Comparative analyses of pentraxins: implications for protomer assembly and ligand binding

N Srinivasan¹, Helen E White¹, Jonas Emsley¹, Steve P Wood¹,
Mark B Pepys² and Tom L Blundell^{1*}

¹Laboratory of Molecular Biology and ICRF Unit of Structural Molecular Biology, Department of Crystallography, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK and ²Immunological Medicine Unit, Department of Medicine, Royal Postgraduate Medical School, Hammersmith Hospital, London W12 0NN, UK

Background: Pentraxins are a family of plasma proteins characterized by their pentameric assembly and calcium-dependent ligand binding. The recent determination of the crystal structure for a member of this family, human serum amyloid P component (SAP), provides a basis for the comparative analysis of the pentraxin family.

Results: We have compared the sequences, tertiary structures and quaternary arrangements of SAP with human C-reactive protein (CRP), Syrian hamster SAP (HSAP) and *Limulus polyphemus* CRP (LIM). These proteins can adopt a β -jelly roll topology and a

hydrophobic core similar to that seen in SAP. Only minor differences are observed in the positions of residues involved in coordinating calcium ions.

Conclusions: Calcium-mediated ligand binding by CRP, HSAP and LIM is similar to that defined by the crystal structure of SAP, but sequence differences in the hydrophobic pocket explain the differential ligand specificities exhibited by the homologous proteins. Differences elsewhere, including insertions and deletions, account for the different (hexameric) quaternary structure of LIM.

Structure 15 November 1994, 2:1017–1027

Key words: acute phase response, C-reactive protein, hamster SAP, *Limulus* CRP, subunit aggregation

Introduction

The pentraxins are a family of oligomeric plasma proteins with the capacity for specific calcium-dependent ligand binding and named for the pentameric arrangement of their protomers [1,2]. They have been conserved throughout vertebrate evolution with retention of ligand specificity, subunit arrangement, and substantial amino acid sequence similarity [3–5]. Their production can be modulated by cytokines, including interleukin (IL)-1, IL-6 and tumour necrosis factor (TNF) [6], and they participate in inflammation and host defence, although they may also have other functions [2,7–9]. The phylogenetically oldest member of the family, the C-reactive protein of an arachnid, *Limulus polyphemus* (the horseshoe crab), has a hexameric rather than a pentameric structure [10].

The human plasma pentraxins are C-reactive protein (CRP), the classical acute phase reactant, and serum amyloid P component (SAP), which is also a normal extracellular matrix protein and a universal constituent of the tissue deposits in the pathological condition of amyloidosis [2]. Human CRP and SAP are clearly products of gene duplication and pentraxins in other vertebrates generally resemble more closely one or other of CRP or SAP. However, there is much overlap in precise ligand specificity, behaviour as acute phase proteins, and deposition in amyloid. Another human protein, PTX3 [11] or TSG-14 [12], with some sequence similarity to the pentraxins has lately been described but has not yet been well characterized, and it is not known whether it shares

the hallmark features of pentameric assembly or calcium-dependent ligand binding.

The calcium-dependent ligand specificity of human CRP for phosphocholine (PC) [13] is of particular interest as it is not shared by human SAP (WL Hutchinson and MB Pepys, unpublished data), although they both bind to phosphoethanolamine (PE) [14,15]. On the other hand, human SAP specifically binds to the 4,6 cyclic pyruvate acetal of β -D-galactose (MO β DG) [16] and also to all types of amyloid fibrils [17], which are not recognized by human CRP. The major pentraxin in Syrian hamsters, formerly known as female protein [18] (here designated HSAP), is the hamster counterpart of human SAP, in that it binds to hamster amyloid [19,20], but it shares with human CRP the capacity to bind to PC as well as PE [18,21].

Recently, the three-dimensional (3D) structure of human SAP has been elucidated at high resolution by X-ray analysis [22]. The protomer has a flattened β -jelly roll topology which is remarkably similar to legume lectins such as concanavalin A [23,24] and pea lectin [25], bacterial 'lectins' such as glucanase [26], endo 1,4- β -xylanase II [27] and cellobiohydrolase I [28], human S-Lac lectin [29] and vertebrate lectin [30]. However, the metal ions and carbohydrate ligand binding sites in SAP and legume lectins are located in different positions on the common fold and the quaternary structures are different (N Srinivasan *et al.*, & TL Blundell, unpublished data). Crystal structures of SAP complexes show

*Corresponding author.

that SAP binding to MO β DG and PE is mediated by two calcium ions [22].

In view of the high degree of sequence similarity between human SAP, human CRP and pentraxins in other species it should be possible to identify structure–function relationships in these other proteins on the basis of the human SAP structure. Using the comparative modelling approach encoded in COMPOSER [31–35] we have built 3D models for the sequences of human CRP [36,37], HSAP [38] and *Limulus* CRP (LIM) [39,40] on the basis of the crystal structure of human SAP. We have constructed pentamers of CRP and HSAP based on the pentamer defined in the crystal structure of SAP and identified the interactions that stabilize subunit aggregation. By operating a six-fold axis, whose position is determined empirically, we have generated a model of the hexamer of LIM and analyzed inter-protomer interactions. Extrapolation of ligand binding modes of SAP to the 3D structure of CRP suggests explanations for the relative affinities of CRP for PC, PE and MO β DG. These comparative analyses of structure and function of the pentraxin family provide a basis for the design of specific ligands of high affinity.

Results and discussion

Comparative analysis of the pentraxin protomers

Fig. 1 shows the alignment of the sequence of SAP with those of CRP, HSAP and LIM; sequence identities with SAP are 52%, 67% and 30%, respectively. Sequences of CRP, HSAP and LIM can be modelled on the fold of SAP with the β -strands and the α -helix conserved. The root mean square deviations (rmsds) between topologically equivalent C α atoms of SAP and the various models are 0.15 Å (CRP), 0.06 Å (HSAP) and 0.65 Å (LIM), with the largest differences occurring as expected in the loop regions. Fig. 2 compares the tertiary structure of SAP defined by X-ray analysis, with those of CRP, HSAP and LIM defined by comparative modelling. Regions with significantly different amino acids relative to SAP are highlighted. Most amino acid substitutions occur in loops, with some in β -sheets, especially in LIM.

CRP versus SAP

The insertion of Ser44A–Thr44B (SAP numbering) in CRP relative to SAP has been modelled by replacing ELSSSTR (42–45) with the segment 411–416 of taka amylase (PDB entry code 2TAA). This segment was chosen based on a search for suitable segments in a database of known structures. The insertion region 42–45 of CRP is spatially proximal to another region, Pro89–Val90A which differs from SAP. Pro89 of SAP is in the *cis* conformation, but no equivalent proline is present in CRP. Hence the region Val87–Thr88–Val89 of CRP has been modelled, after a general search of known 3D structures, on 52–54 of subtilisin *Novo* (PDB entry code 2SNI).

Most of the more radical sequence differences in CRP compared with SAP appear to stabilize the structure. For example, the hydrophobic Val16, Phe124 and Leu170, which are surprisingly exposed in SAP, are replaced in CRP by the more hydrophilic residues aspartate, threonine and asparagine, respectively. Thr24, which is buried and hydrogen bonded in SAP, is replaced by alanine in CRP and is in an apolar environment provided by Leu26, Val182 and Leu183. The side chain of Thr90 is hydrogen bonded to that of Asp70, thereby replacing the hydrophobic interaction between Phe90 and Val70 in SAP. Other differences in interactions that may stabilize the structure of the protomer are listed in Table 1.

HSAP versus SAP

There are two regions in HSAP which significantly differ from SAP, one corresponding to the *dis*-proline region of SAP and the other corresponding to the eight-residue carboxy-terminal extension. As in the case of CRP, the former region of HSAP was modelled on a segment from 2SNI. A seven-residue overlap with the preceding conserved region was used to identify the segment, starting from Thr233 of the B-chain of rat mast cell proteinase (PDB entry code 3RP2), as the most suitable carboxy-terminal extension.

Despite high sequence identity between HSAP and SAP, there are some amino acid substitutions that result in differences in interactions between residues (Fig. 1 and Table 1). For example, although Val16 of SAP is replaced by glutamate in HSAP, the side chain of glutamate can form a hydrogen bond to the carbonyl oxygen at Phe144. Glu70 of HSAP is involved in two ion pairs; one with Lys81 and the other with His83 which are threonine and lysine, respectively, in SAP. Phe144 in HSAP packs with the Tyr19, which is a histidine in SAP. Side chains of Asn21 and Arg193 interact in SAP and these residues are changed to lysine and asparagine, respectively, in HSAP.

LIM versus SAP

LIM has many insertions and deletions, reflecting its more distant relationship with SAP. The nine regions with most structural variation from SAP were successfully modelled on the basis of fragments from proteins of known 3D structures in the Brookhaven Protein Data Bank [41] and are listed in Table 2. Amino acid substitutions in LIM relative to SAP occur throughout the molecule, including the β -sheets forming the jelly roll fold. These differences alter the packing and nature of the interactions between the side chains, but they allow the overall 3D structure to be maintained. For example, the interaction between two hydrophobic residues Phe8 and Leu188 in the β -strands in SAP is replaced by a polar–polar interaction between Lys9 and Thr188 in LIM. The interaction between the side chains of Lys65 and Tyr71 in SAP is replaced by a hydrophobic contact between Ile65 and Phe71 in LIM. The interaction between the polar side chains Glu66 and Ser72 in SAP is replaced by the hydrophobic interaction between Leu64 and Leu72 in LIM. In SAP, the side chain of Ser41 is

Table 1. Predicted differences in interactions within a protomer.

Residue or residue pair in:	Interacting groups in: ^a
SAP	CRP
Ile23–Glu191	Lys23–Glu191
Asn32–Ser82	Ala32–Phe82
Asn53–Tyr140	Ala53–Phe140
Lys65–Val92	Trp65–Val92
Val68–Phe88	Asp68–Thr88 & Thr44B
Gly69–Val84	Ile69–Val84
Val70–Phe90	Asp70–Thr90
Val96	Thr96 H-bonded to MC at Ser97
Leu114–Ile109	Arg114–Tyr71
Non-H-bonded Lys143	Asn143–Ser139
SAP	HSAP
Val16–Phe144	Glu16 H-bonded to MC at Phe144
His19–Phe144	Tyr19–Phe144
Asn21–Arg193	Lys21–Asn193
Glu70–Thr81 & Lys83	Glu70–Lys81 & Arg83
Ser82–Leu114	Val82–Trp114
Val96	Thr96 H-bonded to MC at Val95
Leu170	Lys170 H-bonded to MC at Gly111
SAP	LIM
Phe8–Leu188	Lys9–Thr188
Lys65–Tyr71	Ile65–Phe71
Glu66–Ser72	Leu64–Leu72
Phe11–Ile154	Pro11–Leu154
Val196–Val20 & Ile192	Val196–Pro20 & Ala192
Ser41 H-bonded to MC at Val151	Val41–Leu150
Asn53–Asp58	Ala53–Val140
Val63–Ser82	Thr63–Gln82
Ala108–Ile169	Trp108–Ile169
Trp160–Ala172 & Tyr158	Trp160–Leu172 & Tyr173

^aUnless specified the interactions are between side chains. MC = main chain.

Table 2. Modelling of structurally variable regions (SVRs) in LIM.

SVRs of LIM ^a	Modelled on: ^b
– 2 to – 1	2GD1 Ala O 0
11 to 13	2TBV Asp B 217
29 to 30	4TMS Arg 178
42A to 44	1WSY Ala B 269
75A to 78	4APE Gly 21
85 to 91A	1LZ1 Gly 55
114 to 116	1BMV Gly 1 1078
174 to 180	2GD1 Asp P 293
204 to 204A	2CRO Gly 62

^aResidue numbering is based on SAP (see Fig. 1). ^bThe protein fragments used for modelling are derived from proteins of known structure. Their PDB codes are given along with the first residue number of the chosen segment. Chain identifiers in the structures with more than one polypeptide chain in the PDB file are specified before the residue number.

in SAP that is retained in CRP by the replacements Lys→Arg and Glu→Asp, respectively.

Calcium binding occurs on one side of the pentamer, on the surface composed of the more concave β -sheets. On the opposite more planar face, each subunit has an

α -helical segment [22]. The existence of SAP in pentameric or decameric form is dependent on pH and the nature and concentration of buffer salts [42] (WL Hutchinson and MB Pepys, unpublished data). The decamer is thought to be formed by packing two pentamers in such a way that the faces containing helices are packed with an angle between the helices from different pentamers; in this arrangement the calcium-containing β -sheets of each pentamer are accessible [22]. The charge distribution at the helix-containing face of the SAP pentamer is shown in Fig. 3a. Both positive and negative charges permit electrostatic interactions between the two pentamers of the decamer. The presence of more negative than positive charges means that not all charges are paired and is indicative of a weak association between the two pentamers. Fig. 3b shows the charge distribution in the helix-containing face of the CRP pentamer. A conspicuous feature is the cluster of strong negative charge at the centre of the pentameric ring. This distribution of charge would lead to repulsion of pentamers in a putative decamer similar to that of SAP. Such repulsion could not be eliminated by the rotation of one of the pentamers with respect to the other as the strong negative charge is evenly spread at the centre. This may explain the observation that CRP does not form a decamer [43].

Most inter-protomer contacts, including ion pairs and hydrophobic contacts, are similar in HSAP and SAP (Table 3). For example, substitution of Ile85 in SAP by valine in HSAP allows retention of the hydrophobic contact with the conserved Phe88. Subunit interactions in HSAP are increased by the presence of a carboxy-terminal extension. Fig. 4 shows the superposition of SAP and HSAP in this inter-protomer contact region.

LIM forms hexamers rather than pentamers [10], which we have modelled as in Fig. 5. The inter-protomer contacts in the hexamer of LIM are listed in Table 3. Two regions where there are deletions of one residue in LIM compared with SAP, 10–12 and 83–89 (see Fig. 1), and the region of two inserted residues in LIM after position 114 of SAP, are in the inter-protomer contact regions and the differences may accommodate hexamer rather than pentamer formation. Although the precise details of inter-protomer interactions might vary depending upon the method of hexamer generation, the aggregation of LIM as a pentamer seems unfavourable. If LIM forms a pentamer similar to SAP, apart from the loss of some of the inter-protomer contacts due to deletions, the insertion of lysine and glycine after position 114 would result in steric clashes with an adjacent protomer. This is also supported by a model of a hypothetical pentamer of LIM which has been constructed in an analogous fashion to the pentameric SAP.

Calcium binding

Calcium coordination is essential for ligand binding by pentraxins [2] and for the resistance of pentraxins to proteolytic degradation [44,45]. There are two calcium

Table 3. Inter-protomer interactions.^a

SAP	CRP	HSAP	LIM
Val10-Ile104	Val10-Ile104	Val10-Ile104	Phe10-Trp101
Pro12(M)-Gly118(M)	Pro12(M)-Ser118(M)	Pro12(M)-Gly118(M)	Lys40-Cys114(M)
Tyr40-Pro113(M)	Tyr40-Pro113(M)	Tyr40-Pro113(M)	Lys40-Glu104
Tyr40-Val115	Tyr40-Val115	Tyr40-Val115	Asn42-Asn115(M)
Asp42-Lys83(M)	Glu42-Glu83(M)	Asp42-Arg83(M)	Asn42-Lys114A
Asp42(M)-Lys83	Ser44(M)-Glu83	Asp42(M)-Arg83	
Asp42-Lys117	Glu42-Lys117	Asp42-Lys117	
	Ser44 & Ser44A-Lys67		
	Thr44B-Asp68		
	Val88A-Pro85		
Ile85-Phe88 & Pro89	Pro85-Val88A	Val85-Phe88 & Ala88A	
			Lys91-Lys114A(M)
Glu99-Lys199	Glu99-Lys199	Glu99-Arg199	Trp101-Val197
Ser102(M)-Tyr195	Ser118-Tyr195	Ser102-Tyr195	Glu102-Ser199
Ile104-Ile197	Ile104-Phe197	Ile104-Val197	Thr106-Glu202
Pro113-Trp203	Pro113-Trp203	Pro113-Val203	His113-Glu202
Val115(M)-Gly152(M)	Val115(M)-Gly152(M)	Val115(M)-Gly152(M)	
Lys116-Glu153	Arg116-Asp 153	Lys116-Glu153	Ala116M-Glu151
			Val121-Val197
Pro166-Val202	Pro166-Leu202		His166-Glu202
		Lys112-Ser204B	

^aInteractions involving the main chain are indicated by 'M' at the residue position.

Fig. 3. Representation of the charge distribution calculated using the program GRASP that incorporates the DelPhi algorithm [56]. The pentameric face, containing the helical segments, is shown for (a) SAP and (b) CRP. The charge distribution is represented with the colour code red for negative and purple for positive charges.

ions bound to SAP. Calcium 1 is coordinated to side chains of Asp58 (both oxygens), Asn59, Glu136, Asp138 and the main chain carbonyl of Gln137. The other ligand is an acetate in four of the subunits and the side chain of Glu167 from an adjacent pentamer in the crystal lattice in the fifth subunit. Calcium 2 is less extensively coordinated by SAP, with ligands to Glu136, Asp138, Gln148, and acetate. The coordination is completed by Glu167 and two water molecules [22].

The equivalent residues in CRP, HSAP and LIM are indicated in Fig. 1.

Unlike SAP, CRP binds to the two calcium ions with equal affinity at neutral pH [44]. The backbone at the calcium-binding site of CRP is similar to that in SAP. However, there are some residue changes which could affect the affinity for calcium. In SAP, residue 145 is aspartate, which is not liganded to a calcium ion, but in

Fig. 4. Inter-protomer region of HSAP (green) superimposed on the inter-protomer region of SAP (pink). The two carboxyl termini are indicated. (Figure generated using SETOR [53].)

Fig. 5. Hexameric model of LIM. The β -strands and helical segment are represented in pink and blue respectively. Disulphides are shown in yellow. (Figure generated using SETOR [53].)

CRP the larger side chain of glutamate could coordinate the calcium.

In HSAP, the calcium ligands Asn59, Glu136, Gln137, Asp138 and Gln148 are conserved although Asp58 of SAP is replaced by glutamate in HSAP, a change which is unlikely to have much effect on the binding of calcium ions. In a similar way, all calcium-binding residues in LIM are identical to those in SAP with the exception of Glu136 in SAP which is replaced by aspartate.

Calcium-mediated ligand binding

SAP shows calcium-dependent binding of PE but does not bind PC. In the crystal structure of the SAP-PE complex [22], the residues Tyr74, Lys79 and Glu66 are situated close to PE with the side chain hydroxyl of Tyr74 forming a hydrogen bond with the nitrogen of PE. Glu66 and Lys79 form a salt bridge.

PE and PC differ only at the nitrogen where three methyl groups are attached in PC. Modelling of the three methyl groups (PC) resulted in short contacts ranging from 2.1 Å to 2.7 Å between the hydroxyl group of Tyr74 and the methyl carbons. These could not be relieved by rotation around the ethyl nitrogen bond as the side chain hydroxyl of Tyr74 is situated roughly on the rotation axis (Fig. 6) and there is no space for the PC to re-position itself.

Thr74 of CRP is topologically equivalent to Tyr74 of SAP. A shorter side chain at 74 (Tyr→Thr) results in longer distances between the side chain atoms of Thr74 and the three methyl groups of PC (distances range between 3.8 Å and 6.5 Å). Glu66 and Lys79, which form an ion pair in SAP, are replaced by serine and glutamate, respectively. Distances between the side chain atoms of these residues and methyl groups of PC range between 4.9 Å and 8 Å, thus accommodating the methyl groups of PC (Fig. 6).

In the complex of PE with SAP, the hydroxyl of Tyr74 forms a hydrogen bond with the nitrogen of PE. In

Fig. 6. Docking of phosphocholine. Side chains of SAP (pink) and CRP (blue) are shown superimposed. Phosphocholine is shown in yellow and calcium ions in red. (Figure generated using SETOR [53].)

CRP, Thr74 is too far from the nitrogen of PE to directly stabilize the complex. However, Lys79 in SAP is replaced by a glutamate in CRP, whose side chain could form a hydrogen bond with the nitrogen of PE.

These observations provide an explanation for the binding properties for PE and PC of SAP and CRP. Liu *et al.* [46] and Swanson and Mortensen [47] suggest that the residues 50–60 of CRP are important in binding PC while the mutagenesis work of Agrawal *et al.* [48] implicated Lys55, Arg56 and Trp67 as the key residues. However, it is likely that changes at positions 55, 56 and 67 will have resulted in a structural perturbation at the calcium-binding site.

HSAP resembles CRP in binding both PC and PE [15]. Tyr74 and Glu66, which are close to PC, are conserved, but the third side chain (Lys79 in SAP) that is close to PC is replaced by glycine in HSAP, thus creating more space to accommodate PC. LIM also binds both PC and PE. Tyr74 in SAP, which binds directly with the nitrogen of PE, is replaced by asparagine in LIM. Other changes include Lys79 and Glu66 in SAP which are Leu79 and Asp66 in LIM. Despite these substitutions, which alter the shape/charge of the binding site, both PC and PE can be bound by LIM.

Binding of MO β DG by SAP involves two hydrogen bonds formed between the 4,6 oxygen atoms of the carboxyethylidene ring and the side chains of Gln148 and Asn59 respectively, in addition to the direct interaction between calcium ions and the acidic group of MO β DG [22]. In the crystal structure without an added ligand, acetate from the buffer binds calcium in a similar way. CRP binds MO β DG very weakly [49]. Interestingly, both Gln148 and Asn59 are conserved in CRP (Fig. 1) but there are changes in residues that define the hydrophobic pocket: Tyr64 is a phenylalanine and Tyr74 is a threonine (Fig. 7). These changes alter the size of the pocket as well as the charge distribution and they may account for weaker binding of MO β DG. The higher affinity of CRP for calcium, possibly mediated through Gln145, may also lead to an alteration of the position of calcium 2, and thereby weaken binding of MO β DG.

HSAP binds weakly to MO β DG compared with binding of MO β DG by SAP [21]. The residues in SAP that directly interact with MO β DG (Asn59 and Gln148) and the residues forming the hydrophobic pocket (Leu62, Tyr64 and Tyr74) are all conserved in HSAP (see Fig. 1). However, replacement of Lys79 in SAP by glycine in HSAP may account for the reduced binding of MO β DG.

LIM does not bind to MO β DG (GA Tennent and MB Pepys, unpublished data). Although residues Asn59, Gln148 and Leu62 which interact with MO β DG in SAP, are conserved in LIM, the other hydrophobic residues in the pocket are significantly different in LIM, i.e., Tyr64 and Tyr74 in SAP are Leu64 and Asn74 in LIM (Fig. 1).

Fig. 7. Superimposition of residues in SAP and CRP in the vicinity of the MO β DG binding site. The backbone is shown in green. The side chains of SAP are shown in pink and those of CRP in blue. MO β DG is represented in yellow and calcium ions in red. (Figure generated using SETOR [53].)

Biological implications

The pentraxins are a family of pentameric vertebrate plasma proteins that have been stably conserved in evolution and share sequence homology, similar subunit assembly and the capacity for specific calcium-dependent ligand binding. The phylogenetically oldest known protein with sequence homology is C-reactive protein (LIM) from the arachnid, *Limulus polyphemus* (the horseshoe crab), which differs from its vertebrate counterparts in being hexameric. Human C-reactive protein (CRP) is the classical cytokine-regulated acute phase protein which is produced in response to inflammation and tissue damage. It is universally measured in clinical practice as an objective marker of disease activity. Human serum amyloid P component (SAP), the other human pentraxin, is a normal constituent of the extracellular matrix as well as a plasma protein. It is also invariably found in the deposits associated with the pathological condition of amyloidosis that underlies such common and important clinical disorders as Alzheimer's disease and diabetes mellitus. SAP probably contributes to persistence of amyloid *in vivo* and radiolabelled SAP is used routinely for diagnosis and monitoring of systematic amyloidosis. Serum amyloid P component from the Syrian hamster (HSAP) is of interest because it shares properties of both human CRP and SAP.

Comparative analyses of the sequences and structures of SAP, CRP, HSAP and LIM, based on the high-resolution crystal structure of SAP provide strong evidence for a highly conserved tertiary fold, despite differences in quaternary structure

and ligand specificity. The multimeric assembly of protomers with their binding sites on one planar face of the pentamer (or hexamer in the case of LIM) enables multivalent attachment. The predominantly β -sheet arrangement of the protomer fold with short, tightly bound joining loops, together with the multiple coordination of calcium ions by residues in solvent-exposed segments, account for the remarkable proteinase resistance of pentraxins. This resistance may be advantageous in the proteinase-rich microenvironment of inflammatory lesions, and it is of interest that CRP is even more proteinase resistant than SAP (GA Tennent and MB Pepys, unpublished data) and that most of the residue differences appear likely to further stabilize the CRP protomer structure.

The models also provide a structure-based explanation of the different ligand specificities among pentraxins, in particular the recognition of phosphocholine by CRP, HSAP and LIM but not by SAP, the recognition of cyclic pyruvate acetals only by SAP, and the recognition of phosphoethanolamine by all of them. Final confirmation will require crystallographic analyses of ligand complexes, but mutagenesis experiments, currently in progress, should provide an experimental test of the proposed interactions. Better understanding of pentraxin-ligand binding may help to elucidate their physiological and pathological functions, and contribute to the development of new diagnostic and therapeutic agents in amyloidosis and inflammation.

Materials and methods

Comparative modelling

The 3D structures of CRP, HSAP and LIM were modelled on the basis of the crystal structure of SAP determined at 2 Å resolution and refined to an R-factor of 0.18 [22]. The suite of programs, encoded in the comparative modelling software COMPOSER [31,32] and incorporated in SYBYL (version 6.0 – Tripos Inc.), was used to obtain initial 3D models. Conserved regions of CRP, HSAP and LIM were modelled by extrapolating from the equivalent regions of SAP; they comprise 15 stretches of β -strands, most of which are also present in lectins, a long α -helical segment at 166–174 which is unique to pentraxins, and some conserved loops. The insertion and deletion regions were modelled by searching a data base of fragments, extracted from known 3D structures, for an appropriate segment satisfying end-to-end inter-C α distances of flanking conserved regions. Steric contact with the rest of the protein was also a consideration in the loop selection. The template matching approach was used in ranking the candidate loops [50]. The selected loop segment was modelled on to the three-residue overlap with flanking SCRs (structurally conserved regions), using an algorithm to relax the structure by 'tweaking' at the ends of SCRs (F Eisenmenger, personal communication). Side chain conformations have been predicted using a set of 1200 rules based

on the conformers in equivalent residues in SAP or from a rotamer library, depending on the nature of substitutions and the type of secondary structure [34].

Energy minimization

The models were energy minimized using the AMBER force-field [51] as incorporated in SYBYL. Hydrogen atoms were also considered in the energy minimization. For every course of energy minimization, 20 cycles of Simplex method and a further 100 cycles of Powell algorithm were employed. During the initial cycles of energy minimization, the backbone atoms were maintained fixed and electrostatic energy terms were not considered. This strategy was employed primarily to relieve short contacts and to correct geometry, particularly at the 'anchor' regions of various loops. In the subsequent cycles, all atoms were allowed to move and electrostatic energy terms were included. A distance-dependent dielectric constant with a cut-off of 9 Å and an ϵ value of 4 was used. Minimization was performed until all inconsistencies in geometry were rectified and all short contacts were relieved.

Modelling of LIM

As LIM is more distantly related to SAP (30% sequence identity) many, usually conservative, amino acid replacements are present, even in the solvent-inaccessible core region. Hence a slightly different strategy, which combines constrained energy minimization at every stage of modelling, was employed to model LIM.

Initially, the structurally conserved regions of LIM, which comprise the β -strands and helix, were built using COMPOSER with SAP as the basis structure. Then the backbone of LIM was kept rigid and the side chains alone were allowed to optimize their positions during energy minimization. Each of the variable regions was built individually and checked for geometry and steric clashes with the 'anchor' regions and rectified by a simulated annealing protocol in SYBYL. Many loop conformations were adjusted manually using interactive graphics and a careful check has been made for the compatibility of conformation at each position [50]. Energy minimization of the (local) loop regions enabled steady improvement in the stereochemical quality of the model. The two disulphides that are present in LIM, in addition to the disulphide in common with SAP, were modelled after using distance constraints on the selection of fragments from the database [52]. Finally, the completed model was energy minimized by allowing all the atoms to move and with an electrostatic term included in the force field.

Generation of quaternary structures

The pentamers of CRP and HSAP were generated by superimposing the model on to each protomer of SAP; special care was taken not to alter the coordinate axes of reference. The generated pentamer model was then energy minimized by moving side chains alone, in order to relieve short contacts at the inter-protomer interfaces. Electrostatic terms were also included in these minimization cycles.

In order to generate the hexamer of LIM, a hypothetical pentamer of LIM was first generated by superposing the LIM model on to each of the protomers of SAP. The centres of C α atomic positions of the protomers define a pentagon

Fig. 8. Generation of a hexameric subunit arrangement for *Limulus* CRP (LIM) based on the pentameric SAP structure. **(a)** The hypothetical pentamer of LIM with the centre of the C α atoms in each of the protomers marked as A to E. The point P refers to the centre of the pentagon of points A to E and the five-fold axis passes through it. **(b)** The triangle ABP from the pentamer is used to identify a point H away from the five-fold axis and passing through the six-fold axis. **(c)** A preliminary model of the hexameric LIM.

whose vertices and centroid are in a plane (points A to E and P in Fig. 8a). The distance from the point P (and hence its coordinates) to a point H in the six-fold axis and in the mean plane formed by the points A to E can be calculated by considering the triangles ABP and ABH where the angles of the triangles and the distances AB, AP and PB are known (Fig. 8b). Using the six-fold axis that passes through H and parallel to the five-fold axis, a preliminary model for a hexamer of LIM was generated (Fig. 8c). The individual protomers were then systematically translated in small steps away from the six-fold axis. For every step of translation, the number of distances, within 10 Å between C β atoms across the two adjacent protomers, was monitored. Translation was terminated when the number of interacting C β atoms was same as in the pentamer of SAP. The hexameric model was then energy minimized by keeping the backbone fixed and allowing side chains to move.

Generation of CRP models bound to PE, PC and MO β DG was accomplished by alterations in side chain positions at the binding site, as guided by the structures of the complexes, SAP-PE and SAP-MO β DG [22].

All computations and development of programs for modelling and analysis were performed on a network of IRIS-SG workstations. Three-dimensional models were analyzed using interactive graphics programs such as SETOR [53] and SYBYL (Tripos Inc., St. Louis) on IRIS workstations and FRODO [54] on an Evans and Sutherland system.

Atomic coordinates of the models are being deposited in the Protein Data Bank [41].

Acknowledgements: This research was supported in part by MRC and ICRF grants to TLB and MRC programme grant G790051 to MBP. NS is supported by Tripos Inc. We thank Ms Beth Sontrop for expert preparation of the manuscript.

References

- Osmand, A.P., Friedenson, B., Gewurz, H., Painter, R.H., Hofmann, T. & Shelton, E. (1977). Characterisation of C-reactive protein and the complement subcomponent Clt as homologous proteins displaying cyclic pentameric symmetry (pentaxins). *Proc. Natl. Acad. Sci. USA* **74**, 739–743.
- Pepys, M.B. & Baltz, M.L. (1983). Acute phase proteins with special reference to C-reactive protein and related proteins (pentaxins) and serum amyloid A protein. *Adv. Immunol.* **34**, 141–212.
- Pepys, M.B., Dash, A.C., Fletcher, T.C., Richardson, N., Munn, E.A. & Feinstein, A. (1978). Analogues in other mammals and in fish of human plasma proteins C-reactive protein and amyloid P component. *Nature* **273**, 168–170.
- Baltz, M.L., et al., & Pepys, M.B. (1982). Phylogenetic aspects of C-reactive protein and related proteins. *Ann. N.Y. Acad. Sci.* **389**, 49–75.
- Pepys, M.B., et al., & Fletcher, T.C. (1982). C-reactive protein and serum amyloid P component in the plaice (*Pleuronectes platessa* L.), a marine teleost, are homologous with their human counterparts. *Biochim. Biophys. Acta* **704**, 123–133.
- Steel, D.M. & Whitehead, A.S. (1994). The major acute phase reactants: C-reactive protein, serum amyloid P component and serum amyloid A protein. *Immunol. Today* **15**, 81–88.
- Pepys, M.B., Rowe, I.F. & Baltz, M.L. (1985). C-reactive protein: binding to lipids and lipoproteins. *Int. Rev. Exp. Pathol.* **27**, 83–111.
- Vigushin, D.M., Pepys, M.B. & Hawkins, P.N. (1993). Metabolic and scintigraphic studies of radioiodinated human C-reactive protein in health and disease. *J. Clin. Invest.* **91**, 1351–1357.
- Tennent, G.A. & Pepys, M.B. (1993). Glycobiology of the pentaxins. *Biochem. Soc. Trans.* **22**, 74–79.
- Tennent, G.A., et al., & Pepys, M.B. (1993). Molecular characterization of *Limulus polyphemus* C-reactive protein. I. Subunit composition. *Eur. J. Biochem.* **214**, 91–97.
- Breviario, F., et al., & Introna, M. (1992). Interleukin-1-inducible genes in endothelial cells. Cloning of a new gene related to C-reactive protein and serum amyloid P component. *J. Biol. Chem.* **267**, 22190–22197.
- Lee, G.W., Lee, T.H. & Vilcek, J. (1993). TSG-14, a tumor necrosis factor- and IL-1-inducible protein, is a novel member of the pentaxin family of acute phase proteins. *J. Immunol.* **150**, 1804–1812.
- Volanakis, J.E. & Kaplan, M.H. (1971). Specificity of C-reactive protein for choline phosphate residues of pneumococcal C-polysaccharide. *Proc. Soc. Exp. Biol. Med.* **136**, 612–614.
- Pontet, M., Engler, R. & Jayle, M.F. (1978). One step preparation of both human C-reactive protein and C1t. *Fed. Eur. Biol. Soc. Lett.* **88**, 172–178.
- Schwalbe, R.A., Dahlback, B., Coe, J.E. & Nelsestuen, G.L. (1992). Pentraxin family of proteins interact specifically with phosphocholine and/or phosphorylethanolamine. *Biochemistry* **31**, 4907–4915.
- Hind, C.R.K., et al., & Pepys, M.B. (1984). Binding specificity of serum amyloid P component for the pyruvate acetal of galactose. *J. Exp. Med.* **159**, 1058–1069.
- Pepys, M.B. (1994). Amyloidosis. In *Samter's Immunological Diseases*. (5th edn) (Frank, M.M., Austen, K.F., Claman, H.N. & Unanue, E.R., eds), pp. 637–655, Little, Brown and Co., Boston.
- Coe, J.E., Margossian, S.S., Slayter, M.S. & Sogn, J.A. (1981). Hamster female protein. A new pentraxin structurally and functionally similar to C-reactive protein and amyloid P component. *J. Exp. Med.* **153**, 977–991.
- Coe, J.E. & Ross, M.J. (1985). Hamster female protein, a sex-limited pentraxin, is a constituent of Syrian hamster amyloid. *J. Clin. Invest.* **76**, 66–74.
- Coe, J.E. & Ross, M.J. (1990). Amyloidosis and female protein in the Syrian hamster. *J. Exp. Med.* **171**, 1257–1267.
- Tennent, G.A., et al., & Pepys, M.B. (1993). Studies of the structure and binding properties of hamster female protein. *Immunology* **80**, 645–651.
- Emsley, J., et al., & Wood, S.P. (1994). Structure of pentameric human serum amyloid P component. *Nature* **367**, 338–345.
- Hardman, K.D. & Ainsworth, C.F. (1972). Structure of concanavalin A at 2.4 Å resolution. *Biochemistry* **11**, 4910–4919.
- Reeke, G.N., Jr, Becker, J.W. & Edelman, G.M. (1975). The covalent and three dimensional structure of concanavalin A. IV. Atomic coordinates, hydrogen bonding and quaternary structure. *J. Biol. Chem.* **250**, 1525–1547.
- Einspahr, H., Parks, E.H., Suguna, K., Subramanian, E. & Suddath, F.L. (1986). The crystal structure of pea lectin at 3.0 Å resolution. *J. Biol. Chem.* **261**, 16518–16527.
- Keitel, T., Simon, O., Borris, R. & Heinemann, U. (1993). Molecular and active-site structure of a *Bacillus* 1,3-1,4-β-glucanase. *Proc. Natl. Acad. Sci. USA* **90**, 5287–5291.
- Torronen, A., Harkki, A. & Rouvinen, J. (1994). Three-dimensional structure of endo-1,4-β-xylanase II from *Trichoderma reesei*: two conformational states in the active site. *EMBO J.* **13**, 2493–2501.
- Divne, C., et al., & Jones, T.A. (1994). The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science* **265**, 524–527.
- Lobsanov, Y.D., Gitt, M.A., Leffler, H., Barondes, S.H. & Rini, J.M. (1993). X-ray crystal structure of the human dimeric S-lac lectin, L-14-II, in complex with lactose at 2.9 Å resolution. *J. Biol. Chem.* **268**, 27034–27038.
- Liao, D.-I., Kapadia, G., Ahmed, H., Vasta, G.R. & Herzberg, O. (1994). Structure of S-lectin, a developmentally regulated vertebrate β-galactoside-binding protein. *Proc. Natl. Acad. Sci. USA* **91**, 1428–1432.
- Blundell, T.L., Sibanda, B.L., Sternberg, M.J.E. & Thornton, J.M. (1987). Knowledge-based prediction of protein structures and the design of novel molecules. *Nature* **326**, 347–352.
- Blundell, T.L., et al., & Sutcliffe, M. (1988). Knowledge-based protein modelling and design. *Eur. J. Biochem.* **172**, 513–520.
- Sutcliffe, M.J., Haneef, I., Carney, D. & Blundell, T.L. (1987). Knowledge based modelling of homologous proteins. Part I. Three-dimensional frameworks derived from the simultaneous superposition of multiple structures. *Protein Eng.* **1**, 377–384.
- Sutcliffe, M.J., Hayes, F.R.F. & Blundell, T.L. (1987). Knowledge based modelling of homologous proteins. Part II. Rules for the conformations of substituted sidechains. *Protein Eng.* **1**, 385–392.
- Srinivasan, N. & Blundell, T.L. (1993). An evaluation of the performance of an automated procedure for comparative modelling of protein tertiary structure. *Protein Eng.* **6**, 501–512.
- Oliveira, E.B., Gotschlich, E.C. & Liu, T.-Y. (1979). Primary structure of human C-reactive protein. *J. Biol. Chem.* **254**, 489–502.
- Lei, K.-J., Liu, T., Zon, G., Soravia, E., Liu, T.-Y. & Goldman, N.D. (1985). Genomic DNA sequence for human C-reactive protein. *J. Biol. Chem.* **260**, 13377–13383.
- Dowton, S.B., Woods, D.E., Mantzouranis, E.C. & Colten, H.R. (1985). Syrian hamster female protein: analysis of female protein primary structure and gene expression. *Science* **228**, 1206–1208.
- Nguyen, N.Y., Suzuki, A., Cheng, S.M., Zon, G. & Liu, T.-Y. (1986). Isolation and characterization of *Limulus* C-reactive protein genes. *J. Biol. Chem.* **261**, 10450–10455.
- Nguyen, N.Y., Suzuki, A., Boykins, R.A. & Liu, T.-Y. (1986). The amino acid sequence of *Limulus* C-reactive protein. Evidence of polymorphism. *J. Biol. Chem.* **261**, 10456–10465.
- Bernstein, F.C., et al., & Tasumi, M. (1977). The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542.
- Wood, S.P., et al., & Pepys, M.B. (1988). A pentameric form of human serum amyloid P component: crystallization, X-ray diffraction and neutron scattering studies. *J. Mol. Biol.* **202**, 169–173.
- Perkins, S.J. & Pepys, M.B. (1986). X-ray and neutron scattering studies on CRP and SAP. In *Protides of the Biological Fluids*. Vol. **34** (Peeters, J., eds), pp. 323–326, Pergamon Press, Oxford.
- Kinoshita, C.M., et al., & Gewurz, H. (1989). Elucidation of a protease-sensitive site involved in the binding of calcium to C-reactive protein. *Biochemistry* **28**, 9840–9848.
- Kinoshita, C.M., et al., & Gewurz, H. (1992). A protease-sensitive site in the proposed Ca²⁺-binding region of human serum amyloid P component and other pentaxins. *Protein Sci.* **1**, 700–709.
- Liu, T.Y., et al., & Goldman, N. (1987). Comparison of protein structure and genomic structure of human, rabbit and *Limulus* C-reactive proteins: possible implications for function and evolution. *J. Protein Chem.* **6**, 263–272.
- Swanson, S.J. & Mortensen, R.F. (1990). Binding and immunological properties of a synthetic peptide corresponding to the phosphorylcholine-binding region of C-reactive protein. *Mol. Immunol.* **27**, 679–687.
- Agrawal, A., Xu, Y., Ansardi, D., Macon, K.J. & Volanakis, J.E. (1992). Probing the phosphocholine-binding site of human C-reactive protein by site-directed mutagenesis. *J. Biol. Chem.* **267**, 25353–25358.

49. Hind, C.R.K., Collins, P.M., Caspi, D., Baltz, M.L. & Pepys, M.B. (1984). Specific chemical dissociation of fibrillar and non-fibrillar components of amyloid deposits. *Lancet* **ii**, 376–378.
50. Topham, C.M., McLeod, A., Eisenmenger, F., Overington, J.P., Johnson, M.S. & Blundell, T.L. (1993). Fragment ranking in modelling of protein-structure. Conformationally constrained environment amino acid substitution tables. *J. Mol. Biol.* **229**, 194–220.
51. Weiner, S.J., *et al.*, & Weiner, P.K. (1984). A new force-field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **106**, 765–784.
52. Sowdhamini, R., Srinivasan, N., Shoichet, B., Santi, D.V., Ramakrishnan, C. & Balaram, P. (1989). Stereochemical modelling of disulfide bridges: criteria for introduction into proteins by site-directed mutagenesis. *Protein Eng.* **3**, 95–103.
53. Evans, S.V. (1993). SETOR: hardware-lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graph.* **11**, 134–138.
54. Jones, T.A. (1978). Graphics model building and refinement system for macromolecules. *J. Appl. Crystallogr.* **11**, 268–272.
55. Overington, J.P., Johnson, M.S., Sali, A. & Blundell, T.L. (1990). Tertiary structural constraints on protein evolutionary diversity: templates, key residues and structure prediction. *Proc. R. Soc. Lond.* **241**, 132–145.
56. Nicholls, A. & Honig, B. (1991). A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson–Boltzmann equation. *J. Comput. Chem.* **12**, 435–445.

Received: 15 Aug 1994; revisions requested: 2 Sep 1994;
revisions received: 13 Sep 1994. Accepted: 14 Sep 1994.