Appl Biochem Biotechnol (2011) 165:497–505
DOI 10.1007/s12010-011-9270-7

Conserved Domains, Conserved Residues, and Surface Cavities of C-reactive Protein (CRP)

Suggula Varun Kumar · Roshini Kulampurathu Ravunny · Chiranjib Chakraborty

Received: 9 January 2011 / Accepted: 18 April 2011 /
Published online: 4 May 2011

Abstract C-reactive protein (CRP) is a highly conserved plasma protein belonging to pentraxins, a superfamily which has significant proinflammatory role. Therefore, CRP can be a good target for drug discovery to prevent disease pathogenesis, especially cardioprotection in acute myocardial infarction and neuroprotection in stroke. Hence, the knowledge of the structure of CRP is very important. In this paper, we have demonstrated three-dimensional structure, conserved domains, and surface structure of human CRP with the help of bioinformatics analysis. We have also depicted that evolutionary relationship of CRP exists among the different species. Simultaneously, WebLogo has been generated to know the more conserved part of this medically important protein.

Keywords Conserved domains · Conserved residues · Surface cavities · C-reactive protein (CRP)

Introduction

C-reactive protein (CRP) is a plasma protein which rises noticeably in a cytokine-mediated reaction during inflammation, infection as well as tissue injury. Clinicians are generally measures CRP values to know the diseases activity which is a clinical practice [1, 2]. In healthy humans, the CRP concentration in is not more than 1 μg/ml. Conversely, with the help of stimulus, it can increase 1,000-fold following an acute phase [3, 4]. However, the values have a tendency to increase with the increase of age, and it has been noted that the value is higher in women than in men. High levels of CRP are prognostic for stroke and

All authors contributed equally to this article.

S. V. Kumar · R. K. Ravunny · C. Chakraborty (✉)
Medical Biotechnology Division, School of Bio-Sciences and Technology, VIT University, Vellore, India
e-mail: drchiranjib@yahoo.com

related to coronary functions [4, 5], as well as associated with the patients with end-stage renal disease and cardiovascular mortality [6].

On the other hand, CRP has an important physiological function and has been noted to have been associated with autoimmunity, especially with antinuclear autoantibody production and with systemic lupus erythematosus [7]. It has also been hypothesized by several scientists that CRP may have significant proinflammatory role (Fig. 1) [8, 9]. So, concentration of this circulating protein is not only a diagnostic marker of disease severity and progression but also contributes to pathogenesis [1, 10]. However, there is possibility that we can assess the hypothesis by developing drugs that specially block CRP binding and will inhibit the proinflammatory effects [11]. Again, CRP is related to serum amyloid P component (SAP) [12]. The SAP is an important ingredient of the amyloid deposits that cause a range of human diseases like Alzheimer's diseases [13, 14]. Interestingly, CRP and amyloid P are associated with many diseases, like the amyloid deposits and neurofibrillary tangles of Alzheimer disease [11]. So, presently, CRP may be a good target for drug discovery to prevent those diseases pathogenesis. Therefore, the information on the structure of CRP, including its three-dimensional structure, conserved domains, and surface structure, is more important to design the novel inhibitors to platform for drug design.

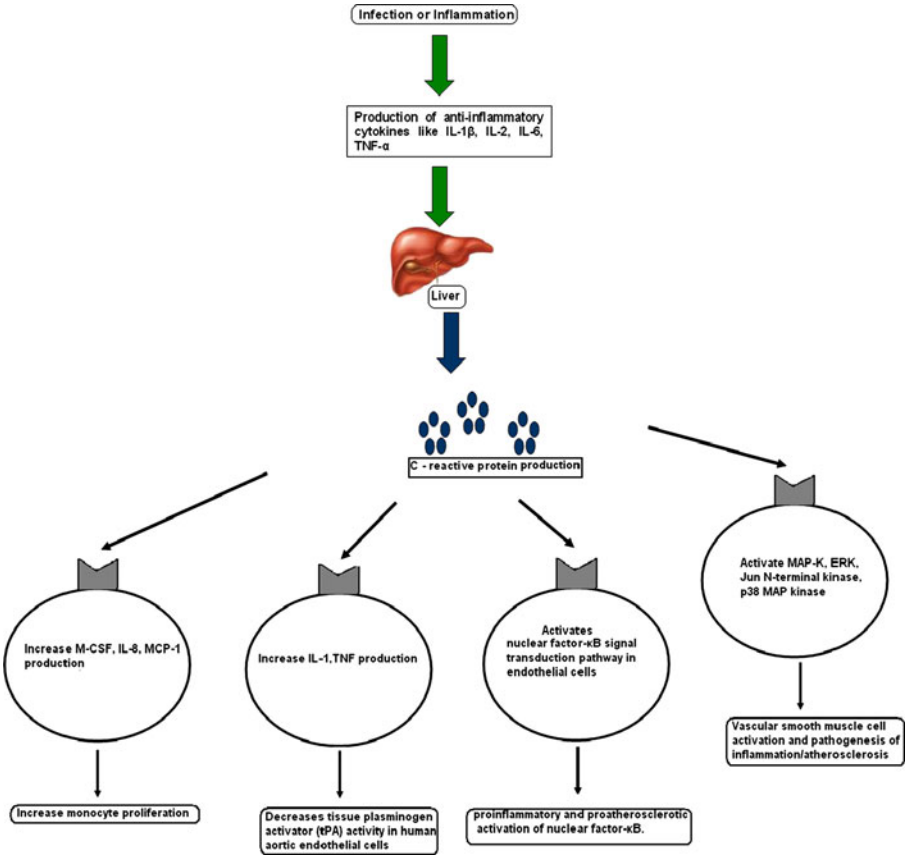

Fig. 1 C-reactive protein (CRP) and its in proinflammatory or inflammatory role

In this study, we have addressed the three-dimensional structure, conserved domains, and surface structure with of human CRP with the help of bioinformatics analysis. We have also found that a developed evolutionary relationship of CRP exists between the different species. Again, WebLogo has been produced to know the more conserved part of this CRP from humans as well as other species.

Materials and Methods

Data Collection

We have collected data on the genes related to the proteins of CRP human as well as different species from the National Center for Biotechnology Information (NCBI) [15]. The functional protein sequences (in FASTA format) were collected with the accession number from the NCBI database and further analyzed. The collected sequence are human (CAA39671.1), African clawed frog (NP_001165686.1), rabbit (AAA75404.1), mouse (CAA31928.1), pig (AAU13779.1), guinea pig (AAC60662.1), golden hamster (AAB19893.2), rat (AAA40964.1), chicken (ABD16281.1), turkey (ABV03362.1), Atlantic horseshoe crab (AAA28270.1), rhesus monkey (XP_001117250.1), chimpanzee (XP_001170785.1), cattle (BAH66497.1), orange spotted groper fish (ADC92292.1), and Chinese horseshoe crab (BAA85653.1).

Conservation Patterns of Structures and Calculation of Highly Conserved Amino Acids in CRP

The conservation patterns of structures in CRP were formed using ConSurf server [16, 17]. The conservation scores at each amino acid position were calculated using the same server. It also calculated the evolutionary conservation of amino acid positions in proteins using an empirical Bayesian inference, starting from protein structure and sequence, respectively. Highly conserved amino acids from proteins were used for further analysis.

Generation of Surface Cavity and Binding Groove Identification

The PyMOL [18] has been used for the generation of surface cavity as well as identification of binding grooves of CRP. We have used ".pdb" files to generate the surface structure and the cavities of those proteins.

Phylogenetic Tree Construction

Based on the results of the sequence of different species of CRP, a phylogenetic tree was constructed using Phylogeny.fr, which is a free, simple-to-use web service [19] software (http://www.phylogeny.fr/version2_cgi/simple_phylogeny.cgi). We have developed two types of phylogenetic tree, i.e., phylogram. The phylogram tree shows the distances between the protein sequences of the factors.

Sequence Logo Formation of Conserved Domains of CRP

Sequence logos were formed using WebLogo software, which is used for graphical representation of amino acid or nucleic acid for displaying the patterns in a set of

aligned sequences [20, 21]. Here, we have used all sequences to generate the WebLogo. Again, we have used ten amino acids each time to see the conserved part of human and other species. This method was used to visualize patterns of aligned sequences as well as the bias amino acid sequences of CRP.

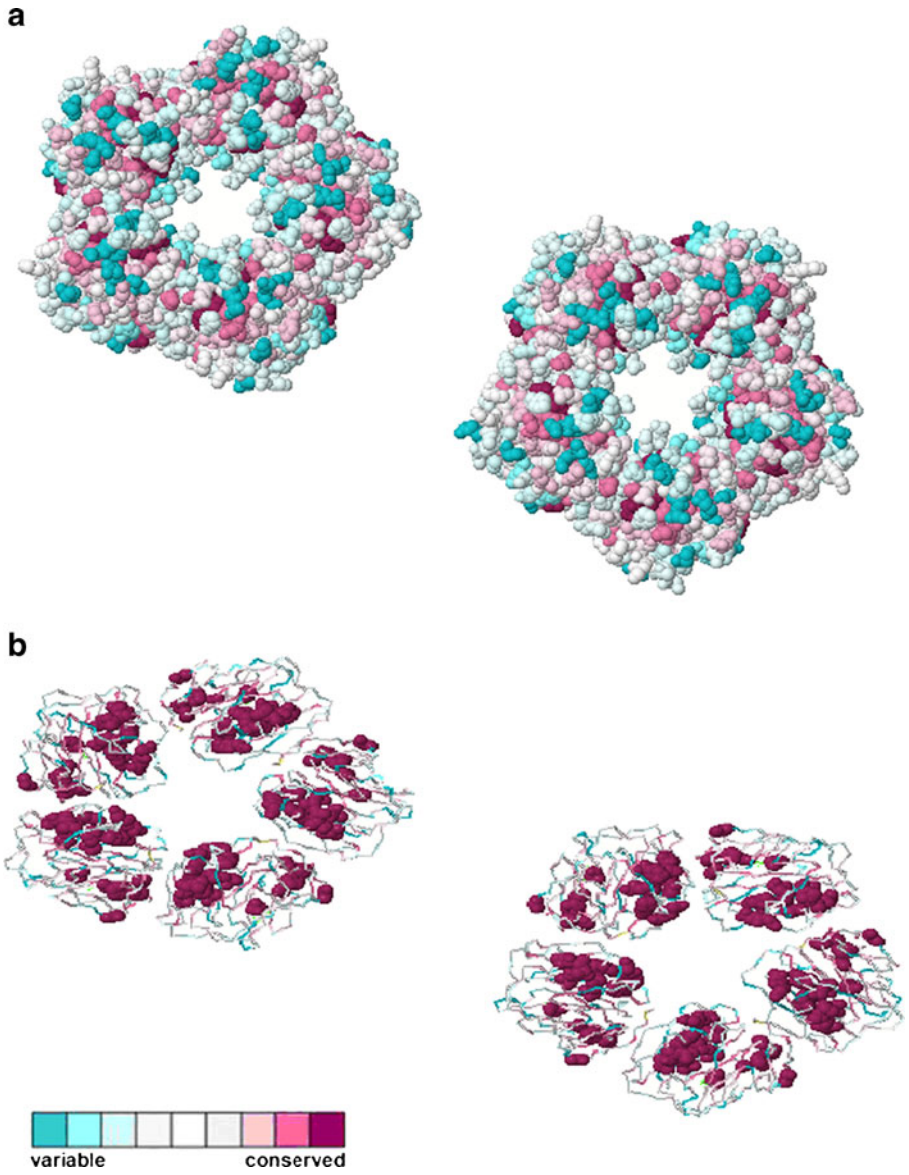

Fig. 2 Conservation patterns and backbone structures of the pentameric structure of CRP. **a** This figure shows the general conservation patterns with highly conserved amino acids in 3D structure of the CRP. Amino acid conservation scores were classified into nine levels. The color scale for residue conservation is indicated in the figure. **b** Backbone structures with highly conserved amino acids of CRP

Results and Discussion

CRP proteins of humans, as well as other species, and their genes were recorded using original data from the NCBI data bank. The conservation patterns of CRP and their backbone structures have been shown in Fig. 2. Nonetheless, we were able to document highly conserved amino acids of the protein which are SER53, HIS95, CYS97, ASP112, GLY113, GLY136, GLY154, VAL165, LEU166, ILE171, and GLY196. Surface cavity and binding grooves has been provided in Fig. 3 which show that there are several surface cavities present on CRP, and these can be the ligand-binding pocket. It has been reported that CRP is a highly evolutionary conserved protein in the human as well as other species. Therefore, we have developed a relationship of CRP in human as well as other species which have been shown in Fig. 4. The phylogenetic tree shows that CRPs of human and

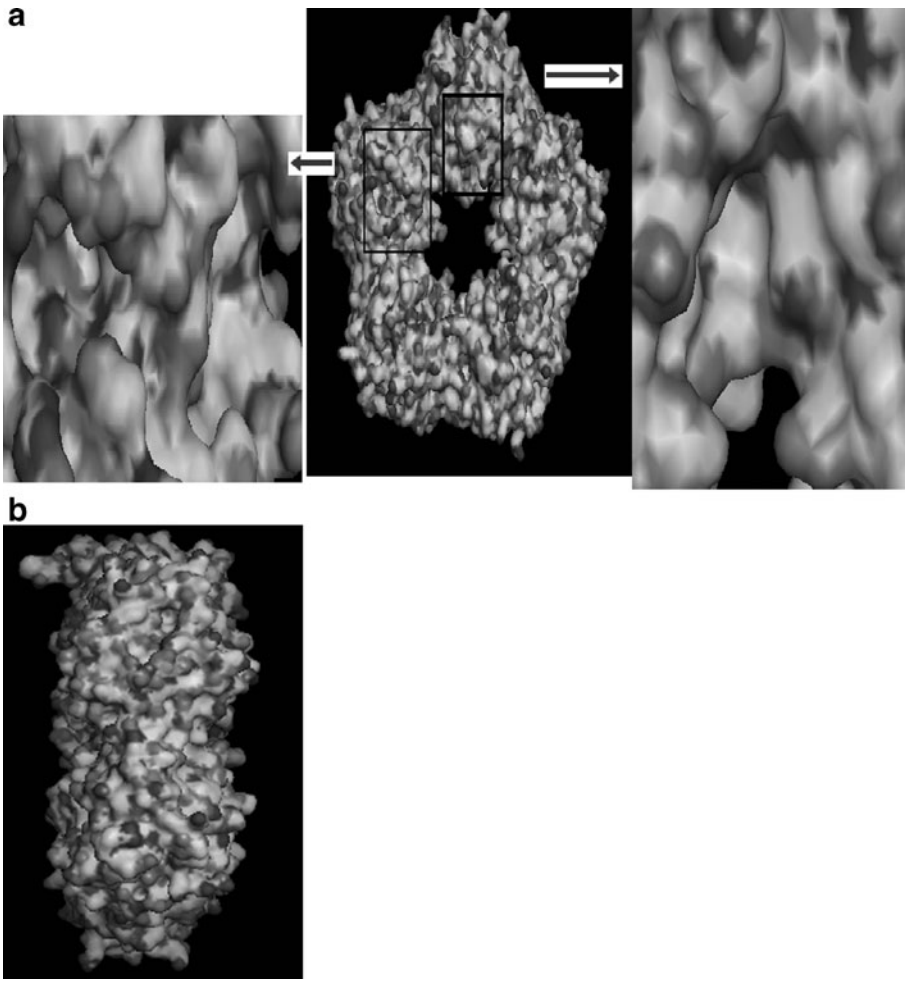

Fig. 3 Surface cavity and binding grooves of the CRP. **a** Front view shows some specific binding pocket/binding grooves. **b** Side view of CRP shows different grooves

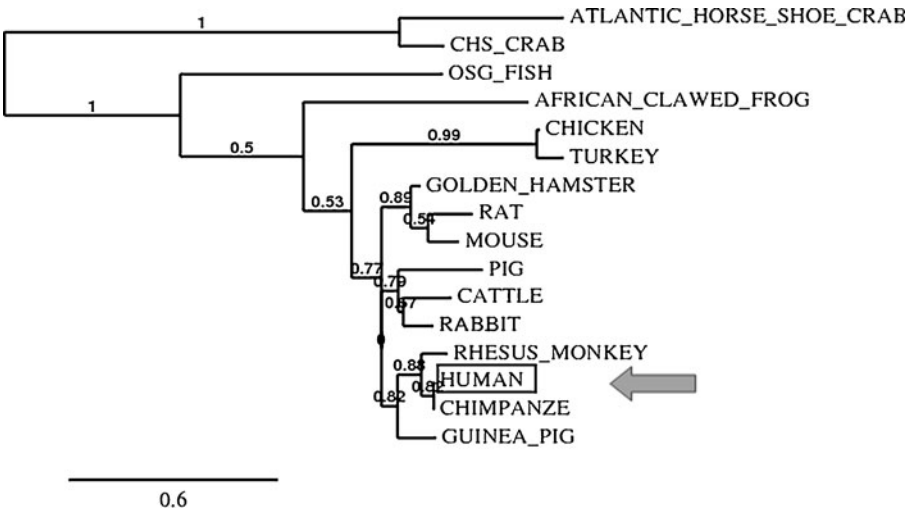

Fig. 4 Phylogenetic tree construction of CRP using Phylogeny.fr software. It shows the relationship between the human CRP and the CRP of other species (horseshoe crab, African clawed frog, rabbit, mouse, pig, guinea pig, golden hamster, rat, chicken, turkey, Atlantic horseshoe crab, rhesus monkey, chimpanzee, cattle, orange spotted groper fish, Chinese horseshoe crab)

chimpanzee are very closely related. To know the conserved domains, we have generated WebLogo (Fig. 5) which shows that this protein from all species is a highly conserved one. We have noted the first ten amino acids are more conserved and that maximum stack height is 2.8 bits and the minimum stack height, 1 bit in this part. No stack has been noted on 24th position WebLogo.

The name of "C-reactive protein" has been given for its ability to precipitate the "C" polysaccharide mainly taken out from the pneumococcal cell wall [22]. This circulating protein seems to have an important physiological role. The physiological role of this protein is unknown because of its structural polymorphism in human CRP, and experimental CRP knockout has not yet been reported. This protein is produced by hepatocytes in response to inflammatory cytokines like interleukin-1 and interleukin-6 [3, 4]. It is formed by the liver as part of the "reorchestration" of hepatic gene expression in response to inflammation and infection [22]. The human CRP gene present in chromosome-1 located between 1q21and 1q23 [23, 24]. The half-life of this protein is relatively long in plasma (19 h) which is quite stable in vitro [25].

CRP is an evolutionary conserved protein and a member of the evolutionarily ancient conserved pentraxin family. Pentraxins are a superfamily which contains multifunctional conserved proteins [26]. This family also contains SAP [12]. Volanakis and Kaplan [27] have demonstrated that the conservation of the structure of CRP and of its calcium-dependent specific binding of ligands containing phosphocholine. This complex may hinder detection of any polymorphism or deficiency of CRP in man [28]. However, in our case, several highly conserved residues/domains have been noted. Conserved domains are distinct units of three-dimensional (3D) structure as well as some distinct functional role [29, 30]. Alternatively, homologous CRP-like pentameric proteins (pentraxin group members) are found in several vertebrates, as well as invertebrates [31]. Therefore, it is a primitive molecule of an innate host-immune protection strategy dependent upon the ability to opsonize pathogenic ligands [32].

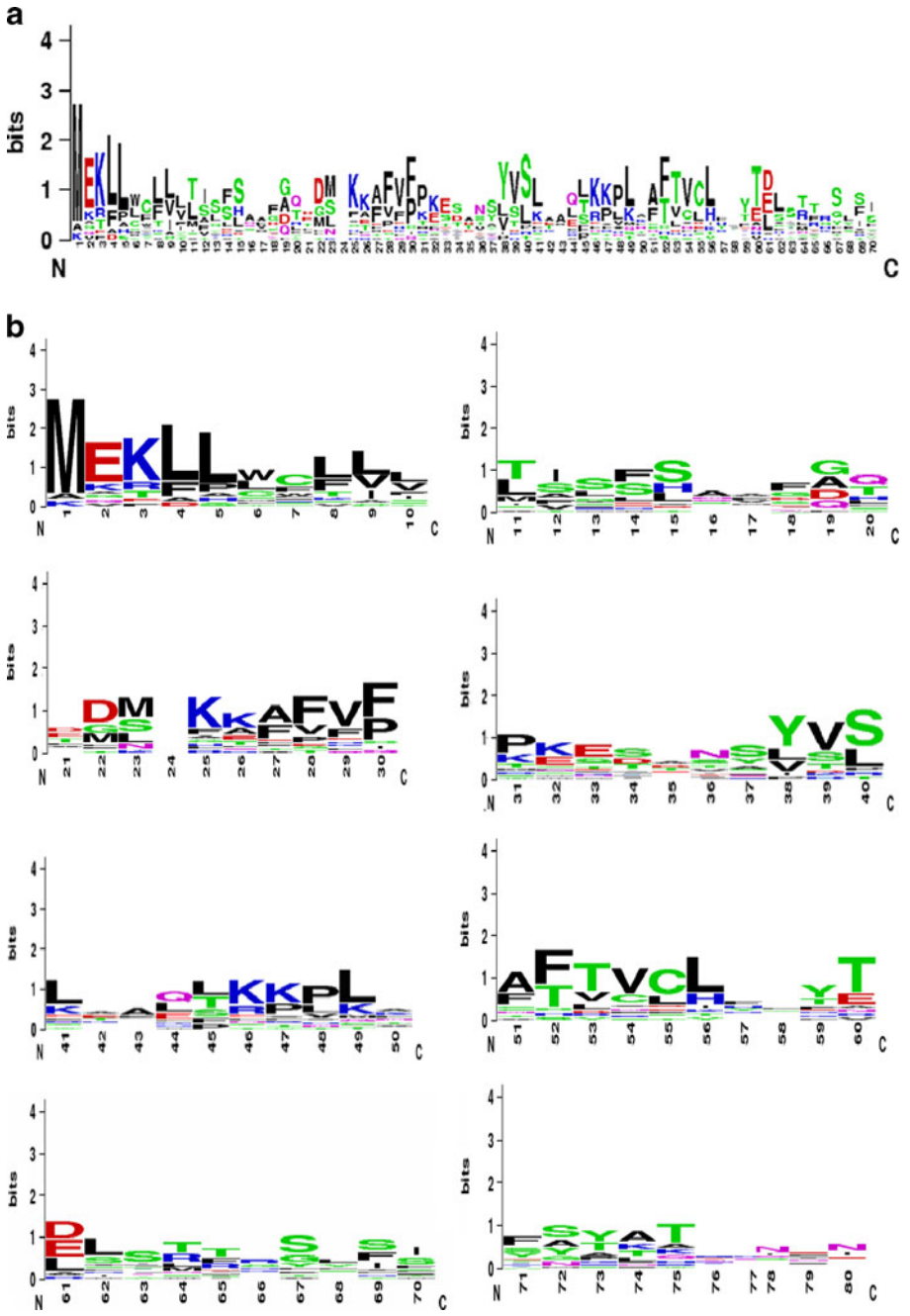

Fig. 5 Generated WebLogo for the CRP and related species. **a** WebLogo has been formed with all the species including human. **b** WebLogo has been formed comparing ten amino acids of humans with other species

CRP is having five identical, noncovalently associated protomers (approximately 23 kDa). These protomers arranged symmetrically around a central pore [22]. It has been

reported that based on the crystal structure of SAP [33], the CRP protomer is very similar to that of SAP. The subunits consist of a two-layered β sheet with a flattened jellyroll topology. The crystal structure of CRP with phosphocholine has been determined [28], and the structure suggests that Phe-66 and Glu-81 are the two key residues mediating the binding of phosphocholine to CRP. The Cα rms fit is 1.3 Å, but it drops to 0.83 Å when three divergent loops involving 19 residues are omitted. The sidewalls of this protein are assembled by the amino acids like Ser5, Arg6, Gln203, Pro206, Trp187, Arg188, Asn160, Gly177, Leu176, Tyr175, His95, and Asp112. The bottom is composed with Asn158, His38, Leu37, Val94, and Asp112 [28]. Short helical regions are formed around residues 43 and 185.

In the complement system, the first component, C1, consists of three distinct proteins named C1q, C1r, and C1s [34]. The CRP binds with C1q in the opposite face, and it is described as the effector face of the pentamer. It is not only the C1q but also immunoglobulin receptors (FcγR) that bind with the effector face. The binding groove extends from the center of the protomer along the boundaries, and this cleft has been shown to be critical for the binding of CRP to C1q [35, 36]. A model has been proposed for C1q binding to CRP. In this model, it has been described that the top of the positively charged C1q head interacts with negatively charged central pore of the CRP pentamer. The optimal C1q binding is related to slight conformational changes in the CRP structure [37], and these structural changes appear to differ depending on the ligand to which CRP is bound [38].

There are two calcium ions bound 4-Å apart by protein side chains coming from loops collected at the concave face [39], and this is the site for ligand binding. A small-molecule inhibitor like 1,6-bis(phosphocholine)-hexane has recently been designed and synthesized for of CRP, and five molecules of this palindromic compound have been bound by two pentameric CRP molecules. The cross-linking of the molecule with the compound/ligand occurs at the ligand-binding B-face of CRP which blocks its functions [40]. Therefore, it has been noted that the therapeutic inhibition of CRP is a novel method to cardioprotection in acute myocardial infarction and can also provide neuroprotection in stroke.

Presently, novel therapeutic development targeting CRP is a good approach for the cardioprotection and neuroprotection. Our three-dimensional structure, conserved domain identification, and surface cavities/binding pocket identification will be helpful in developing many more inhibitors which can give us a better and economical therapeutic approach. Furthermore, the structure of CRP as well its inhibitors may provide more information about the physiological roles of human CRP.

Acknowledgment We would like to thank the senior management of VIT University (Vellore, India) for their kind support and encouragement for our research work.

References

1. Pepys, M. B., & Baltz, M. L. (1983). *Advances in Immunology, 34*, 141–212.
2. Pepys, M. B. (1995). In D. J. Weatherall, J. G. G. Ledingham, & D. A. Warrell (Eds.), *Oxford textbook of medicine* (3rd ed., pp. 1527–1533). Oxford: Oxford University Press.
3. Garlanda, C., Bottazzi, B., Bastone, A., & Mantovani, A. (2005). *Ann Rev Immunol, 23*, 337–366.
4. Du Clos, T. W., & Mold, C. (2004). *Immunologic Research, 30*, 261–277.
5. Hirshfeld, G. M., & Pepys, M. B. (2003). *The Quarterly Journal of Medicine, 96*, 793–807.
6. Arici, M., & Walls, J. (2001). *Kidney International, 59*, 407–414.
7. Pepys, M. B., Lanham, J. G., & De Beer, F. C. (1982). *Clinics in Rheumatic Diseases, 8*, 91–103.
8. Pasceri, V., Willerson, J. T., & Yeh, E. T. (2000). *Circulation, 102*, 2165–2168.
9. Wadham, C., Albanese, N., Roberts, J., Wang, L., Bagley, C. J., Gamble, J. R., et al. (2004). *Circulation, 109*, 2116–2122.

10. Pepys, M. B., & Hirschfield, G. M. (2003). *Journal of Clinical Investigation, 111*, 1805–1812.
11. Yasojima, K., Schwab, C., McGeer, E. G., & McGeer, P. L. (2000). *Brain Research, 887*, 80–89.
12. Pepys, M. B., Dash, A. C., Fletcher, T. C., Richardson, N., Munn, E. A., & Feinstein, A. (1978). *Nature, 273*, 168–170.
13. Pepys, M. B. (1994). In M. M. Frank, K. F. Austen, H. N. Claman, & E. R. Unanue (Eds.), *Samter's immunologic diseases* (5th ed., pp. 637–655). Boston: Little, Brown and Company.
14. Pepys, M. B., Booth, D. R., Hutchinson, W. L., Gallimore, J. R., Collins, P. M., & Hohenester, E. (1997). *Amyloid Int J Exp Clin Invest, 4*, 274–295.
15. Wheeler, D. L., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., et al. (2007). *Nucleic Acids Research, 35*, D5–D12.
16. Ashkenazy, H., Erez, E., Martz, E., Pupko, T., & Ben-Tal, N. (2010). *Nucleic Acids Research, 38*, W529–W533.
17. Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E., et al. (2003). *Bioinformatics, 19*, 163–164.
18. DeLano WL (2002) *The PyMOL molecular graphics system*. Available from: http://www.pymol.org.
19. Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., et al. (2008). *Nucleic Acids Research, 36*, W465–W469.
20. Schneider, T. D., & Stephens, R. M. (1990). *Nucleic Acids Research, 18*, 6097–6100.
21. Crooks, G. E., Hon, G., Chandonia, J. M., & Brenner, S. E. (2004). *Genome Research, 14*, 1188–1190.
22. Black, S., Kushner, I., & Samols, D. (2004). *The Journal of Biological Chemistry, 279*, 48487–48490.
23. Sinico, R. A., Radice, A., Ikehata, M., Giammarresi, G., Corace, C., Arrigo, G., et al. (2005). *Annals of the New York Academy of Sciences, 1050*, 193–200.
24. Marto, N., Bertolaccini, M. L., Calabuig, E., Hughes, G. R., & Khamashta, M. A. (2005). *Annals of the Rheumatic Diseases, 64*, 444–448.
25. Aziz, N., Fahey, J. L., Detels, R., & Butch, A. W. (2003). *Clinical and Diagnostic Laboratory Immunology, 10*, 652–657.
26. Martinez de la Torre, Y., Fabbri, M., Jaillon, S., Bastone, A., Nebuloni, M., Vecchi, A., et al. (2010). *Journal of Immunology, 184*, 5055–5064.
27. Volanakis, J. E., & Kaplan, M. H. (1971). *Proceedings of the Society for Experimental Biology and Medicine, 136*, 612–614.
28. Thompson, D., Pepys, M. B., & Wood, S. P. (1999). *Structure, 7*, 169–177.
29. Marchler-Bauer, A., Anderson, J. B., Cherukuri, P. F., DeWeese-Scott, C., Geer, L. Y., Gwadz, M., et al. (2005). *Nucleic Acids Research, 33*, D192–D196.
30. Branden, C., & Tooze, J. (1999). *Introduction to Protein Structure* (2nd ed.). New York: Garland Publication.
31. Magor, B. G., & Magor, K. E. (2001). *Developmental and Comparative Immunology, 25*, 651–682.
32. Tharia, H. A., Shrive, A. K., Mills, J. D., Arme, C., Williams, G. T., & Greenhough, T. J. (2002). *Journal of Molecular Biology, 316*, 583–597.
33. Srinivasan, N., White, H. E., Emsley, J., Wood, S. P., Pepys, M. B., & Blundell, T. L. (1994). *Structure, 2*, 1017–1027.
34. Svehag, S. E., Manhem, L., & Bloth, B. (1972). *Nature, 238*, 117–118.
35. Agrawal, A., & Volanakis, J. E. (1994). *Journal of Immunology, 152*, 5404–5410.
36. Agrawal, A., Shrive, A. K., Greenhough, T. J., & Volanakis, J. E. (2001). *Journal of Immunology, 166*, 3998–4004.
37. Gaboriaud, C., Juanhuix, J., Gruez, A., Lacroix, M., Darnault, C., Pignol, D., et al. (2003). *The Journal of Biological Chemistry, 278*, 46974–46982.
38. Black, S., Agrawal, A., & Samols, D. (2003). *Molecular Immunology, 39*, 1045–1054.
39. Ashton, A. W., Boehm, M. K., Gallimore, J. R., Pepys, M. B., & Perkins, S. J. (1997). *Journal of Molecular Biology, 272*, 408–422.
40. Pepys, M. B., Hirschfield, G. M., Tennent, G. A., Gallimore, J. R., Kahan, M. C., Bellotti, V., et al. (2006). *Nature, 440*, 1217–1221.