   **Sequences:** 3(3)   **emPAI:** 0.29

Tubulin alpha-1A chain OS=Mus musculus GN=Tuba1a PE=1 SV=1

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
1217	543.3137	1084.6128	1084.6128	0.00	0	51	0.00028	1	U	K.EIIDLVLDR.I 1216
3457	851.4581	1700.9016	1700.8985	1.81	0	30	0.05	1	U	R.AVFVDLEPTVIDEVR.T
5155	803.7415	2408.2026	2408.2012	0.56	0	(23)	0.27	1	U	R.FDGALNVDLTEFQTNLVPYPR.I
5156	1205.1141	2408.2136	2408.2012	5.12	0	67	1e-005	1	U	R.FDGALNVDLTEFQTNLVPYPR.I

     Proteins matching the same set of peptides:
     TBA1B MOUSE   **Mass:** 50804   **Score:** 131   **Matches:** 5(4)   **Sequences:** 3(3)
     Tubulin alpha-1B chain OS=Mus musculus GN=Tuba1b PE=1 SV=2
     TBA1C MOUSE   **Mass:** 50562   **Score:** 131   **Matches:** 5(4)   **Sequences:** 3(3)
     Tubulin alpha-1C chain OS=Mus musculus GN=Tuba1c PE=1 SV=1

---

10. **K2C5 MOUSE**   **Mass:** 61957   **Score:** 105   **Matches:** 7(2)   **Sequences:** 6(1)   **emPAI:** 0.17

Keratin, type II cytoskeletal 5 OS=Mus musculus GN=Krt5 PE=1 SV=1

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
184	414.2200	826.4254	826.4225	3.57	0	15	0.87	1		K.FASFIDK.V
601	466.7500	931.4854	931.4974	-12.87	1	26	0.14	1	U	R.SEIDNVKK.Q
1174	538.3144	1074.6143	1072.5876	1889	1	5	16	1	U	R.LRSEIDNVK.K
1554	597.7888	1193.5630	1193.5676	-3.86	0	11	2.8	1	U	K.YEELQQTAGR.H
1566	601.3500	1200.6854	1200.6826	2.37	2	20	0.5	1	U	R.LRSEIDNVKK.Q
2188	651.8608	1301.7070	1301.7078	-0.67	0	68	9.8e-006	1	U	R.SLDLDSIIAEVK.A 2187

---

11. **ACTB MOUSE**   **Mass:** 42052   **Score:** 99   **Matches:** 7(3)   **Sequences:** 7(3)   **emPAI:** 0.46

Actin, cytoplasmic 1 OS=Mus musculus GN=Actb PE=1 SV=1

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
772	488.7300	975.4454	975.4410	4.55	0	17	0.47	1		K.AGFAGDDAPR.A
857	499.7476	997.4807	997.4790	1.68	0	23	0.14	1		R.DLTDYLMK.I
1389	566.7643	1131.5141	1131.5197	-4.89	0	25	0.054	1	U	R.GYSFTTTAER.E
1463	581.3112	1160.6079	1160.6111	-2.73	0	32	0.042	1		K.EITALAPSTMK.I
1560	599.7600	1197.5054	1197.5150	-7.94	0	2	5.9	1		K.DSYVGDEAQSK.R
4861	1108.0443	2214.0740	2214.0627	5.12	0	68	7.5e-006	1	U	K.DLYANTVLSGGTTMYPGIADR.M
7003	1077.8234	3230.4485	3230.4545	-1.86	0	11	1.4	1	U	R.CPEALFQPSFLGMESCGIHETTFNSIMK.C

     Proteins matching the same set of peptides:
     ACTG MOUSE   **Mass:** 42108   **Score:** 99   **Matches:** 7(3)   **Sequences:** 7(3)
     Actin, cytoplasmic 2 OS=Mus musculus GN=Actg1 PE=1 SV=1

---

12. **K1C27 MOUSE**   **Mass:** 49645   **Score:** 86   **Matches:** 4(3)   **Sequences:** 2(2)   **emPAI:** 0.21

Keratin, type I cytoskeletal 27 OS=Mus musculus GN=Krt27 PE=1 SV=1

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
1239	545.7700	1089.5254	1089.5237	1.62	0	37	0.0072	1		K.VTMQNLNDR.L 1238
4523	699.3559	2095.0460	2093.0779	940	0	(39)	0.0057	1	U	R.TDLEVQLETLSEELAYLK.K
4524	1048.5319	2095.0493	2093.0779	942	0	44	0.0021	1	U	R.TDLEVQLETLSEELAYLK.K

---

13. **H2A1F MOUSE**   **Mass:** 14153   **Score:** 83   **Matches:** 5(3)   **Sequences:** 3(1)   **emPAI:** 0.91

Histone H2A type 1-F OS=Mus musculus GN=Hist1h2af PE=1 SV=3

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
269	425.7655	849.5165	849.5184	-2.25	0	23	0.078	1	U	R.HLQLAIR.N
658	472.7678	943.5210	943.5240	-3.14	0	12	3.2	1	U	R.AGLQFPVGR.V
3957	644.3933	1930.1580	1930.1615	-1.85	0	(40)	0.00034	1	U	R.VTIAQGGVLPNIQAVLLPK.K
3959	966.0877	1930.1608	1930.1615	-0.35	0	57	6.5e-006	1	U	R.VTIAQGGVLPNIQAVLLPK.K 3958

     Proteins matching the same set of peptides:

H2A1H MOUSE   **Mass:** 13942   **Score:** 83   **Matches:** 5(3) **Sequences:** 3(1)
Histone H2A type 1-H OS=Mus musculus GN=Hist1h2ah PE=1 SV=3
H2A1K MOUSE   **Mass:** 14141   **Score:** 83   **Matches:** 5(3) **Sequences:** 3(1)
Histone H2A type 1-K OS=Mus musculus GN=Hist1h2ak PE=1 SV=3
H2A1 MOUSE   **Mass:** 14127   **Score:** 83   **Matches:** 5(3) **Sequences:** 3(1)
Histone H2A type 1 OS=Mus musculus GN=Hist1h2ab PE=1 SV=3
H2A2A MOUSE   **Mass:** 14087   **Score:** 83   **Matches:** 5(3) **Sequences:** 3(1)
Histone H2A type 2-A OS=Mus musculus GN=Hist2h2aa1 PE=1 SV=3
H2A2C MOUSE   **Mass:** 13980   **Score:** 83   **Matches:** 5(3) **Sequences:** 3(1)
Histone H2A type 2-C OS=Mus musculus GN=Hist2h2ac PE=1 SV=3
H2A3 MOUSE   **Mass:** 14113   **Score:** 83   **Matches:** 5(3) **Sequences:** 3(1)
Histone H2A type 3 OS=Mus musculus GN=Hist3h2a PE=1 SV=3
H2AJ MOUSE   **Mass:** 14037   **Score:** 83   **Matches:** 5(3) **Sequences:** 3(1)
Histone H2A.J OS=Mus musculus GN=H2afj PE=1 SV=1

---

14. H4 MOUSE   **Mass:** 11360   **Score:** 66   **Matches:** 4(2) **Sequences:** 3(2) **emPAI:** 1.22
Histone H4 OS=Mus musculus GN=Hist1h4a PE=1 SV=2

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
831	495.2917	988.5689	988.5706	-1.66	0	31	0.039	1	U	K.VFLENVIR.D 830
1395	567.7700	1133.5254	1133.5353	-8.69	0	23	0.14	1	U	R.DAVTYTEHAK.R
1515	590.8130	1179.6115	1179.6135	-1.69	0	48	0.00092	1	U	R.ISGLIYEETR.G

---

15. LDHB MOUSE   **Mass:** 36834   **Score:** 60   **Matches:** 1(1) **Sequences:** 1(1) **emPAI:** 0.09
L-lactate dehydrogenase B chain OS=Mus musculus GN=Ldhb PE=1 SV=2

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
3309	815.4336	1628.8526	1628.8509	1.05	0	60	5e-005	1	U	K.SLADELALVDVLEDK.L

---

16. HNRPK MOUSE   **Mass:** 51230   **Score:** 58   **Matches:** 1(1) **Sequences:** 1(1) **emPAI:** 0.06
Heterogeneous nuclear ribonucleoprotein K OS=Mus musculus GN=Hnrnpk PE=1 SV=1

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
3479	857.9900	1713.9654	1713.9764	-6.38	0	58	4.6e-005	1	U	R.ILSISADIETIGEILK.K

---

17. K1C13 MOUSE   **Mass:** 48066   **Score:** 55   **Matches:** 5(1) **Sequences:** 4(1) **emPAI:** 0.14
Keratin, type I cytoskeletal 13 OS=Mus musculus GN=Krt13 PE=1 SV=2

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
107	404.2010	806.3874	806.3923	-6.08	0	8	6.5	1		R.LAADDFR.L
110	405.2221	808.4296	808.4330	-4.24	0	23	0.21	1		R.LASYLDK.V 111
1564	601.3110	1200.6074	1200.6098	-2.05	0	54	0.00018	1		R.QSVEADINGLR.R
2346	679.3600	1356.7054	1356.7110	-4.06	1	9	5.9	1		R.QSVEADINGLRR.V

---

18. K2C79 MOUSE   **Mass:** 57802   **Score:** 54   **Matches:** 2(1) **Sequences:** 2(1) **emPAI:** 0.12
Keratin, type II cytoskeletal 79 OS=Mus musculus GN=Krt79 PE=1 SV=2

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
184	414.2200	826.4254	826.4225	3.57	0	15	0.87	1		K.FASFIDK.V
2285	665.3661	1328.7177	1328.7187	-0.76	0	58	8.8e-005	1	U	R.NLDLDSIIAEVK.A

---

19. KPYM MOUSE   **Mass:** 58378   **Score:** 51   **Matches:** 1(1) **Sequences:** 1(1) **emPAI:** 0.06
Pyruvate kinase PKM OS=Mus musculus GN=Pkm PE=1 SV=4

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
2677	731.9111	1461.8077	1461.8079	-0.15	0	51	0.00035	1	U	K.IYVDDGLISLQVK.E

---

20. PPIA MOUSE   **Mass:** 18131   **Score:** 51   **Matches:** 2(1) **Sequences:** 2(1) **emPAI:** 0.41
Peptidyl-prolyl cis-trans isomerase A OS=Mus musculus GN=Ppia PE=1 SV=2

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
1082	528.2700	1054.5254	1054.5335	-7.64	0	48	0.00064	1	U	R.VSFELFADK.V
2017	639.7900	1277.5654	1277.5744	-6.98	0	18	0.25	1		K.EGMNIVEAMER.F

---

21. K22E MOUSE   **Mass:** 71336   **Score:** 47   **Matches:** 3(2) **Sequences:** 2(1) **emPAI:** 0.09
Keratin, type II cytoskeletal 2 epidermal OS=Mus musculus GN=Krt2 PE=1 SV=1

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
184	414.2200	826.4254	826.4225	3.57	0	15	0.87	1		K.FASFIDK.V
1416	571.2616	1140.5087	1140.5121	-2.97	0	36	0.0049	1	U	R.DYQELMNTK.L 1417

---

22. EF1A1 MOUSE   **Mass:** 50424   **Score:** 46   **Matches:** 3(1) **Sequences:** 2(1) **emPAI:** 0.14
Elongation factor 1-alpha 1 OS=Mus musculus GN=Eef1a1 PE=1 SV=3

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
325	435.7700	869.5254	869.5334	-9.20	0	5	6.9	1	U	K.QLIVGVNK.M
6292	970.8200	2909.4382	2908.4310	346	0	40	0.0043	1	U	K.NMITGTSQADCAVLIVAAGVGEFEAGISK.N
6367	976.1500	2925.4282	2924.4260	343	0	(21)	0.34	1	U	K.NMITGTSQADCAVLIVAAGVGEFEAGISK.N

Proteins matching the same set of peptides:
EF1A2 MOUSE   **Mass:** 50764   **Score:** 46   **Matches:** 3(1) **Sequences:** 2(1)
Elongation factor 1-alpha 2 OS=Mus musculus GN=Eef1a2 PE=1 SV=1

---

23. RS2 MOUSE   **Mass:** 31497   **Score:** 43   **Matches:** 1(1) **Sequences:** 1(1) **emPAI:** 0.11

156

```
40S ribosomal protein S2 OS=Mus musculus GN=Rps2 PE=1 SV=3
Query  Observed   Mr(expt)   Mr(calc)   ppm  Miss Score Expect Rank Unique  Peptide
2970   784.9160  1567.8174  1567.8134  2.58   0    43  0.0022   1     U    K.ESEIIDFFLGASLK.D
```

24. HSP7C MOUSE   **Mass:** 71055   **Score:** 40   **Matches:** 1(1) **Sequences:** 1(1) **emPAI:** 0.05
```
Heat shock cognate 71 kDa protein OS=Mus musculus GN=Hspa8 PE=1 SV=1
Query  Observed   Mr(expt)   Mr(calc)   ppm  Miss Score Expect Rank Unique  Peptide
6462   999.8273  2996.4599  2996.4502  3.24   0    40  0.004    1     U    R.TLSSSTQASIEIDSLYEGIDFYTSITR.A
```

25. IF4A1 MOUSE   **Mass:** 46353   **Score:** 39   **Matches:** 1(1) **Sequences:** 1(1) **emPAI:** 0.07
```
Eukaryotic initiation factor 4A-I OS=Mus musculus GN=Eif4a1 PE=1 SV=1
Query  Observed   Mr(expt)   Mr(calc)   ppm  Miss Score Expect Rank Unique  Peptide
1327   557.8442  1113.6739  1113.6758  -1.68  0    39  0.0027   1     U    R.VLITTDLLAR.G
```

```
     Proteins matching the same set of peptides:
```
IF4A2 MOUSE   **Mass:** 46601   **Score:** 39   **Matches:** 1(1) **Sequences:** 1(1)
```
Eukaryotic initiation factor 4A-II OS=Mus musculus GN=Eif4a2 PE=1 SV=2
```

26. ACTA MOUSE   **Mass:** 42381   **Score:** 39   **Matches:** 5(1) **Sequences:** 5(1) **emPAI:** 0.25
```
Actin, aortic smooth muscle OS=Mus musculus GN=Acta2 PE=1 SV=1
Query  Observed   Mr(expt)   Mr(calc)   ppm  Miss Score Expect Rank Unique  Peptide
 772   488.7300   975.4454   975.4410  4.55   0    17   0.47    1          K.AGFAGDDAPR.A
 857   499.7476   997.4807   997.4790  1.68   0    23   0.14    1          R.DLTDYLMK.I
1463   581.3112  1160.6079  1160.6111  -2.73  0    32   0.042   1          K.EITALAPSTMK.I
1560   599.7600  1197.5054  1197.5150  -7.94  0     2   5.9     1          K.DSYVGDEAQSK.R
4051   654.3091  1959.9054  1959.9036  0.90   0    12   1.7     1     U    K.YPIEHGIITNWDDMEK.I
```

```
     Proteins matching the same set of peptides:
```
ACTC MOUSE   **Mass:** 42334   **Score:** 39   **Matches:** 5(1) **Sequences:** 5(1)
```
Actin, alpha cardiac muscle 1 OS=Mus musculus GN=Actc1 PE=1 SV=1
```
ACTH MOUSE   **Mass:** 42249   **Score:** 39   **Matches:** 5(1) **Sequences:** 5(1)
```
Actin, gamma-enteric smooth muscle OS=Mus musculus GN=Actg2 PE=1 SV=1
```
ACTS MOUSE   **Mass:** 42366   **Score:** 39   **Matches:** 5(1) **Sequences:** 5(1)
```
Actin, alpha skeletal muscle OS=Mus musculus GN=Acta1 PE=1 SV=1
```

27. HB22 MOUSE   **Mass:** 30315   **Score:** 38   **Matches:** 3(1) **Sequences:** 2(1) **emPAI:** 0.23
```
H-2 class II histocompatibility antigen, E-D beta chain OS=Mus musculus PE=3 SV=1
Query  Observed   Mr(expt)   Mr(calc)   ppm  Miss Score Expect Rank Unique  Peptide
1748   617.3373  1232.6601  1232.6653  -4.19  0    31   0.045   1     U    R.VEPTVTVYPTK.T
2386   695.3877  1388.7609  1388.7664  -3.95  1    21   0.38    1     U    R.RVEPTVTVYPTK.T 2387
```

```
     Proteins matching the same set of peptides:
```
HB2I MOUSE   **Mass:** 30554   **Score:** 38   **Matches:** 3(1) **Sequences:** 2(1)
```
H-2 class II histocompatibility antigen, I-A beta chain OS=Mus musculus GN=H2-Eb1 PE=1 SV=1
```
HB2J MOUSE   **Mass:** 30407   **Score:** 38   **Matches:** 3(1) **Sequences:** 2(1)
```
H-2 class II histocompatibility antigen, I-E beta chain OS=Mus musculus GN=H2-Eb1 PE=1 SV=1
```
HB21 MOUSE   **Mass:** 30489   **Score:** 38   **Matches:** 3(1) **Sequences:** 2(1)
```
H-2 class II histocompatibility antigen, E-B beta chain OS=Mus musculus GN=H2-Eb1 PE=1 SV=1
```
HB23 MOUSE   **Mass:** 26858   **Score:** 38   **Matches:** 3(1) **Sequences:** 2(1)
```
H-2 class II histocompatibility antigen, E-S beta chain (Fragment) OS=Mus musculus GN=H2-Eb1 PE=1 SV=1
```
HB24 MOUSE   **Mass:** 30527   **Score:** 38   **Matches:** 3(1) **Sequences:** 2(1)
```
H-2 class II histocompatibility antigen, E-Q beta chain OS=Mus musculus PE=1 SV=1
```

28. AAAT MOUSE   **Mass:** 59243   **Score:** 36   **Matches:** 7(1) **Sequences:** 2(1) **emPAI:** 0.06
```
Neutral amino acid transporter B(0) OS=Mus musculus GN=Slc1a5 PE=1 SV=2
Query  Observed   Mr(expt)   Mr(calc)   ppm  Miss Score Expect Rank Unique  Peptide
 982   515.2452  1028.4759  1028.4815  -5.45  0    27   0.04    1     U    R.SFATSYEPK.D 979 980 981 983
3603   878.3682  1754.7218  1752.7268  1138   2     2   2.6     3     U    R.EDQSAKAGGCCGSRDR.V 3604
```

29. ANXA2 MOUSE   **Mass:** 38937   **Score:** 35   **Matches:** 1(1) **Sequences:** 1(1) **emPAI:** 0.08
```
Annexin A2 OS=Mus musculus GN=Anxa2 PE=1 SV=2
Query  Observed   Mr(expt)   Mr(calc)   ppm  Miss Score Expect Rank Unique  Peptide
2931   771.9300  1541.8454  1541.8413  2.67   0    35   0.013   1     U    K.GVDEVTIVNILTNR.S
```

30. H12 MOUSE   **Mass:** 21254   **Score:** 35   **Matches:** 1(1) **Sequences:** 1(1) **emPAI:** 0.16
```
Histone H1.2 OS=Mus musculus GN=Hist1h1c PE=1 SV=2
Query  Observed   Mr(expt)   Mr(calc)   ppm  Miss Score Expect Rank Unique  Peptide
1561   599.8359  1197.6573  1197.6605  -2.70  0    35   0.014   1     U    K.ASGPPVSELITK.A
```

```
     Proteins matching the same set of peptides:
```
H13 MOUSE   **Mass:** 22086   **Score:** 35   **Matches:** 1(1) **Sequences:** 1(1)
```
Histone H1.3 OS=Mus musculus GN=Hist1h1d PE=1 SV=2
```

31. RL10A MOUSE   **Mass:** 25072   **Score:** 33   **Matches:** 1(1) **Sequences:** 1(1) **emPAI:** 0.13
```
60S ribosomal protein L10a OS=Mus musculus GN=Rpl10a PE=1 SV=3
Query  Observed   Mr(expt)   Mr(calc)   ppm  Miss Score Expect Rank Unique  Peptide
```

2527	710.4103	1418.8060	1418.8021	2.76	0	33	0.014	1		U	K.FLETVELQISLK.N

---

32. EWS_MOUSE **Mass:** 68705   **Score:** 32   **Matches:** 4(0)   **Sequences:** 4(0)   **emPAI:** 0.15
RNA-binding protein EWS OS=Mus musculus GN=Ewsr1 PE=1 SV=2

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
2630	725.8400	1449.6654	1449.6624	2.13	0	21	0.21	1	U	K.GDATVSYEDPPTAK.A
3423	562.2697	1683.7873	1683.7893	-1.18	1	17	0.57	1	U	K.AAVEWFDGKDFQGSK.L
3953	642.6629	1924.9669	1924.9683	-0.75	2	22	0.34	1	U	K.AAVEWFDGKDFQGSKLK.V
7594	876.6895	3502.7290	3502.7290	0.00	2	0	31	1	U	R.TGQPMIHIYLDKETGKPKGDATVSYEDPPTAK.A

---

33. K22O_MOUSE **Mass:** 63319   **Score:** 30   **Matches:** 2(0)   **Sequences:** 2(0)   **emPAI:** 0.11
Keratin, type II cytoskeletal 2 oral OS=Mus musculus GN=Krt76 PE=1 SV=1

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
184	414.2200	826.4254	826.4225	3.57	0	15	0.87	1		K.FASFIDK.V
1158	533.2624	1064.5102	1064.5138	-3.40	0	28	0.057	1	U	K.AQYEDIAQK.S

---

34. K1C42_MOUSE **Mass:** 50444   **Score:** 30   **Matches:** 2(1)   **Sequences:** 2(1)   **emPAI:** 0.07
Keratin, type I cytoskeletal 42 OS=Mus musculus GN=Krt42 PE=1 SV=1

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
107	404.2010	806.3874	806.3923	-6.08	0	8	6.5	1		R.LAADDFR.T
2161	651.3319	1300.6492	1300.6510	-1.40	0	30	0.049	1	U	R.ALEEANADLEVK.I

---

35. HYEP_MOUSE **Mass:** 52714   **Score:** 25   **Matches:** 2(0)   **Sequences:** 1(0)   **emPAI:** 0.06
Epoxide hydrolase 1 OS=Mus musculus GN=Ephx1 PE=1 SV=2

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
711	478.7914	955.5682	955.5814	-13.84	1	20	0.27	1	U	K.LLAQDIRK.F 712

---

36. HS90A_MOUSE **Mass:** 85134   **Score:** 22   **Matches:** 1(0)   **Sequences:** 1(0)   **emPAI:** 0.04
Heat shock protein HSP 90-alpha OS=Mus musculus GN=Hsp90aa1 PE=1 SV=4

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
1818	621.8541	1241.6937	1241.6979	-3.37	0	22	0.24	1	U	K.ADLINNLGTIAK.S

Proteins matching the same set of peptides:
HS90B_MOUSE **Mass:** 83571   **Score:** 22   **Matches:** 1(0)   **Sequences:** 1(0)
Heat shock protein HSP 90-beta OS=Mus musculus GN=Hsp90ab1 PE=1 SV=3

---

37. TBB2A_MOUSE **Mass:** 50274   **Score:** 20   **Matches:** 1(0)   **Sequences:** 1(0)   **emPAI:** 0.07
Tubulin beta-2A chain OS=Mus musculus GN=Tubb2a PE=1 SV=1

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
3231	808.4223	1614.8301	1614.8287	0.84	0	20	0.48	1	U	R.AILVDLEPGTMDSVR.S

Proteins matching the same set of peptides:
TBB2B_MOUSE **Mass:** 50377   **Score:** 20   **Matches:** 1(0)   **Sequences:** 1(0)
Tubulin beta-2B chain OS=Mus musculus GN=Tubb2b PE=1 SV=1
TBB3_MOUSE **Mass:** 50842   **Score:** 20   **Matches:** 1(0)   **Sequences:** 1(0)
Tubulin beta-3 chain OS=Mus musculus GN=Tubb3 PE=1 SV=1
TBB5_MOUSE **Mass:** 50095   **Score:** 20   **Matches:** 1(0)   **Sequences:** 1(0)
Tubulin beta-5 chain OS=Mus musculus GN=Tubb5 PE=1 SV=1

---

38. RL11_MOUSE **Mass:** 20468   **Score:** 20   **Matches:** 1(0)   **Sequences:** 1(0)   **emPAI:** 0.16
60S ribosomal protein L11 OS=Mus musculus GN=Rpl11 PE=1 SV=4

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
5172	809.0600	2424.1582	2422.1554	827	0	20	0.43	1	U	K.NNFSDTGNFGFGIQEHIDLGIK.Y

---

39. TEX2_MOUSE **Mass:** 125718   **Score:** 20   **Matches:** 1(0)   **Sequences:** 1(0)   **emPAI:** 0.03
Testis-expressed sequence 2 protein OS=Mus musculus GN=Tex2 PE=1 SV=2

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
829	495.2548	988.4951	988.5077	-12.68	0	20	0.49	1	U	K.SLSTEVEPK.E

---

40. ZN488_MOUSE **Mass:** 37628   **Score:** 18   **Matches:** 1(0)   **Sequences:** 1(0)   **emPAI:** 0.09
Zinc finger protein 488 OS=Mus musculus GN=Znf488 PE=2 SV=2

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
2259	442.2200	1323.6382	1321.6548	1501	1	18	0.61	1	U	K.KMSLVDSDTAAGK.G

---

41. CALM_MOUSE **Score:** 16   **Matches:** 1(0)   **Sequences:** 1(0)   **emPAI:** 0.20
Calmodulin OS=Mus musculus GN=Calm1 PE=1 SV=2

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
166	411.2100	820.4054	820.4113	-7.11	0	16	1.3	1	U	K.ELGTVMR.S

Proteins matching the same set of peptides:
TNNC2_MOUSE **Score:** 16   **Matches:** 1(0)   **Sequences:** 1(0)

---

42. TMPSD_MOUSE **Mass:** 60851   **Score:** 16   **Matches:** 1(0)   **Sequences:** 1(0)   **emPAI:** 0.05

Transmembrane protease serine 13 OS=Mus musculus GN=Tmprss13 PE=2 SV=2

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
511	453.7537	905.4929	905.4971	-4.61	0	16	1.6	1	U	K.NKPGVYTK.V

---

43. G3P_MOUSE   **Mass:** 36072   **Score:** 15   **Matches:** 1(0)   **Sequences:** 1(0)   **emPAI:** 0.09

Glyceraldehyde-3-phosphate dehydrogenase OS=Mus musculus GN=Gapdh PE=1 SV=2

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
5595	865.7924	2594.3555	2594.3527	1.08	0	15	1.1	1	U	K.VIHDNFGIVEGLMTTVHAITATQK.T

---

44. RS27A_MOUSE   **Mass:** 18282   **Score:** 15   **Matches:** 1(0)   **Sequences:** 1(0)   **emPAI:** 0.18

Ubiquitin-40S ribosomal protein S27a OS=Mus musculus GN=Rps27a PE=1 SV=2

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
3681	894.4650	1786.9155	1786.9200	-2.55	0	15	1.7	1	U	K.TITLEVEPSDTIENVK.A

Proteins matching the same set of peptides:
RL40_MOUSE   **Mass:** 15004   **Score:** 15   **Matches:** 1(0)   **Sequences:** 1(0)
Ubiquitin-60S ribosomal protein L40 OS=Mus musculus GN=Uba52 PE=1 SV=2
UBB_MOUSE   **Mass:** 34348   **Score:** 15   **Matches:** 1(0)   **Sequences:** 1(0)
Polyubiquitin-B OS=Mus musculus GN=Ubb PE=2 SV=1
UBC_MOUSE   **Mass:** 82614   **Score:** 15   **Matches:** 1(0)   **Sequences:** 1(0)
Polyubiquitin-C OS=Mus musculus GN=Ubc PE=1 SV=2

---

45. ROA2_MOUSE   **Mass:** 37437   **Score:** 14   **Matches:** 1(0)   **Sequences:** 1(0)   **emPAI:** 0.09

Heterogeneous nuclear ribonucleoproteins A2/B1 OS=Mus musculus GN=Hnrnpa2b1 PE=1 SV=2

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
3711	900.4600	1798.9054	1797.9149	551	0	14	1.9	1	U	K.LFIGGLSFETTEESLR.N

Mascot: http://www.matrixscience.com/

159