Adaptive diversity of innate immune receptor family short pentraxins in Murinae

Yan Li, Judith H. Robins, Jianping Ye, Zhiyong Huang, Qiang Wen, Guojie Zhang

Abstract

The short pentraxins C-reactive protein (CRP) and serum amyloid P component (SAP) constitute a group of innate immune receptors that trigger immune activation by detecting molecules of the microbial cell wall. Here, we examined the evolution of short pentraxins in Murinae lineages. By molecular evolutionary analysis, CRP and SAP have been experiencing rapid diversification, driven by adaptive selection. Further, our protein modeling demonstrates that adaptively selected amino acids lie in the ligand-binding region and contact region between subunits. Our findings suggest that rapid diversification of these regions could contribute to the determinants of recognizing specificity and the interaction between subunits.

1. Introduction

The innate immune system is the first line of host defense against pathogens and is critical in activating and orienting adaptive immunity. The recognition of pathogens by receptors is the principle aspect of the innate immune system. The innate immune system recognizes pathogens through germline-encoded pattern recognition receptors (PRRs) [1].

Pentraxins are a superfamily of innate immune receptors characterized as having five or ten identical subunits, which are arranged in pentameric symmetry [2]. Based on the length of the subunits, the pentraxins are divided into two groups: long pentraxins and short pentraxins. PTX3 together with PTX4, NP1, NP2, and NPR constitute the long pentraxins. The short pentraxins include two members: CRP and SAP. CRP was the first pentraxin characterized in the 1930s and named for its ability to bind in a Ca2+-dependent manner [2-5]. The crystal structures of human CRP and SAP have been determined, supporting that they are related [2,5]. The subunit (protomer) consists of a two-layered, antiparallel β-sheet and a single, short α-helix with a flat topology. One face, designated A, is characterized by the presence of a furrow. Mutagenesis studies have demonstrated that certain residues in the furrow of CRP make contact with C1q [7-9] and the IgG receptors FcγRI (CD64) and FcγRII (CD32) [10-13]. The other face, designated B, contains a ligand-binding concave face. The ligand-binding site consists of several residues that form a hydrophobic pocket and two Ca2+ ions [14,15]. The contact domain between subunits comprises the open-core end of the two-layered β-sheet of one protomer and the N- and C-terminal strands of another [16].

Short pentraxins recognize ligands and protect against various pathogens, including parasites, fungi, yeasts, and bacteria [17-19]. The direct interaction between CRP and phosphorylcholine (PC) mediates its binding to several microorganisms, including the C-polysaccharide of pneumococcus [20], the lipopolysaccharide (LPS) of Hemophilus influenzae [21], and the repeating phosphorylated disaccharide of Leishmania donovani [22]. Moreover, CRP binds carbohydrates on the surface of Plasmodium falciparum, Hymenolepsis diminuta, and L. donovani in a Ca2+-dependent manner [23-26]. Like CRP, SAP recognizes various microorganisms, such as Streptococcus pyogenes, Neisseria meningitidis, and the influenza virus [27-29]. The ligands for SAP include core polysaccharide moieties of bacterial LPS and phosphoethanolamine (PE) and chemical...
components on lipid A [29–32]. The recognition of ligands by short pentraxins initiates a series of innate immune reactions, such as the activation of the classical complement pathway [33,34] and the stimulation of macrophages [12,13,35–38].

Pentraxins are generally regarded as a superfamily of evolutionarily conserved proteins characterized by their structural and physiological role in organisms. However, many previous studies have suggested that pathogens have evolved a range of anti-immune molecular mechanisms to overcome both innate and adaptive immunity [39]. In the case of the short pentraxins, there are conflicting reports of a possible protective role of CRP toward ligands. In the in vivo plasma clearance experiment, pneumococcal C-polysaccharide (CPS) a ligand to which CRP binds, was not cleared immediately in rats, indeed the half-life of the ligand was rather slower than it was in mice [40]. In another study, passively administered mouse CRP is unable to provide protection against norovirus, which claimed the life of the host [42]; and similarly, human CRP administration is protective in some experiments [41].

2. Materials and methods

2.1. Samples

To provide comprehensive information about the evolutionary patterns and processes of CRP and SAP in Murinae, we collected animals from nine species (Apodemus agrarius, A. draco, M. caroli, M. pahari, Niviventer cremeriventer, R. exulans, R. norvegicus, R. rattus, T. tanzeumi). The sequences of M. musculus were obtained from GenBank (see below). The diversification events of these species have occurred between 12 and less than 0.3 My ago (Fig. S1). The species names and registry numbers are presented in Table S1.

2.2. Sequence and alignment

The entire CRP and SAP coding region of all samples was amplified from genomic DNA using primers in the conserved 5'-UTR and 3' UTR regions. For CRP, the primers were: forward 5'-CTTAG TCCRAGATCCAGCAG-3' and reverse 5'-CTCTA GTGTGAGGACCACA-3'; for SAP, the primers were: forward 5'-GCTGTCGCTCACAAT-3' and reverse 5'-AGGATT TATTTTGGCAGATATG-3'. The PCR products were purified and cloned using the pMD 19-T Vector kit (Takara). The inserts were sequenced using the vector primer. The sequences of a laboratory M. musculus strain (strain C57BL/6) and R. norvegicus (strain Sprague–Dawley), obtained from GenBank were used. Additional sequences of CRP and SAP genes in 32 individuals from five species (n = 10 R. norvegicus, 4 R. tanezumi, 10 A. agrarius, 4 N. cremeriventer, and 4 M. caroli) were determined with the same method, and the resulting sequences were aligned using ClustalW [44] and checked manually. The accession numbers of sequences are presented in Table S1.

2.3. Evolutionary analysis

To examine positive and negative selection within short pentraxins during Murinae evolution, we first analyzed the substitution patterns between pairs of species. Combining the sequences of M. musculus and R. norvegicus, we calculated the ratios of the number of non-synonymous substitutions per non-synonymous site (dN/dS) and the number of synonymous substitutions per synonymous site (dS) for each coding region using Pamilo–Bianchi–Li’s method [45]. The Z-test was used to detect deviations of dN/dS from neutrality.

After detecting adaptive selection in the coding regions, we examined positive selection at individual sites in the data set previously used in evaluating substitution patterns between species pairs (Table S1) with CODEML in the PAML package [46,47]. We implemented three pairs of models: M0 versus M2, M7 versus M8, and M2 and M8 allowed a proportion of sites to be under positive selection with dN/dS > 1. In all tests, a likelihood ratio test (LRT) was used to compare the two models [48,49]. With the same maximum likelihood approach, we then analyzed allelic datasets of CRP and SAP from R. norvegicus and A. agrarius to assess whether positive selection had influenced the variation within species.

2.4. Modeling

Because no crystal structure of murine pentraxins have to date been reported, we simulated their tertiary structure with SWISS-MODEL (Automated Protein Modeling Server http://swissmodel.expasy.org/) [50,51]. The tertiary structure confidence (QMEAN4 Z-score) was assessed using Structure Assessment.

3. Results and discussion

3.1. The short pentraxins share the same gene structures in Murinae lineages

The molecular signature of adaptive selection can be identified through comparative sequence analysis among species. We amplified short pentraxin genes in eight species and cloned the products into vectors. One clone per species was sequenced and used to disclose the evolutionary patterns of the short pentraxin family in Murinae. The sequences had the same gene structures as CRP and SAP in M. musculus and R. norvegicus, based on their Ensembl entries.

3.2. Adaptive selection could have effects on the evolution of short pentraxins in Murinae lineages

Because the continual challenge posed by pathogenic diseases undoubtedly played an essential role in shaping the variation of present-day host immune systems, we analyzed the sequence variations in the entire coding region of CRP and SAP in these Murinae species. Across the CRP and SAP data sets, we observed a large number of non-synonymous substitutions. In the possible pairwise comparisons, the dN/dS ratios of CRP and SAP were calculated. For the CRP sequences, we observed that dN exceeded dS in R. norvegicus versus N. cremeriventer (P = 0.039) and R. exulans versus N. cremeriventer (P = 0.046) significantly. Further, the dN/dS ratios of R. rattus versus N. cremeriventer, R. tanezumi versus N. cremeriventer, M. caroli versus N. cremeriventer, and A. agrarius versus A. draco exceeded 1, which deviated from neutral expectation.

For the SAP sequences, using the Z-test, none of the pairwise comparison ratios exceeded one significantly. However, the ratios of N. cremeriventer to R. norvegicus, R. rattus, and R. tanezumi were similar to that of CRP, for which dN/dS > 1. In addition, the dN/dS ratios between M. pahari and Rattus greatly exceeded 1. These observations suggest that short pentraxins experienced Darwinian positive selection in Murinae lineages.
Because interspecific and intraspecific comparisons could provide information about evolutionary patterns and processes acting over different timescales [52–54], we performed an allelic comparison of DNA sequence polymorphisms. We sequenced the coding region of CRP and SAP of 32 individuals per gene from five species (n = 10 R. norvegicus, 4 R. tanezumi, 10 A. agrarius, 4 N. cremorventer, and 4 M. caroli). Notably, CRP and SAP harbored polymorphisms shared by different species (Fig. S1), with two amino acid-altering polymorphisms in each gene.

There are several possible sources that could account for the existence of interspecific shared polymorphisms, such as maintenance of trans-species polymorphisms, convergent evolution, and persistence of polymorphisms in recently diverged species. It is feasible that the same polymorphisms detected between any two species may have developed by chance when the species were young. Because some of these polymorphisms are present in both species, it is difficult to decide between the role of convergent evolution and trans-species polymorphisms. The long-term maintenance of trans-species polymorphisms could be due to balancing selection, as shown for the major histocompatibility complex genes in jawed vertebrates [57–61]. Convergent change is common at amino acid sites that have been influenced by either positive selection [62–65] or relaxation of negative selection over long periods of time. In order to detect the evolutionary forces in this situation, we analyzed the within-species nucleotide variation in R. norvegicus and A. agrarius (see below).

Adaptive evolution typically occurs at few sites, as most amino acids in a protein are under functional constraints [45]. Using a codon-based test in a maximum likelihood framework [46,47], we narrowed down the individual sites that were under positive selection in CRP and SAP during Murinae evolution. We compared three pairs of models with and without positive selection using likelihood ratio tests, and found that in each comparison, the model that includes sites with dN/dS > 1 fits the data significantly better than the neutral one for both genes (Table 1). Both genes contained several sites for which dN/dS > 1 (Table 1). Six sites in CRP and seven sites in SAP had dN/dS ratios that rejected a neutral expectation in favor of adaptive selection (P < 0.95).

With the allelic datasets of CRP and SAP from R. norvegicus and A. agrarius, we repeated maximum likelihood analyses to assess whether positive selection had influenced the variation within species. For each of the genes, the proportion of sites under selection was relatively low (Table S2). However, three interspecific shared polymorphism sites showed dN/dS > 1 with a posterior probability of 0.8. These results suggested that natural selection played an important role in shaping these polymorphisms.

It is notable that some of the indicated amino acid changes at these sites are radical in terms of their striking physicochemical properties (Table 2). We used Grantham's distance, which relies on size, polarity, and carbon composition of amino acid side chains, to measure the underlying physicochemical differences between the exchanging residues [66]. Based on Grantham's distance, an amino acid substitution can be classified as either conservative or radical. Several amino acid changes at these sites were radical, strengthening the argument for a model of adaptive selection. For the M. musculus/R. norvegicus comparison, there were 4 putative selected sites in CRP that were different, of which the change at site 32 was radical with regard to its physicochemistry (Lys32 in

### Table 1
Likelihood values and sites inferred to be under positive selection for short pentraxins genes across Murinae lineages.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Model</th>
<th>I</th>
<th>-2AL</th>
<th>Sites under selection identified by PAML</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRP</td>
<td>M0: one-ratio</td>
<td>-2225.62</td>
<td>41.70** (M2 versus M0)</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>M2: selection</td>
<td>-2204.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M7: beta</td>
<td>-2209.90</td>
<td>16.24** (M8 versus M7)</td>
<td>Not allowed</td>
</tr>
<tr>
<td></td>
<td>M8a: beta &amp; fixed ω</td>
<td>-2209.86</td>
<td>10.16** (M8 versus M8a)</td>
<td>Not allowed</td>
</tr>
<tr>
<td></td>
<td>M8: beta &amp; ω</td>
<td>-2204.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAP</td>
<td>M0: one-ratio</td>
<td>-2005.76</td>
<td>70.40** (M2 versus M0)</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>M2: selection</td>
<td>-1970.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M7: beta</td>
<td>-1983.83</td>
<td>26.46** (M8 versus M7)</td>
<td>Not allowed</td>
</tr>
<tr>
<td></td>
<td>M8a: beta &amp; fixed ω</td>
<td>-1983.78</td>
<td>26.36** (M8 versus M8a)</td>
<td>Not allowed</td>
</tr>
<tr>
<td></td>
<td>M8: beta &amp; ω</td>
<td>-1970.60</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. -2AL, twice the log likelihood difference between pair of models. Double asterisks correspond to P < 0.01 (one-tailed t-test). The numbers of amino acid sites represent their position in the interspecies sequences alignment.

### Table 2
The amino acid substitutions at the adaptively selected sites of CRP and SAP across Murinae lineages.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Position</th>
<th>aa 1</th>
<th>aa 2</th>
<th>aa 3</th>
<th>aa 4</th>
<th>aa 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRP</td>
<td>32</td>
<td>Gly</td>
<td>Glu</td>
<td>Lys</td>
<td>Thr</td>
<td></td>
</tr>
<tr>
<td></td>
<td>65</td>
<td>Val</td>
<td>Ile</td>
<td>Ala</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>76</td>
<td>Lys</td>
<td>Thr</td>
<td>Glu</td>
<td>Asn</td>
<td></td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>Gin</td>
<td>Glu</td>
<td>Lys</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>213</td>
<td>Thr</td>
<td>Ala</td>
<td>Lys</td>
<td>Val</td>
<td></td>
</tr>
<tr>
<td></td>
<td>214</td>
<td>His</td>
<td>Gln</td>
<td>Arg</td>
<td>Trp</td>
<td></td>
</tr>
<tr>
<td>SAP</td>
<td>25</td>
<td>Asn</td>
<td>Asp</td>
<td>Lys</td>
<td>Tyr</td>
<td></td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>Gln</td>
<td>Trp</td>
<td>Lys</td>
<td>Arg</td>
<td>Ser</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>Trp</td>
<td>His</td>
<td>Arg</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>105</td>
<td>Leu</td>
<td>Lys</td>
<td>Arg</td>
<td>Met</td>
<td>Ile</td>
</tr>
<tr>
<td></td>
<td>108</td>
<td>Phe</td>
<td>Tyr</td>
<td>Pro</td>
<td>His</td>
<td></td>
</tr>
<tr>
<td></td>
<td>159</td>
<td>Thr</td>
<td>Ser</td>
<td>Ala</td>
<td>Asn</td>
<td></td>
</tr>
<tr>
<td></td>
<td>198</td>
<td>Pro</td>
<td>Leu</td>
<td>Val</td>
<td>Ile</td>
<td></td>
</tr>
</tbody>
</table>

References:

[52–54]: Coding region of CRP and SAP of 32 individuals per gene from five species (n = 10 R. norvegicus, 4 R. tanezumi, 10 A. agrarius, 4 N. cremorventer, and 4 M. caroli).

[55, 56]: Strongly suggesting that the occurrence of such polymorphisms in these two genes is not due to an exceptionally young age of the species. Because of the long time since divergence of these species from their common ancestor, it is difficult to decide between the role of convergent evolution and trans-species polymorphisms.
Similarly, all seven selected sites in SAP varied between *M. musculus* and *R. norvegicus*, of which the substitution at site 45 (45His in *M. musculus* and 45Trp in *R. norvegicus*, Grantham’s distance: 115) was radical. Overall, these evolutionary changes in CRP and SAP might have contributed to the functional divergence of these innate immune receptors in murine species.

### 3.3 Sites under adaptive selection are located on the ligand-binding and subunit interaction region of the short pentraxins

To gain insight into the structural and functional significance of the sites with evidence of selection, we located them on the three-dimensional structures of CRP and SAP. The Murinae short pentraxins shared the greatest similarity with human CRP (PDB ID: 1b09C) and SAP (PDB ID: 3kqrB) proteins. Pairwise alignments of targets against their templates showed that they share high sequence identity (greater than 60%) (Fig. S2), which confirmed that the determined structures of the templates could be used for modeling. The reliability of modeling was assessed, and the resulting QMEAN4 Z-scores (−1.22 and −0.83, respectively) ensured that the predicted structures were valid [67].

On mapping the putatively selected sites onto the structural models of CRP and SAP (Fig. 1), we observed that they fall in disparate functional domains. One group of sites (65, 76, and 90) is located on the side walls of the concave face for ligand-binding and run along the surface of CRP [16]. The position of PC on the LPS affects the susceptibility to CRP in *H. influenzae* [68]. In addition to PC, short pentraxins bind to a wide variety of carbohydrates on microbial surfaces [19]. Thus, the distribution of variable sites in the side walls of the concave face may help to diversify the pathogenic selectivity of short pentraxins, allowing them to recognize disparate ligands. In a comparison of the CPR sequences, we found that the residues at positions 65 and 90 in *M. musculus* differed from those in *R. norvegicus* (Fig. S3), which could contribute to the different plasma clearance rate of CPS in these species [40]. Further, comparing the CRP sequences of *M. musculus*, *Homo sapiens*, and *O. cuniculus*, we found that *O. cuniculus* and *H. sapiens* share the same residues at position 65 and 90, which are different from that

---

**Fig. 1.** Stick model showing the locations of adaptively selected amino acid sites in short pentraxins. The positions of adaptively selected amino acids are shown on the predicted three-dimensional structures of *R. norvegicus* CRP (A and B) and SAP (C and D) pentamers in red. (A and C): bottom view of the concave face for ligand-binding and nearby selected residues (65, 76, and 90 of CRP in (A); 105, 108, and 159 of SAP in (C)). (B and D): side view of the selected amino acids. In each case, the surfaces that associate to form a pentamer and contain sites under selection are shown in the squares. There are 21 amino acids missing in CRP and 20 amino acids missing in SAP compared with the template.
of M. musculus (Fig. 53). The differences among these residues are consistent with the phenomenon that the CRPs from M. musculus and the two other species provide the innate immune system's distinct specificity to LPS [29,41–43].

Another group of sites: 23, 32, 142, 213, and 214 in CRP and sites 25, 26, 45, 198, and 207 in SAP, lie in the interaction region of the protomer (Fig. 1B and D). In particular, sites 142 and 214 of Murine CRP corresponded to sites 123 and 197 in the human ortholog, which form an intermolecular ion pair [16]. Two sites adopt different strategies: the shared polymorphisms (Lys and Glu) in site 142 presented striking physicochemical properties, and rapid changes at site 214 overlapped basic, non-polar, and uncharged polar residues. Interestingly, with regard to site 14 in the N-terminus of SAP, all four interspecific shared polymorphism sites fell in the interaction region (Fig. S1 and Fig. 1). In addition, amphipathic amino acids appeared frequently at these sites. One potential explanation for this is that the coexistence of polymorphisms and the frequent presence of amphipathic amino acids provide greater physicochemical variability for interacting residues in the adjacent protomer. Because the structural organization of protomers could play an important role in the activation of complement pathway [16,69], it is possible that such conversions constitute a functional adaptation that counters microbial insults that block the interaction between protomers [39]. Future functional studies are needed to determine the effects of these amino acid variations. If the hypothesis is supported, it might reveal more insights into the mechanism of the arms race between hosts and pathogens, as well as the coevolution of proteins and their endogenous interaction partners.

The above analysis provides evidence that adaptive selection drove the diversification at a number of sites, which were distributed on the different functional domains of short pentraxins. Understanding the evolution of these immune receptors in Murinae could provide important insights regarding the physiological function of these receptors and the mechanisms of the arms race between host immune systems and diverse pathogens.

Acknowledgments

We thank Scott Edward for his helpful comments and English editing of the manuscript. We thank Yu-Teh Kirk Lin for providing the M. caroli samples; Annie Orth for providing the M. pahari sample; Vernon Tintinger for providing the R. rattus and R. tanezumi DNA samples; and Huailiang Xu and Yongfang Yao for providing the equipment. This work was supported by R. tanezumi and the frequent presence of amphipathic amino acids provide greater physicochemical variability for interacting residues in the adjacent protomer. Because the structural organization of protomers could play an important role in the activation of complement pathway [16,69], it is possible that such conversions constitute a functional adaptation that counters microbial insults that block the interaction between protomers [39]. Future functional studies are needed to determine the effects of these amino acid variations. If the hypothesis is supported, it might reveal more insights into the mechanism of the arms race between hosts and pathogens, as well as the coevolution of proteins and their endogenous interaction partners.

The above analysis provides evidence that adaptive selection drove the diversification at a number of sites, which were distributed on the different functional domains of short pentraxins. Understanding the evolution of these immune receptors in Murinae could provide important insights regarding the physiological function of these receptors and the mechanisms of the arms race between host immune systems and diverse pathogens.

Acknowledgments

We thank Scott Edward for his helpful comments and English editing of the manuscript. We thank Yu-Teh Kirk Lin for providing the M. caroli samples; Annie Orth for providing the M. pahari sample; Vernon Tintinger for providing the R. rattus and R. tanezumi DNA samples; and Huailiang Xu and Yongfang Yao for providing the equipment. This work was supported by Scientific Research Fund of Sichuan Provincial Education Department (Grant No. 102A058 to Y.L.).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febsl.2012.01.048.

References
