Structural biochemistry of the pentraxins

Noone, D.P.

1 Immunity, the complement system and pentraxins

Dylan P. Noone[1] and Thomas H. Sharp[1,*]

[1]Department of Cell and Chemical Biology, Leiden University Medical Center, 2300 RC Leiden, The Netherlands

*To whom correspondence should be addressed: t.sharp@lumc.nl

1.1 Immunity

The immune system has many functions, ranging from protection against pathogens (1, 2), to the repair and maintenance of tissues (3). Additionally, the human immune system can be divided into the innate and adaptive immune system (1) which both respond to motifs on dangerous objects to aid in their removal or destruction (4). The adaptive immune system learns how to recognise specific epitopes via processes such as somatic hypermutation (5) and forms the basis of immunological memory, via effector cells and proteins such as B-cells and antibodies. On the other hand, the innate immune system is conventionally thought not to be able to learn and takes the form of anatomical barriers such as the skin, the inflammatory response and pattern recognition receptors that detect assemblies associated with danger (4). However, the boundaries between innate and adaptive immunity may be more porous than previously thought, with recent data showing a limited ability of the innate immune system to form an immunological memory of events, in a process termed trained immunity (6).

1.2 The complement system

The complement system is classically thought of as a fluid phase based arm of the innate immune system that is activated on dangerous membranes (7). However, it can also receive inputs from the adaptive immune system, for example via activation by hexameric or pseudo-hexameric antibody platforms (8, 9), and can in turn send outputs that affect adaptive immunity, such as C3-mediated enhancement of the B-cell response (10-12). As such, in reality the complement system sits at the interface between the innate and adaptive wings of immunity.

The complement system consists of three pathways; the classical, lectin and alternative(7). The classical cascade can be activated by C1 binding to ligands such as antibodies (9) or pentraxins (13), whereas the lectin pathway recognizes oligosaccharides via mannose binding lectin (MBL), collectin-11 or the ficolins (1, 2 and 3) (14) (**Fig. 1A**). Interestingly, pentraxins can bind to the stalk domain of MBL and cause activation of the classical cascade, enhancing the lectin pathway via inter-cascade cross talk (15). A third form of initiation, the alternative pathway, is characterized by a low tick-over mechanism whereby C3 is cleaved spontaneously by a water molecule (**Fig. 1A**). This is usually inhibited on host cells by a multitude of regulatory proteins, such as factor H and decay accelerating factor (DAF), to prevent aberrant complement activation (7). The three initiation pathways converge at the cleavage of C3, albeit with different C3 convertases (**Fig. 1A**), and allow the complement system to be activated in a controlled manner in a wide range of situations.

The classical initiating C1 complex is produced locally at the sites of inflammation by dendritic cells, macrophages and monocytes(16) forming a hexameric bouquet of flowers structure (8, 9). C1 becomes activated after binding to nanopatterned adaptors such as IgG1 (8), or directly to targets such as phosphatidylserine(17) or LPS(18) via its detection module C1q. This leads to activation of the protease domains C1r and C1s (7, 9), hypothesized to be driven by a

compaction of the protease platform(8) or a sliding mechanism(9), resulting in autoactivation of C1r and C1r mediated activation of C1s.

Active C1s proceeds to cleave C4, exposing a thioester in the proteolytic fragment C4b and results in covalent linkage to the target membrane. With the later C1 mediated proteolysis of C4 bound C2, the classical pathway C3 convertase (C4b2b) is formed (**Fig. 1A**). This can cleave fluid phase C3 resulting in opsonization of the membrane with C3b in a similar manner to C4b (**Fig. 1A**). In addition, deposited opsonins such as antibodies, pentraxins, C1, MBL, C3b or C4b can act as "eat-me" signals, interacting with receptors on professional phagocytes to increase the phagocytosis of these targets (**Fig. 1B**) (19-24). The alternative pathway can also amplify activation by the classical or lectin cascade. The spontaneous tick over mechanism produces covalently membrane bound C3b, which can associate with a cleavage product of Factor B to form the alternative pathway C3 convertase (C3bBb). This in turn cleaves more C3 acting as a positive feedback amplifying loop (7)(**Fig. 1A**). The increasing density of C3b results in the formation of the classical/lectin or alternative C5 convertases, C4b2a3b or C3bBbC3b respectively, and initiation of the terminal pathway (**Fig. 1A**). This ultimately leads to the formation of the membrane attack complex (MAC) lysing the target membrane (7, 25).

Moreover, complement can be fine-tuned via a host of regulatory proteins (26), such as Factor H, allowing tailoring of the immune response to silent clearance of apoptotic cells, characterized by less inflammation and lower rates of lysis(27), or to the complete destruction of the target via the MAC pore (28) and high amounts of inflammation. Complement can directly cause inflammation via the anaphylatoxins produced from the cleavage C3 and C5 (C3a and C5a), which also act as chemo-attractants for leukocytes and cause other physiological responses, such as increased blood vessel permeability(29).

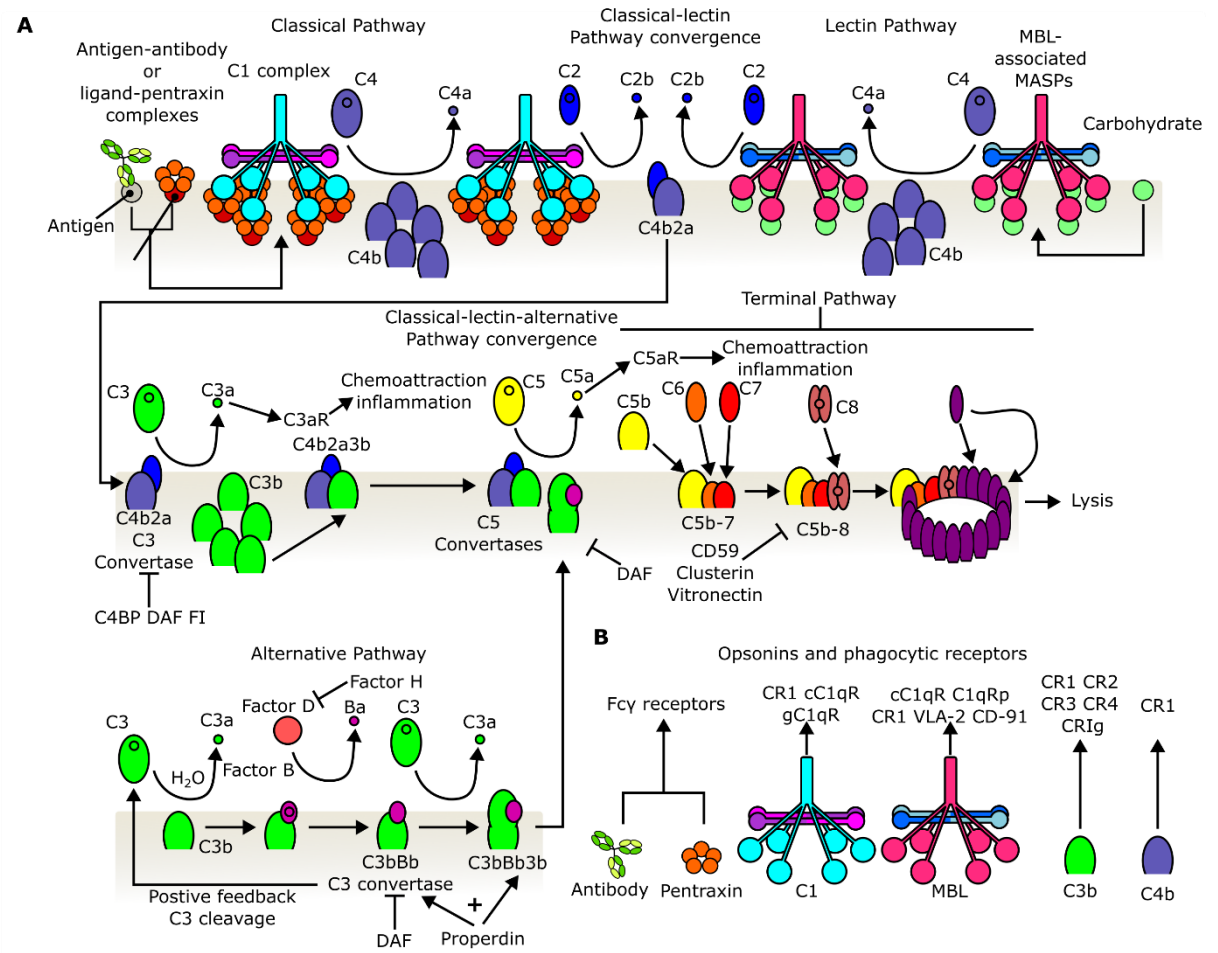

Figure 1: Schematic of the complement system. *(**A**) The classical, lectin and alternative, with convergence at the C5 convertase step and start of the terminal pathway, ending in perforation and lysis of the target. Regulatory molecules, such as Factor H and decay accelerating factor (DAF), are shown as labels as well as the receptors for cleavage products with their downstream effects e.g. C3aR. Only mannose binding lectin (MBL) mediated lectin pathway activation is shown for simplicity. (**B**) The phagocytic receptors for the opsonin molecules resulting from complement activation.*

1.3 Pentraxins

The pentraxins are a family of innate C1 adaptor proteins that are able to activate the classical complement cascade (30-32) and are evolutionarily conserved since arthropods (33). They each contain a pentraxin domain (**Fig. 2A**) which contains the canonical pentraxin motif (**Fig. 2B,** HxCxS/TWxS/T). This motif has a conserved His, Cys and Trp, with all experimentally resolved pentraxins to date showing the Trp buried into the hydrophobic interior of the protein, and the His and Cys residues situated close to one another (**Fig. 2B**). The pentraxin family can be further divided into the long and short pentraxins, the former containing an extra N-terminal region in addition to the pentraxin domain (**Fig. 2A**). All resolved pentraxin domains form a jelly roll structure with two-layered β sheets (34-38) (**Fig. 3A**), which in the short pentraxins are non-covalently associated to form pentamers (**Fig. 3B**).

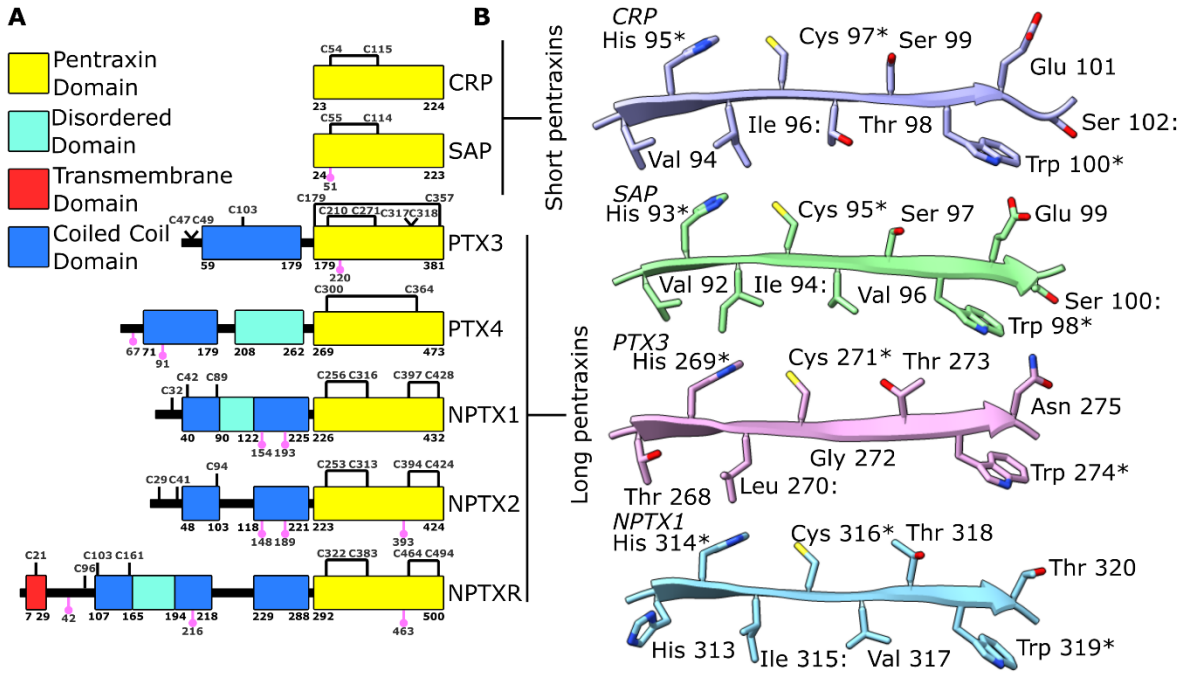

Figure 2: Schematic of the primary sequence of each pentraxin and the structure of the pentraxin motif. *(**A**) Diagrams of the pentraxin's primary sequence showing the predicted pentraxin domain (yellow), predicted coiled coil domains (blue), predicted disordered domains (cyan), and transmembrane domains (red). Residues numbers show the borders of domains in black and the locations of intra-molecular disulfide bonds (closed black line) and inter-molecular disulfide bonds (black lines) are denoted by C then the residue number. Glycosylation is shown as green lines with a circular tip. (**B**) Structures of all resolved pentraxin motifs to date (HxCxS/TWxS/T). Conserved residues contain are marked by an asterisk (*), whilst chemically conserved residues are demarcated by a colon (:). PDB codes of models used CRP:7PKE, SAP:1SAC, PTX3:7ZL1 and NPTX1:6YPE.*

The pentraxin family is split into two subgroups: the short and the long pentraxins (**Fig. 2A**) (39). The short pentraxins are entirely made up of a pentraxin domain, whereas the long pentraxins have an extra N-terminal domain with relatively little known about their structural arrangements. Both subgroups form higher order oligomers and, as will be shown in ***chapter 6***, the only complete long pentraxin resolved to date forms novel quaternary arrangements as compared to the short pentraxins (**Fig. 3B&C**). The pentraxin domains of each family

member have low to intermediate primary sequence similarity (24.2 – 69.8%, **Table 1**) but are probably all highly structurally homologous (0.51 – 1.09 Å RMSD, **Table 1, Fig. 3A**). For instance, the pentraxin domain of NPTX1 has been resolved (38), and despite low sequence homology to the short pentraxins (**Table 1**) it shows a high degree of structural homology (**Fig. 2A, Table 1**).

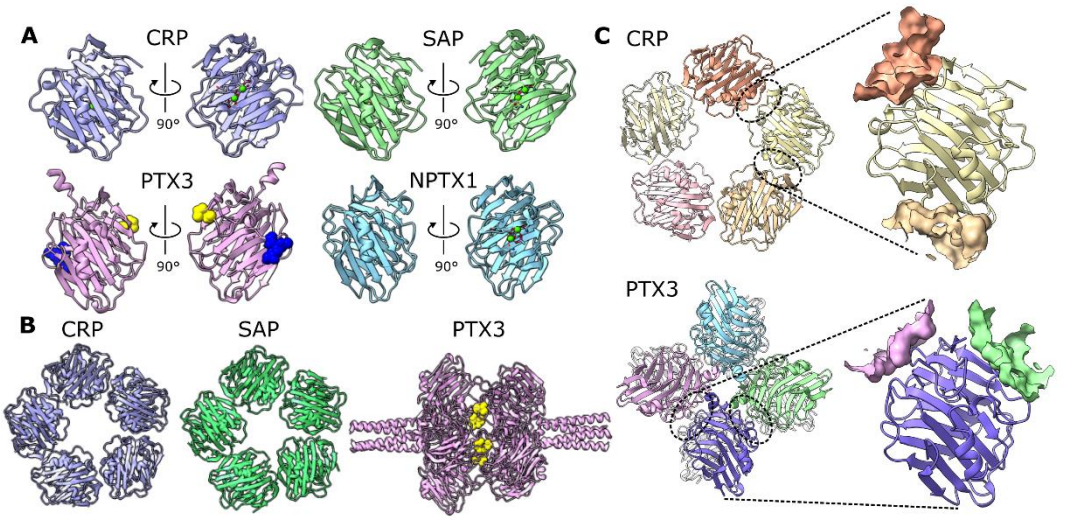

Figure 3: Tertiary and quaternary structures of each pentraxin resolved to date and interfacial regions between short and long pentraxin domains. *(**A**) Tertiary and (**B**) quaternary structures of CRP, SAP, PTX3 and NPTX1 (for NPTX1 only the tertiary structure is shown given the lack of experimental data). On PTX3 C317 and C318 are highlighted as yellow spheres and the site of the N-linked glycan (N220) as blue spheres. (**C**) Differences in the interfacial regions between the pentraxin domains of the short pentraxin CRP and the long pentraxin PTX3 are highlighted by enlarged circled sections from the atomic models. PDB codes of models used CRP:7PKE, SAP:1SAC, PTX3:7ZL1 and NPTX1:6YPE.*

1.4 The long pentraxin N-terminal domain

The additional N-terminal domain present in the long pentraxins are unrelated to one another at the primary sequence level (39) (**Table 1**). However, greater similarity is seen between the N-terminal domains of the neuronal pentraxins (NPTXs), suggesting the presence of a subclass within the long pentraxins (39) (**Table 1**). ***Chapter 6*** will reveal that in PTX3 the N-terminal domain is composed of a tetrameric coiled coil with interspersed disordered regions. The N-terminal region is also known to mediate oligomerization of the complex (36, 40, 41). Similarly, the N-terminal domains of the other long pentraxins contain predicted coiled coil and disordered sections (**Fig. 2A**), with these regions having been shown to regulate quaternary structure in the NPTXs (42). However, whether these conform to previously identified pentraxin architectures or arrange in entirely different ways remains to be seen. These regions may also be biologically important given that disordered regions often participate in protein-protein interactions (43), with the N-terminal domains of several long pentraxins being implicated in the binding of a range of ligands (41, 42, 44-46). Taken together this suggests the N-terminal domains may produce novel quaternary structures and expanded ligand binding repertoires, therefore adding new functionalities to the long pentraxins compared to the short.

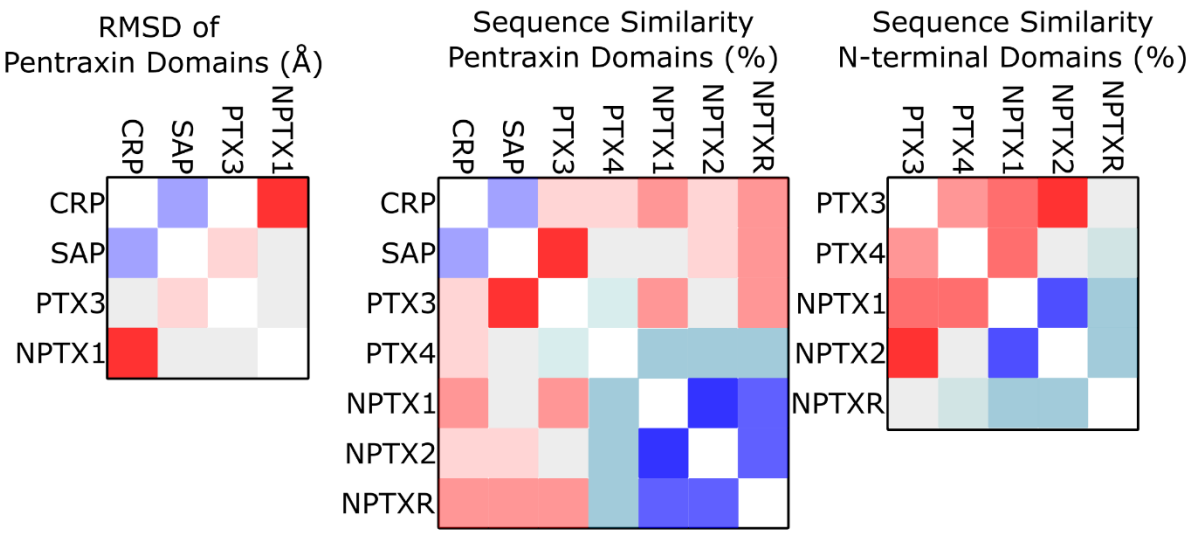

Table 1: Structural and sequence comparison of the pentraxins. *(**Left**) Root-mean-square deviation (RMSD Å) between both the experimentally determined pentraxin domain structures (**Middle**) Comparison of the primary sequence of the pentraxin domains of each pentraxin. (**Right**) Comparison of the primary sequence of the N-terminal domains of the long pentraxins, PTX3, PTX4, NPTX1, NPTX2 and NPTXR. More and less similar are represented by blue and red respectively.*

1.5 Cysteines, post-translation modifications and calcium binding

Cysteine bonds and post-translational modifications, such as glycosylation, play important roles structurally and functionally in the pentraxin family. For instance, disulfide bonds are present in each pentraxin, but whereas the short pentraxins only contain one intra-molecular disulfide bond in the pentraxin domain, the long pentraxins appear to have more complicated networks of disulfide bonding (Fig. 2A). For instance, PTX3 contains a network of disulfide bonds that play important roles in the oligomerisation of the complex (36, 40), with this theme appearing to continue for NPTX1 and NPTX2, where three cysteine residues in the N-terminal domain are also implicated in oligomerisation (42). Additionally, in other organisms such as rats, the short pentraxins can also contain inter-molecular disulfide bonds (47), further showing that disulfide bonds play important roles in the oligomeric states of pentraxins. The importance of disulfide bonds may also form the basis for the reported redox sensitivity of some of the pentraxins (48), having implications for the modulation of pentraxin bioactivity at the sites of inflammation where reactive oxygen species can modulate the local redox environment (49).

Post-translational modifications such as glycosylation are also common, with all the human pentraxins except CRP being glycosylated. Glycans have been show to modulate the biological activities of several pentraxins, for example the effect of sialylation on the interaction between PTX3 and C1q (50). Additionally, glycosylation can expand the ligand binding repertoire of both PTX3 (51, 52) and serum amyloid P component (SAP) (53), enhancing antiviral activity in both. CRP is the only non-glycosylated pentraxin, possibly reflecting the fact it is already temporarily regulated by the virtue of being the only acute phase protein in humans (54), which may negate the need for additional regulation by glycosylation.

Another common theme in the pentraxin family is the ability to bind calcium ions. In the case of CRP and SAP, each monomer binds to two calcium ions, which in turn enable them to bind to their lipid ligands (34, 37). In CRP calcium also plays a structural role, helping rigidify the protein and making it resistant to protease degradation (34, 55). The structure of the pentraxin domain of NPTX1 shows that this also binds to calcium (38), with NPTX2 also shown to have similar calcium binding properties (56). However, if these calcium ions also play structural roles in the NPTXs in addition to their biological importance remains to be seen. On the other hand, the calcium binding site in PTX3 is replaced by inter-tetramer disulfide bonds (36, 40), illustrating how this region of the pentraxin domain is commonly used for association with other molecules even in the absence of metal ions.

1.6 Pentraxins, the classical complement cascade and disease

Pentraxins are adaptor proteins of the C1 complex and play importance functions in pathogen defense(57) and in regulating autoimmune disease(58). Pentraxin mediated complement activation typically does not result in terminal pathway activation and lysis of the target. This occurs especially on apoptotic cells via interactions with regulatory proteins such as factor-H (27, 44, 59). The lack of MAC pore formation biases complement activation towards opsonization and phagocytosis and away from the lytic pathway, which is hypothesized to tailor the immune response to the silent clearance of apoptotic cells and thus prevents exposure of the immune system to autoantigens. This can occur via complement activation and deposition of C3 “eat-me” signals (60), but also via direct opsonization by pentraxins themselves, which can interact with receptors on phagocytes(19, 22-24). PTX3 has even been reported to sequester apoptotic cells away from dendritic cells, to prevent the immune system developing a response to the antigens they may display (61). In addition, CRP has also been observed to bind to necrotic cells at the sites of inflammation in rheumatoid arthritis patients (62) and to nuclear components such as chromatin in a calcium dependent manner (63). This suggests CRP also plays a role in the scavenging of possible autoantigens and taken together this reflects the important role of pentraxins in the clearance of damaged self-cells and tissues.

This gentler, less inflammatory, mode of pentraxin based complement activation is correlated with pathologies such as systemic lupus erythematosus (SLE), where a defect in cellular clearance is thought to be a major factor in the development of the pathology (58). SLE patients often have low CRP serum levels(64) with genetic deficiencies in complement receptors on phagocytes(21) and C1q(65) linked to the disorder. Further highlighting the role of pentraxins in SLE is the observation that addition of CRP to mouse models has been reported to improve symptoms(66). This is juxtaposed with increased antibody mediated complement activation (67) and deposition of MAC components in patients with lupus nephritis, a severe manifestation of SLE (68, 69). Indeed factor-H, has been shown to have a protective role in lupus(68) and is known to be bound by several pentraxins to dampen terminal pathway activation (27, 31, 44). Additionally, direct comparisons of antibody and pentraxin mediated complement activation on human cells illustrate that CRP appears to favour activation of the early components of the classical complement cascade, whereas

antibodies mediated complement activation resulted in MAC formation(59). In summary this suggests that pentraxins can activate complement in a less inflammatory manner than antibodies and are tailored more towards phagocytosis and silent clearance of possible auto-antigens, which may underpin pathologies such as SLE. Whether there is a structural reason as to why there is a difference between pentraxin and antibody-based initiation complexes is unknown. However, given the different protein architectures of pentraxins compared to antibodies, there may be a structural basis for this divergence in physiological outcomes.

However, exceptions to this model do exist, as natural IgM antibodies can also result in silent clearance of self-antigens(70, 71). This indicates that the early components of the immune response, characterized by low affinity but multivalent binding to ligands, may produce a different physiological response than the later highly adapted antibodies like IgG1 and IgG3 that bind with high affinity to their cognate antigens. Furthermore, when pentraxin based complement activation occurs in dysregulated or disease states, a more inflammatory form of immune activation with MAC deposition is seen. For instance, PTX3 induces complement mediated inflammation in cancer, attributed to loss of interactions with regulatory proteins like factor H(31, 44) and has been linked to atherosclerotic inflammation (72). Similarly, CRP colocalises with MAC pore in atherosclerotic plaques (73) and in the heart during myocardial infarction (74). Taken together this suggests that context is also important in determining how pentraxin activate complement, for example in chronic versus acute inflammatory environments. This demonstrates that studies of pentraxins in conditions mimicking the inflammatory microenvironment such as acidosis(75-82) may reveal how these proteins change in terms of structure and function(30, 35, 83-85) to become dysregulated under certain conditions. Moreover, pentraxin research could have far reaching repercussions, with the recently highlighted role of complement based opsonophagocytosis of synapses in the development of Alzheimer's (60) and the ability of the NPTXs to bind C1 at synapses (32).

In summary, understanding the differences between pentraxin and antibody mediated complement activation, and pentraxin mediated complement activation under different physiologically relevant microenvironments, would enable a deeper understanding of disorders ranging from autoimmunity to cancer to Alzheimer's disease. This in turn could open avenues in the development of drugs aiming to modulate pentraxin mediated complement activation and bias it towards different outcomes.

1.7 Overview of the thesis

The aim of this thesis is to interrogate the structural and biochemical immunology of pentraxins and their interactions with the complement system, with particular focus on the short pentraxin CRP and the long pentraxin PTX3. This thesis contains novel structural information regarding both CRP and PTX3, as well as preliminary data on NPTXR, with an aim to start a conversation on how pentraxins work at a structural level, and how manipulation of these proteins may allow the development of new therapies. Chapter 2, Cryo-Electron Microscopy and Biochemical Analysis Offer Insights Into the Effects of Acidic pH, Such as Occur During Acidosis, on the Complement Binding Properties of C-Reactive Protein, explores the

structural and biochemical changes CRP undergoes in conditions of acidosis such as those found in the inflammatory microenvironment. This has implications for the possible environmental inputs CRP uses to selectively activate the immune system. Chapter 3, Structural basis for surface activation of the classical complement cascade by the short pentraxin C-Reactive Protein, shows how CRP binds to liposome cell mimetics to produce higher-order oligomers that bind and activate the C1 complex. This chapter also deconvolutes the structural basis for agglutination, with implications for the role of CRP in phagocytosis. Chapter 4, An acidic microenvironment causes CRP to bind to both LDL and the complement cascade, explores how CRP binds to LDL via apoβ in normal human serum (NHS) at an acidic pH and causes deposition of proteins dependent on both the classical and alternative pathway. This is proposed to start a positive feedback loop via pH dependent interactions with complement components such as C3 and MAC, with implications for conditions such as atherosclerosis. Chapter 5, A pipeline for screening CRP based therapies: a nanobody and TRAP based approach, describes a short biochemical pipeline enabling the screening of nanobodies for their ability to modulate the function of CRP. This may allow the future screening of CRP based therapies ranging from small peptides to monoclonal antibodies. Chapter 6, PTX3 structure determination using a hybrid cryoelectron microscopy and AlphaFold approach offers insights into ligand binding and complement activation, represents the first time a complete atomic model of a long pentraxin has been resolved to date, using a hybrid methodology of traditional single particle analysis (SPA) and AlphaFold-based structure prediction. This shows how the long pentraxins may have radically different quaternary structures and therefore functions compared to the short pentraxins, CRP and SAP. This chapter also presents a useful way to resolve and model flexible regions of protein complexes. Chapter 7, A platform for the interrogation of membrane bound PTX3, presents a liposome cell mimetic system for conjugating tagged PTX3 for interrogation via cryoelectron tomography. This chapter also presents the discovery of a novel native membrane ligand for PTX3, with thoughts on how this can be adapted to image PTX3 based C1 activation in the future. Chapter 8, General discussions– immunity on the mind discusses the data presented in this thesis and their possible implications. It also presents some low-resolution data of the long pentraxin NPTXR and a possible model to how the NPTXs function.