Turun yliopisto
University of Turku

Structural modeling and analysis of non-mammalian CA VI

Abdullah Al-Farabi
Master's Thesis of M.Sc. Bioinformatics
Master's Programme in Digital Health and Life Science
Department of Future Technologies
University of Turku
Finland

Acknowledgements

First of all, thanks to Almighty ALLAH for giving me the strength and ability to understand, learn and complete this report.

I would also like to express my deepest thanks to Dr. Martti Tolvanen for his constant guidance and support for completing this project work and being a supervisor for this thesis. This thesis would not have been possible without his timely suggestions despite his busy schedule.

Finally, I would like to acknowledge everyone who played a role in my academic accomplishment and my parents, who supported me with love and understanding. Without you, I could never have reached this current level of success. Thanks to all for unwavering support.

Master's thesis

Place	University of Turku
	Master's Programme in Digital Health and Life Science
	Department of Future Technologies
Authors	Abdullah Al-Farabi
Title	Structural modeling and analysis of non-mammalian CA VI
Pages	47 p. and 1 p. appendix
Date	November 2020

Abstract

Pentraxins are a family of host defense components of the innate immune system and phylogenetically conserved pattern recognition proteins. The short pentraxins are serum amyloid P component (SAP) and C-reactive protein (CRP), and the long pentraxins are pentraxin 3 (PTX3), which are only pattern recognization molecules.

In non-mammalian species, carbonic anhydrase VI (CA VI) contains an additional short pentraxin domain, and short pentraxins form pentamer units. Our study is an effort to understand the interaction properties pattern of the pentameric model of zebrafish CA VI+PTX complex.

Our results based on the comparisons between SAP and CRP structures and the cluster between their protomers seem different as the list of contacts varies. Between the pentameric model of zebrafish CA VI+PTX complex and CRP electrostatic surfaces, significant hydrophilic and hydrophobic differences are observed among the CA VI+PTX complex and CRP. The counted conserved residues in both pentraxins are the same, but the fact that conserved residues in similar positions and conserved between SAP and CRP. The SAP has a lot more conservation, which is buried, and it seems to have more contact structure because many of these essential residues are buried, at least 95%. The interface residues are not highly conserved but whereas the calcium-binding sites are relatively conserved. The Ca^{2+} binding sites of pentraxins are organized in a similar mode, and Ca^{2+} ions coordinating residues are the same in both pentraxins.

Finally, compare the contact between SAP and CRP. The amount of contact or amount of contact residue is similar, only that the mode of binding is very different. Because of the difference and low conservation of contacting residues, it is concluded that the mode of Ca6 associated PTX cannot be fully predicted based on this study. However, multiple van der Waals contacts and one ionic contact were observed to be conserved between CRP and SAP, which may form a basis for understanding contacts between CA VI +PTX monomers.

Abbreviations

CA	Carbonic Anhydrase
SAP	Serum Amyloid P Component
CRP	C-reactive Protein
PTX3	Pentraxin 3
MSA	Multiple Sequence Alignment
NCBI	National Center for Biotechnology Information
PDB	Protein Data Bank
RCSB	The Research Collaboratory for Structural Bioinformatics
UCSF	University of California, San Francisco
Cβ	Carbon-beta
SNPs	Single nucleotide polymorphisms
SLE	Systemic lupus erythematosus

Amino acid codes

Ala	A	Alanine
Cys	C	Cysteine
Asp	D	Aspartic acid
Glu	E	Glutamic acid
Phe	F	Phenylalanine
Gly	G	Glycine
His	H	Histidine
Ile	I	Isoleucine
Lys	K	Lysine
Leu	L	Leucine
Met	M	Methionine
Asn	N	Asparagine
Pro	P	Proline
Gln	Q	Glutamine
Arg	R	Arginine
Ser	S	Serine
Thr	T	Threonine
Val	V	Valine
Trp	W	Tryptophan
Tyr	T	Tyrosine

Table of Contents

1 Introduction

The carbonic anhydrases (CAs, EC 4.2.1.1) are a family of metalloenzymes that catalyze the dehydration of carbon dioxide or reversible hydration or bicarbonate ions. CAs plays an important role in our life and are connected in several physiological processes for example; calcification, bone respiration, respiration, and photosynthesis (Messerchmidt A *et al.,* 2004; Domsic JF *et al.,* 2008). CAs divided into six genetically individual families: (α, β, γ, δ, ζ and η), among them, η-CAs recently discover and mainly three classes (α, β, and γ) of CA are structurally different (Scozzafava A *et al.,* 2006).

Proteins are molecular units, on the nanometer scale, where the biological functions are exercise (Lesk AM 2001). They are the building blocks of all cells in our bodies and all living beings in all kingdoms. The frequency of twenty natural amino acids is higher with particular functions, and these amino acids can be grouped together, forming polypeptide chains or proteins, and in the different procedure determined by the genetic code (Xie J and Schultz PG 2005). Structures of protein are more conserved than sequences of protein, and several methods have been developed to assess the similarity of protein structure (Kolodny *et al.,* 2005).

From the structural aspects, structural characterization of macromolecular assemblies generally gestures a more difficult challenge than the structure determination of individual proteins (Russell RB *et al.,* 2004). The characteristics between the interface and non-interface proteins of a protein surface, such as secondary structure, proportions of amino acids, sequence conservation, side-chain conformational entropy, and solvent accessibility, are mostly used to characterize the specificity of topical structures relating to protein binding function. Ligand binding is a principal fact of protein functions. Proteins recognize their naturalistic ligands for signal transduction, transportation, or catalysis (Campbell SJ *et al.,* 2003).

Pentraxins are a family of host defense components of the innate immune system and phylogenetically conserved pattern-recognition proteins (Garlanda C *et al.,* 2005). Based on the length of their primary structure, the pentraxins are divided into two groups: short pentraxins are serum amyloid P component (SAP) and C-reactive protein (CRP), and the long pentraxins are pentraxins 3 (PTX3). The long pentraxin group serves as a prototype protein (Osmand AP *et al.,* 1977). The characteristic of CRP is acute phase protein in humans and the plasma concentration

of CRP enhancement in the both acute and chronic inflammatory environment but in SAP is not an acute phase protein in human (Tillett WS *et al.,* 1930; Hirschfield GM *et al.,* 2003). SAP and CRP have been found in all vertebrates where they have been investigated (Pepys MB *et al.,* 1978). CRP is located in the hemolymph of invertebrates for example, the arthropod *Limulus polyphemus* and the mollusca *Achatina fulica*. In all vertebrates, at least one short pentraxin is present, as well as in some invertebrates. Humans have got both SAP and CRP (Agrawal A *et al.,* 1990).

SAP and CRP are primarily originated in the liver, and during inflammation, PTX3 has produced various tissues. The major functions of short pentraxins are to recognize several pathogenic agents and then to either remove them or impartial their detrimental effects by utilizing the macrophages and complement pathway in the host (Hurlimann J *et al.,* 1966). PTX3 interacts with various ligands along with extracellular matrix components, growth factors, and selected pathogens, playing a role in complement activation and facilitating pathogen recognition by phagocytes.

Pentraxins are pattern recognition molecules that play important roles in the host defense and control inflammation. CRP measurement of blood has been used for minor infection and inflammation and predict cardiovascular disease. Recent studies consult that every pentraxin plays a role in regulating inflammation and tissue fibrosis (Zacho J *et al.,* 2010).

Contacts of protein are to go beyond the contact of computing networks, for the query of contacts in structures. Protein contacts enable the exploration of contacts within a ligand, single protein, or between a protein and nucleic acids, a protein complex, or other small molecules (Melis K *et al.,* 2018). Contacts can be identified using a distance threshold. A pair of amino acids are in contact if the distance between their particular atoms (carbon beta or carbon-alpha) is lower than a distance threshold (generally 8Å). Even though proteins can be better reconstructed with carbon-beta atoms but even now, carbon-alpha is widely used as a backbone atom (Duarte JM *et al.,* 2010). At present, the problem of precisely predicting contracts and using them to build 3-D models is mostly undecided, but the area of contract-based structure prediction is quickly going in advance.

Conservation analysis and identification across the α-CA family is the most important common functional elements. They identified highly conserved residues that are shared between all alpha carbonic anhydrase. In conservation analysis, the method determined the conserved residues within a group of homologous protein sequences. The ordinary multiple sequence alignment method only can show the conserved residues for a group of homologous proteins.

2 Literature Review

2.1 Alpha carbonic anhydrase

Carbonic anhydrase (CA. EC 4.2.1.1) catalyzes the reversible hydration of CO_2 (Sly W.S and Hu PY 1995). Three types of CA gene families are α-CA, β-CA, and γ-CA (Hewett-Emmett D and Tashian RE 1996). Nine CA isoforms have found within the α-CA, and from mammals, three CA-related proteins have been discovered and designated CA I to CA XII (Sly W.S and Hu P.Y 1995; Hewett-Emmett D and Tashian RE 1996). All the proteins vary in many aspects for example, tissue distribution, domain structure, activity, and subcellular localization, and all twelve proteins in the CA domain share significant sequence similarity and its nearly 260 amino acid residues (Weiping Jiang and Dwijendra Gupta 1999).

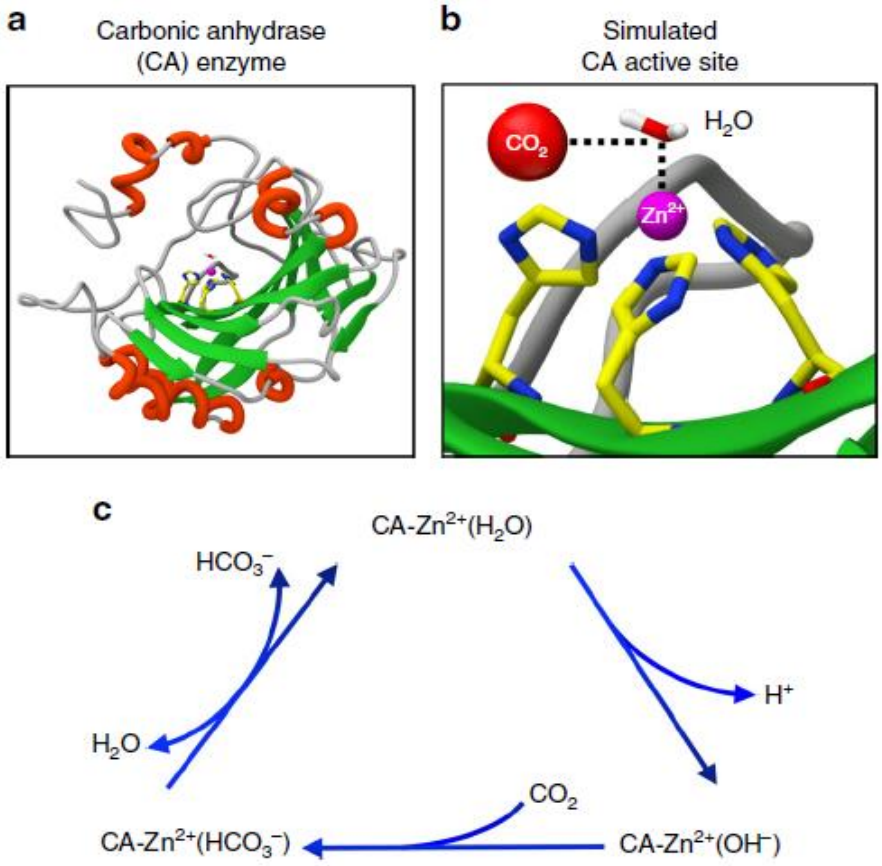


Figure 2-1: Overview regeneration mechanism of Carbonic anhydrase enzyme and its CO_2. **(c)** Overall implementation of the catalytic cycle, for CO_2 hydration to HCO_3^- with zinc as the metal in the CA active site.

Source:https://www.researchgate.net/figure/Carbonic-anhydrase-enzyme-and-its-CO2-capture-and-regeneration-mechanism-a-Ribbon_fig3_323614490

2.2 Carbonic anhydrase VI

Carbonic anhydrase VI is specifically revealed in the salivary gland of several mammalian species. All the sequences of the amino acid are highly conserved across the species (Fernley RT *et al.,* 1998; Aldred P *et al.,* 1991). Such as the mature bovine CA VI is 87% and 68% identical with the human protein and sheep, respectively (Jiang W et al., 1996). Also, in all active CA family members, three histidine residues are conserved (Figure 2). The proposed functions of CA VI include the maintenance of pH and bicarbonate levels in saliva (Parkkila S and Parkkila A-K 1996). It has been demonstrated that CA VI is identical to gustin, approximately 3% of human parotid saliva protein constituting a Zn-metalloprotein, which is reduced in patients with loss of taste and is connected with pathological morphology of taste buds (Thatcher BJ *et al.,* 1998).

A recent study suggests that in dental biofilm, CA VI neutralizes acidity and thereby defends teeth from carries (Kimoto M *et al.,* 2006; Kivela J *et al.,* 1999). Salivary concentrations low of CA VI seem to be associated with an increased outbreak of caries and acid peptic diseases (Kivela J *et al.,* 1999). In chromosome 1, the human CA VI (HCA VI) gene is located. Encoding the gene of human CA VI has 7 introns and 8 exons (Jiang W and Gupta D 1999). In recent years, many studies have identified CA VI in several different species. Examples of these species are Human, Horse, Pig, Dog, Cow, Sheep, Rat, and Mouse.

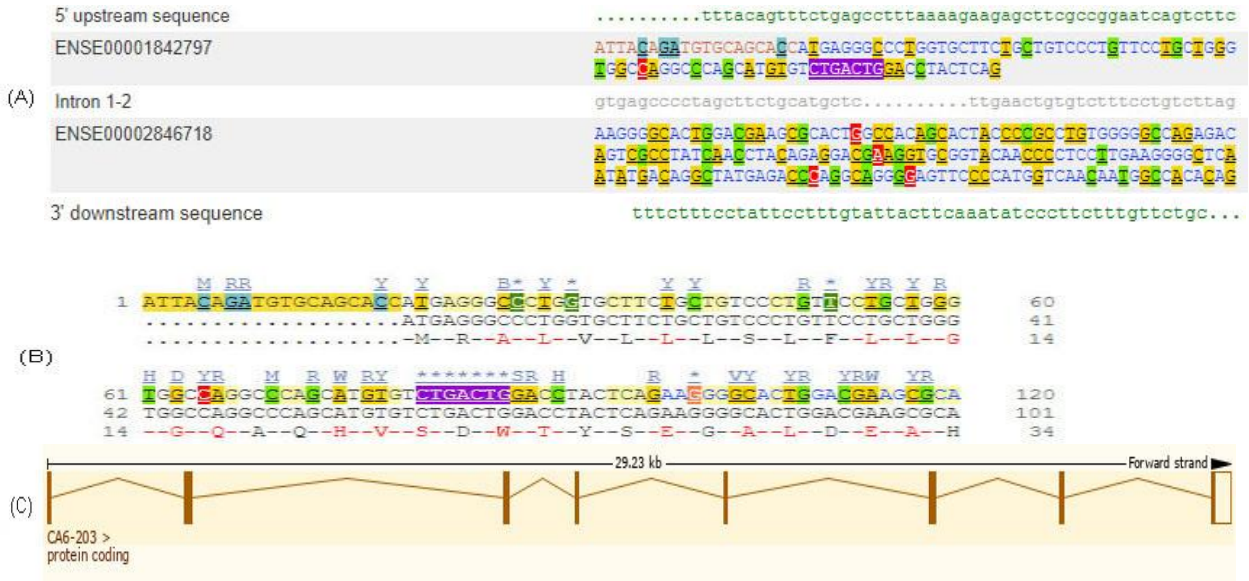


Figure 2-2A: Outline of the Human CA6 gene. (A) The exon/intron sequences and (B) The cDNA sequences. (C) The protein coding (exon/intron organization). The number of introns and exons is represented by boxes and lines

respectively. The statistics of protein coding: Exons: 8, Coding exons: 8, Transcript length: 1334bps, and Translation length: 308 residues.

Source:https://www.ensembl.org/Homo_sapiens/Transcript/Summary?db=core;g=ENSG00000131686;r=1:8931257-8989712;t=ENST00000377443

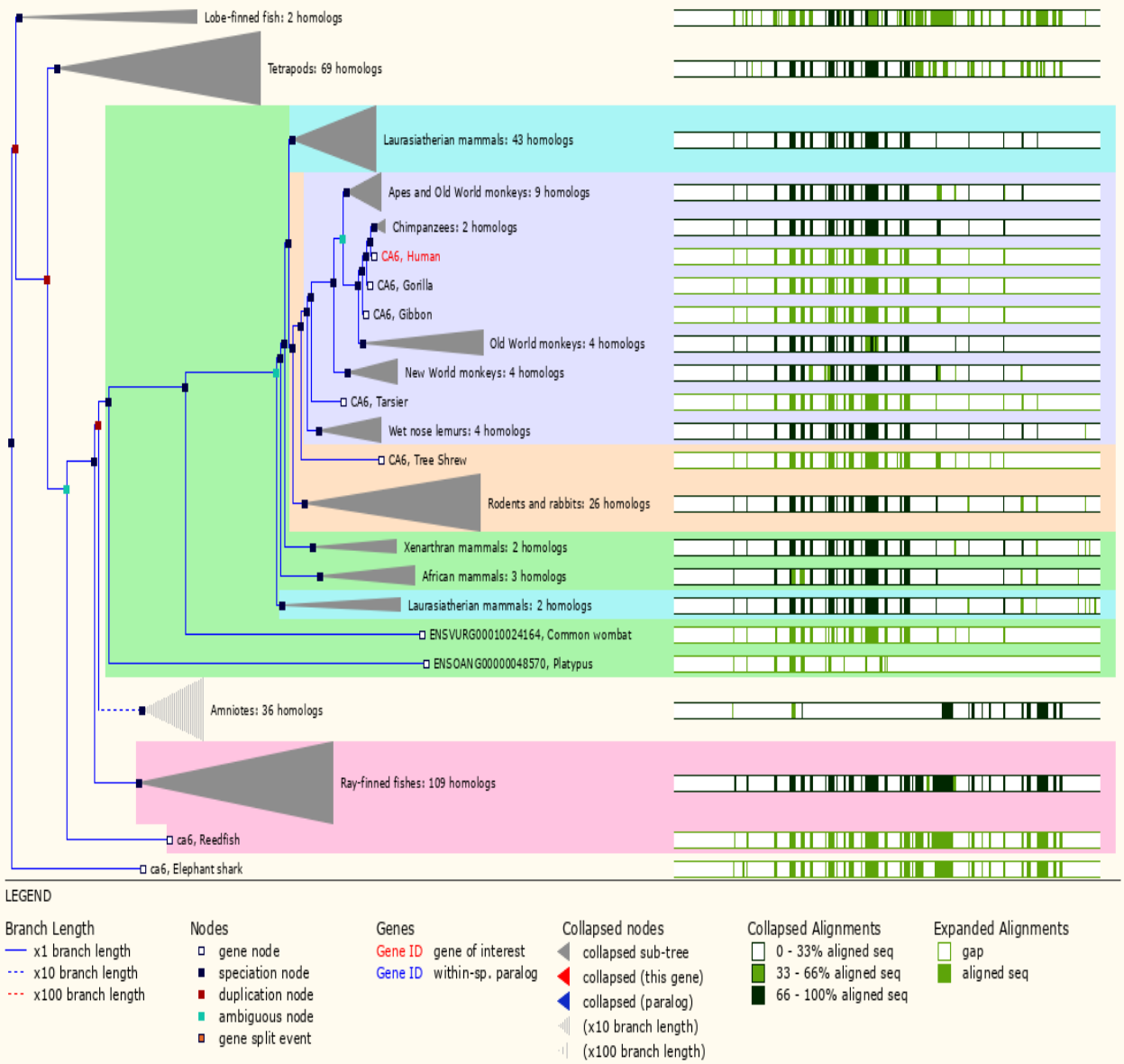


Figure 2-2B: The gene tree of carbonic anhydrase 6 (Gene: CA6 **ENSG00000131686**). Number of genes 324, number of speciation nodes 282, number of duplications 16, number of ambiguous 23, and number of gene split events 2.

Source:https://www.ensembl.org/Homo_sapiens/Gene/Compara_Tree?db=core;g=ENSG00000131686;r=1:8931257-8989712;t=ENST00000377443

2.3 Pentraxin

The pentraxins are an evolutionary conserved superfamily of fluid phase pattern recognition molecules and characterized by a cyclic multimeric structure. C-reactive protein (CRP) and serum amyloid P component (SAP) comprise the short pentraxin of the superfamily. Most of the mammalian species are found in these two types of pentraxins. Pentraxin 3 (PTX3) is the prototype of the long pentraxin arm. Genetic and epigenetic studies in humans advise that PTX3 plays essential non-redundant roles in innate immunity and inflammation as well as remodeling of tissue (Barbara Bottazzi1 *et al.,* 2016).

Most mammals express both SAP and CRP, primarily synthesized by the liver, and there is now considerable proof for local synthesis as well. The initial stages of CRP are defined by different factors along with promoter single nucleotide polymorphism (SNPs) and variations of a gene that influenced CRP synthesis. Primarily synthesis is controlled at the transcription level, although post-synthesis mechanisms increase CRP release from the endoplasmic reticulum quite dramatically during the acute phase reaction. SAP is not an acute phase reactant and is the same in patients with systemic lupus erythematosus (SLE).

More than 80 years ago, the pentraxins were discovered, but the understanding of their complete roles in the biology of infection, inflammation, and autoimmunity induction an exciting challenge to researchers in the field. Figure 2-3, shows an overall explanation of the immune system of short pentraxins interactions and activities (Mold *et al.,* 2012).

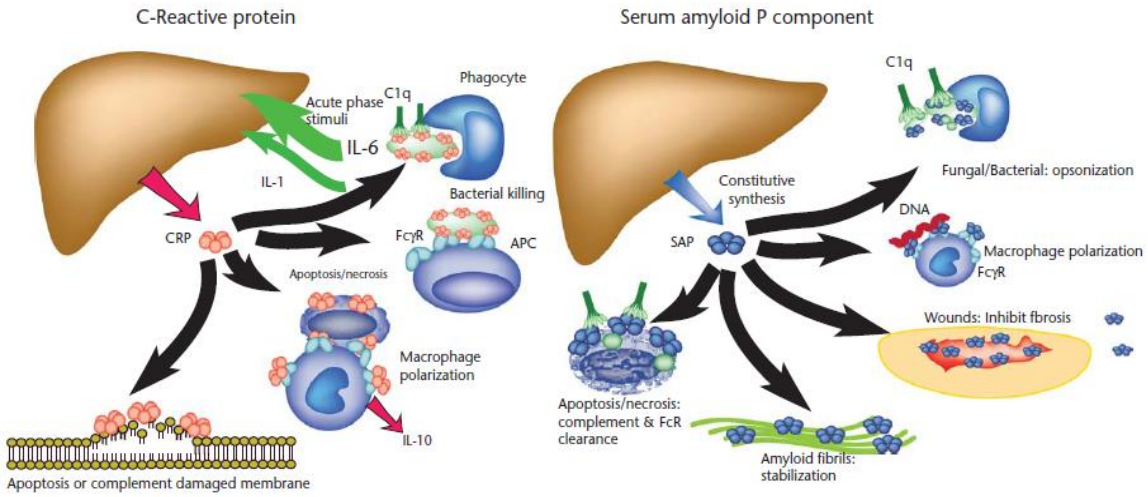


Figure 2-3: Functional overview of interactions and proposed activities of pentraxins in the immune system. Source: https://images.app.goo.gl/R1UTXDcuPVzUeEX67

2.4 Contacts

Contacts are important to several types of protein studies from, Bioinformatics to Structural Biology. Contacts are indispensable to the accuracy and reliability of applications, including protein structure prediction, network contacts analysis, quality assessment of protein structures, folding prediction, protein-protein, and protein-ligand interactions. Different types of contacts analyze in protein-protein interfaces, such as aromatic stackings, hydrogen bonds, hydrophobic and ionic interactions (Pedro M. Martins *et al.,* 2018). In 3-D protein structures, residue-residue contacts are pairs of spatially close residues. A pair of amino acids are in contact if the distance between their specific atoms is less than a distance threshold (usually 8Å) figure 2-4 (Duarte JM *et al.,* 2010). A pair of residues is defined as a contact if the distance among their carbon-beta (Cβ) atoms is less than or equal to 8Å. In recent work, a pair of residues is expressed to contact if their Cα atoms are separated by at least 7Å with no minimum sequence separation distance defined (Jones DT *et al.,* 2012). Based on the type of information, the existing method for residue contact prophecy can be classified into five categories: (1) machine learning method-based, (2) hybrid methods, (3) physiochemical information-based, (4) coevolution-derived information-based, and (5) template-based (Schneider M and Brock O 2014). Other authors classify contact into four groups: (a) 3-D model-based, (b) correlated mutations, (c) machine learning, and (d) template-based (Di Lena P *et al.,* 2012). On the opposite, classified into three groups: (a) statistical methods, (b) template-based, and (c) machine learning (Björkholm P *et al.,* 2009).

Also, covalent bonds and non-covalent bond contacts among the residues (residue-residue contacts) are significant for stability, conformational flexibility, co-operative folding of biomolecules, and molecular recognition (Ken Nishikawa TO *et al.,* 1972). Representations of non-covalent contacts and their comparison between relevant structures have provided insights into protein stability, ligand binding, allosteric mechanisms, conformational switching, and the determinants of protein fold and protein complex assembly (Ken Nishikawa TO *et al.,* 1972; Vishveshwara S *et al.,* 2002; Suel GM *et al.,* 2003; Kornev AP *et al.,* 2006). So thus, a residue contact-based analysis and representation of protein structure enable to recognize of critical contacts and holds the possibility for understanding how biomolecules responsible in new ways and their activity (Doncheva NT *et al.,* 2011; Martin AJ *et al.,* 2011; Zhang X *et al.,* 2013).

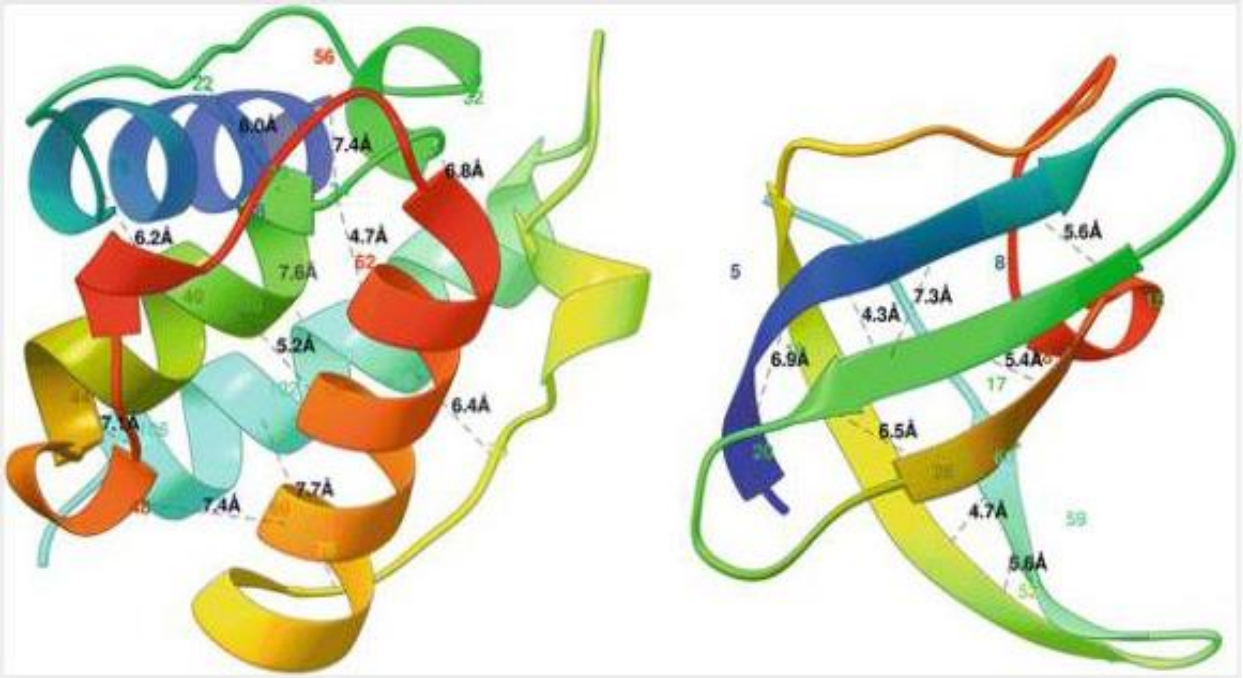


Figure 2-4: Contact analysis of the two globular proteins, black dotted lines show the contacts between the alpha-helical and beta-sheets protein and the distance in Armstrong. The alpha-helical 1bkr (left) protein has many long ranges, and the beta-sheet 1c9o (right) protein has more medium and short-range contacts.
source: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4894841/figure/F1/

2.5 Hydrogen bonds

Hydrogen bonds or H-bonds are an attraction between an electronegative atom and a proton (hydrogen) atom. Hydrogen bonds are weaker than both ionic and covalent bonds but stronger than van der Waals forces. Hydrogen bonds can exist in different molecule atoms (intermolecular) or parts of the same molecule (intramolecular) (Jeffery GA 1997). Ordinarily, nitrogen, oxygen, or fluorine atom is covalently bonded to a hydrogen atom (-NH, -OH, or -FH) whose electrons share unequally, and its high electron affinity causes the hydrogen to accept a slight positive charge. The other pairs of the atom are also typically N, O, or F has an unshared electron pair and its slight negative charge. Hydrogen bonds are also important in the formation of protein fold. The proteins are made; they have to twist and fold into specific shapes to perform their functions. Hydrogen bonds are mostly responsible for the way the proteins fold, and without them, proteins would not be functional.

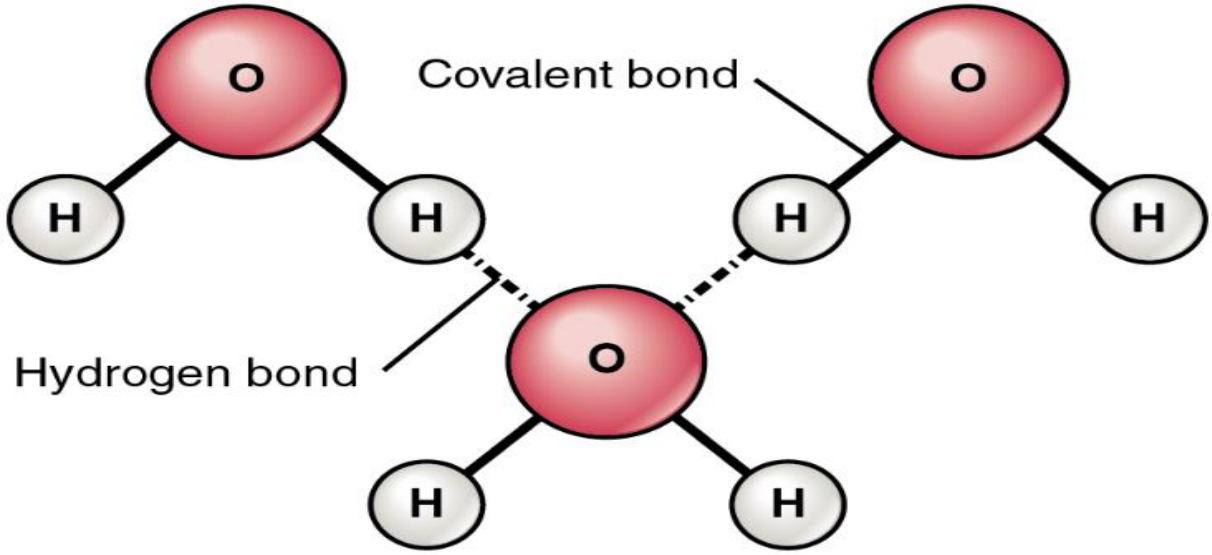


Figure 2-5A: The positive hydrogen is attracted to the electronegative oxygen. Source: https://commons.wikimedia.org/w/index.php?curid=30131145

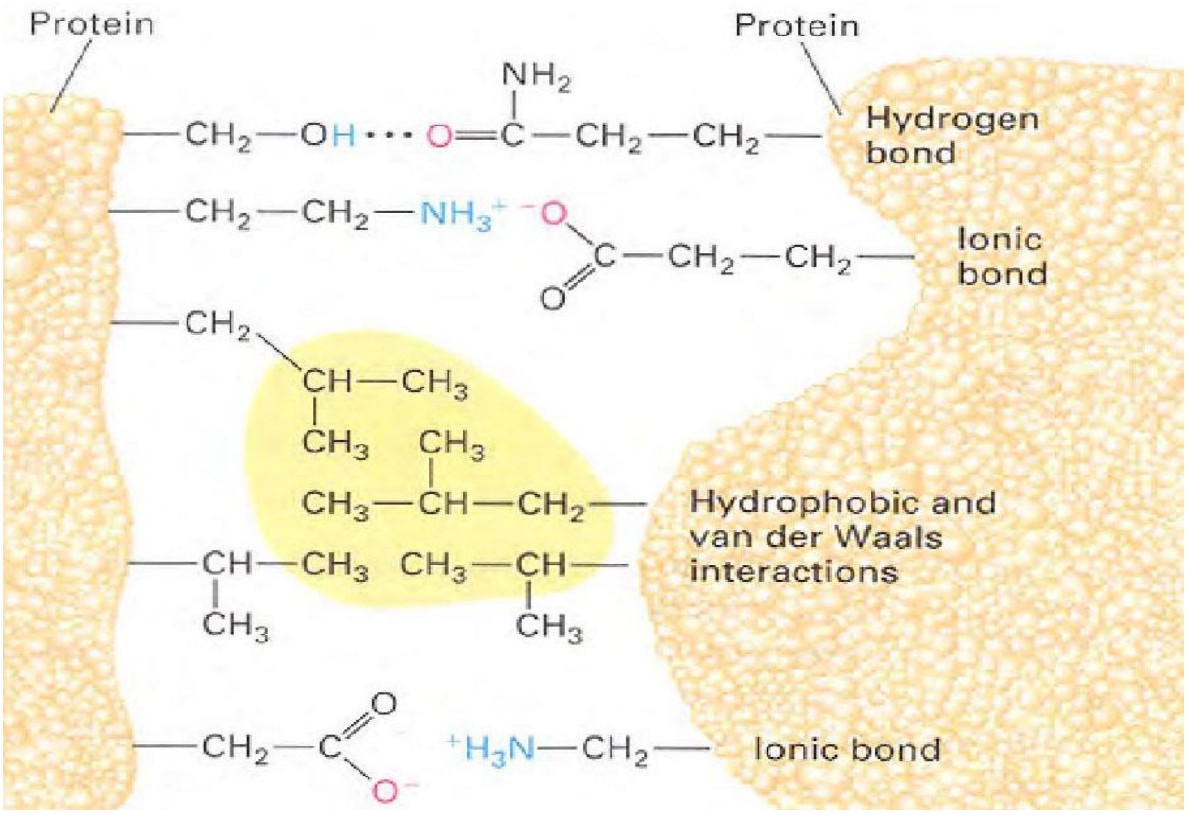


Figure 2-6B: Pair of binding protein by one hydrogen bond, two ionic bonds, and a combination of van der Waals and hydrophobic interactions.

Source: https://www.ncbi.nlm.nih.gov/books/NBK21726/figure/A299/?report=objectonly

2.6 Hydrophobic interaction

Hydrophobic interactions are inherently weak and occur between two or more non-polar molecules in polar environments, they play an important role in many biological processes like lipid bilayer formation, protein folding, and protein-protein interaction. There are also hydrophobic reactions within clusters in amphiphilic/ amphipathic molecules, such as membranes and phospholipid bilayer. Hydrophobic interactions are comparatively stronger than other intermolecular forces (i.e., Hydrogen bonds or van der Waals interaction). The strength of interactions depends on different factors, including; temperature (however, hydrophobic interaction will denature at an extreme temperature), number of carbons on the hydrophobes, and hydrophobes shape. For protein folding, hydrophobic interactions are also important. This is significant in holding a stable and biological action because it allows the protein to reduce in the surface area and decrease the unaccepted interactions with water. Many other biological substances depend on hydrophobic interactions for their survival and functions (Chang and Raymond 2005; Garrett *et al.,* 2005). They exist experimental evidence that the strength of hydrophobic interactions is dramatically affected by two biologically related cations, guanidinium, and ammonium when these ions are immobilized close to hydrophobic patches of molecules (Ma CD *et al.,* 2015).

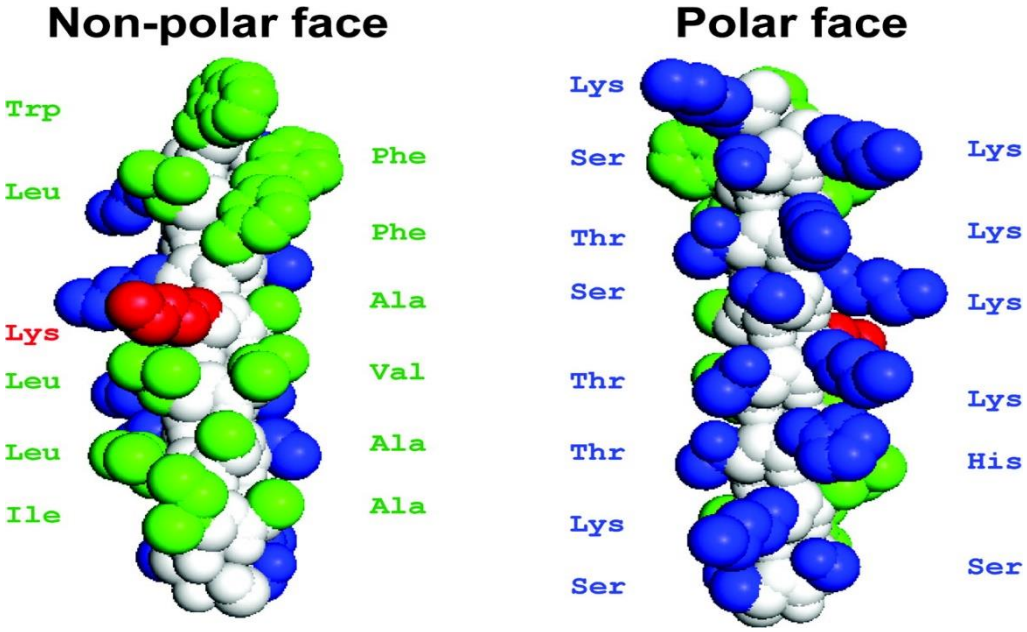


Figure 2-7A: Hydrophobic amino acids on the non-polar face of the helix are colored green; hydrophilic amino acids on the polar face of the helix are colored blue and the peptide backbone is colored white. The Lys substitution on the non-polar face of the helix is colored red. Source: https://aac.asm.org/content/aac/51/4/1398/F1.large.jpg

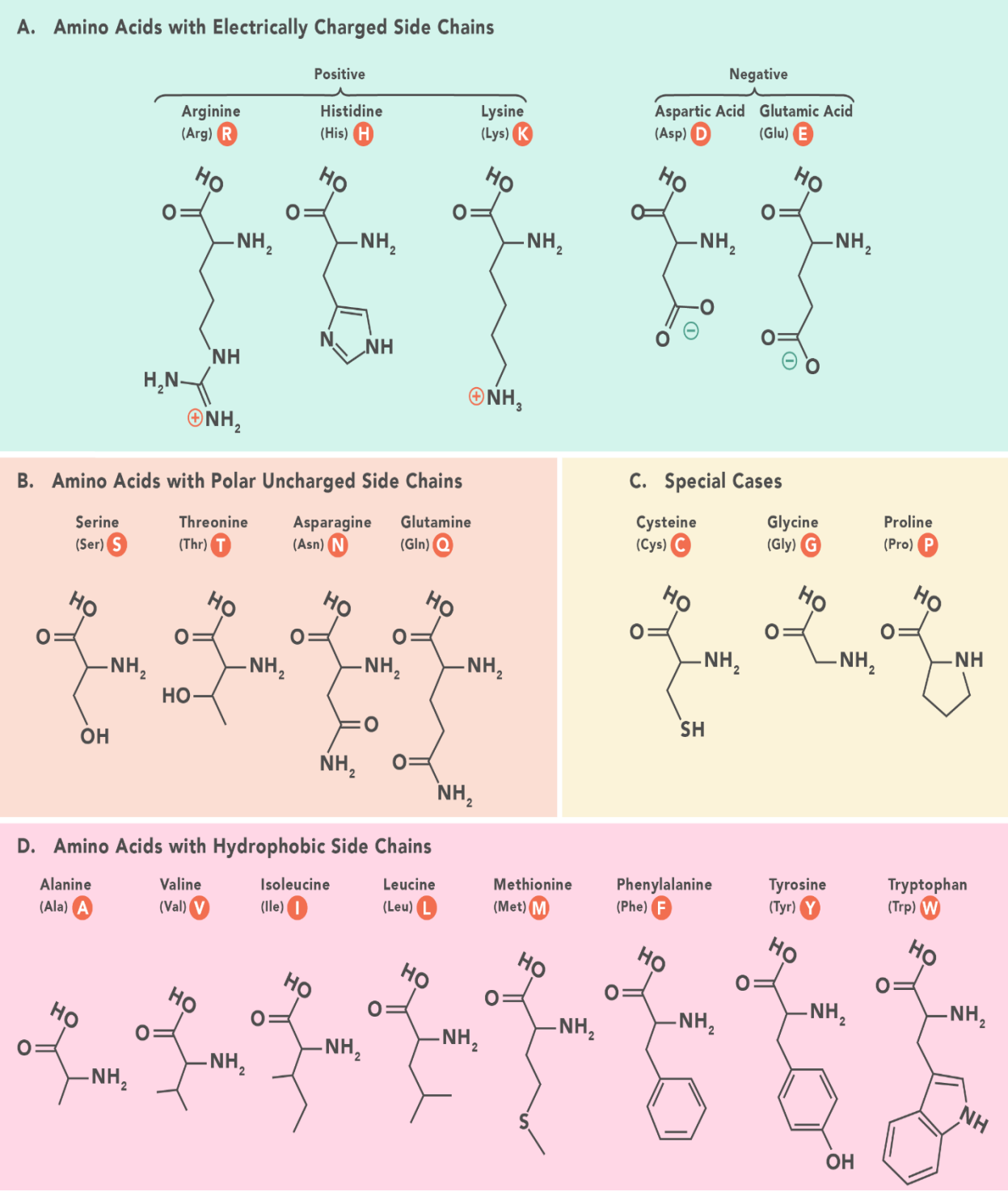


Figure 2-8B: Chemical structures of the 20 amino acids that make up proteins.

Source:https://www.technologynetworks.com/applied-sciences/articles/essential-amino-acids-chart-abbreviations-and-structure-324357

3 Aim of the Study

- I. Identifying PTX-PTX interfaces in CRP and SAP and analyzing their modes of binding.
- II. Finding conservation in PTX-PTX interface residues (CA-linked PTX model vs. CRP vs. SAP).
- III. Same but focusing on Ca^{2+} binding residues of PTX (CA-linked vs. CRP vs. SAP).
- IV. Visualizations of sequence conservation on protein surfaces.
- {V. Discovery of potentially functional residues in CA VI-PTX by comparison of sequence groups for mammals vs. non-mammals to see possible domain-domain interaction residues (conserved surface residues, not conserved in mammals).}

4 Methodology

4.1 Sequence retrieval

From Ensembl and NCBI, all the protein sequences were retrieved. Ensembl is a genome browser for a vertebrate genome that supports research in sequence variation, comparative genomics, and transcriptional regulation. Genomic sequence information mostly vertebrates, particularly focus on the key model organisms for example, zebrafish, humans, and mouse.

For the protein database, NCBI is an important source. It maintains the text record for distinct protein sequences, derived from several resources such as GenBank, PDB, NCBI Reference Sequence (RefSeq) project, and UniProtKB/SWISS-Port (Sayers E 2013).

4.2 Decrease redundancy

Protein sequence databases have some additional or unessential residues, and many different studies we used for removing their additional part. Redundancy is one kind of technique, and it occurs in a dataset when different types of similar data sets are present at the same time. Redundancy in a collection of sequences occurs when the same data set one or more similar/homologous sequences are present (Hobohm U *et al.,* 1992). Five different computer programs were used: Decrease redundancy, Pisces, cd-hit, SkipRedundant, and BlastClust. FASTA format input files accept all the programs.

4.3 Clustal Omega

Multiple sequence alignments (MSA) are indispensable and extensively used computational procedures for biological sequence analysis in bioinformatics. Clustal Omega is the most recent MSA algorithm from the Clustal family. Only protein sequence alignment is used in this algorithm. The precision of Clustal Omega on small numbers of sequences is equivalent to other high-quality aligners. This method produces a multiple sequence alignment by using the k-tuple method, firstly producing pairwise alignments. Later the sequences are clustered by using the mBed method. This is followed by the k-means clustering method. The guide tree is the next constructed using the UPGMA procedure.

4.4 Web logo

Web Logo is a web-based application and generates the sequence logos. The logo generates from the (https://weblogo.berkeley.edu/logo.cgi). Sequence logos are a graphical representation of the patterns within a multiple sequence alignment provided in either FASTA (Pearson W.R and

Lipman DJ 1988) or CLUSTAL formats (Higgins DG and Sharp PM 1988). Every logo consists of stacks of letters, one stack for each position in the sequence. The overall height of every stack indicates the sequences conservation at that position, whereas the height of symbols within the stack indicates the relative frequency of the corresponding amino or nucleic acid (Gavin E *et al.,* 2004). The sequence logo provides a richer and more precise explanation of sequence similarity than consensus sequences.

4.5 RCSB Protein Data Bank

The RCBS Protein Data Bank provides multiple tools for deposition, analysis, query, annotation, molecular visualization, and educational resources to use the PDB archive. The Worldwide Protein Data Bank organization guides the PDB archive (wwPDB; http://wwpdb.org). The management of the PDB In October 1998, it becomes the Research Collaboratory for Structural Bioinformatics (Helen M. Berman *et al.,* 2000). Initially, PDB had been used by a limited small group of structural research experts. Now, PDB has varying expertise in cryo-electron microscopy techniques, X-ray crystal structure determination, and theoretical modelling (Helen M. Berman *et al.,* 2000). The innovation of the RCSB is to generate modern technology on a resource-based and simplify the structural data analysis. I have used SAP (PDB ID: 4AVS) and CRP (PDB ID: 3PVN) protein structures to compare the different species and reliable structural analysis.

4.6 Pentamer model

Based on the Prajwol Manandhar pentamer (SAP) model, I have performed all of my project work (Manandhar Prajwol 2015). Source: http://urn.fi/URN:NBN:fi:uta-201512022478

4.7 UCSF Chimera

UCSF Chimera is a highly extensible program for interactive analysis and visualization of molecular structures and related data. Including docking results, sequence alignments, density maps, supramolecular assemblies, conformational ensembles, and trajectories. The structural analysis also includes contacts, hydrogen bonds, and clash detection; surface area, distance, bond rotation; morphing between different conformations of protein and volume measurements; structure building.

5 Results and discussion

5.1 Protein sequence alignment of mammals and non-mammals

Figure 5-1: The protein sequence aligns the colour of random things we could have them different and by the colour we can easily see the alignment is correct position. The fully conserved (100% identical) reasons are important residues in every alpha CA domain. The Q, S, P and we have the catalytic histidine are important for the active sides.

We have retrieved 120 mammal protein sequences and 103 non-mammal protein sequences from NCBI and Ensembl. Finally selected 75 protein sequences and the rest of the sequences were deleted because some of the sequences were short, had X characters, long gaps or insertions.

5.2 Sequence logo of mammals and non-mammals

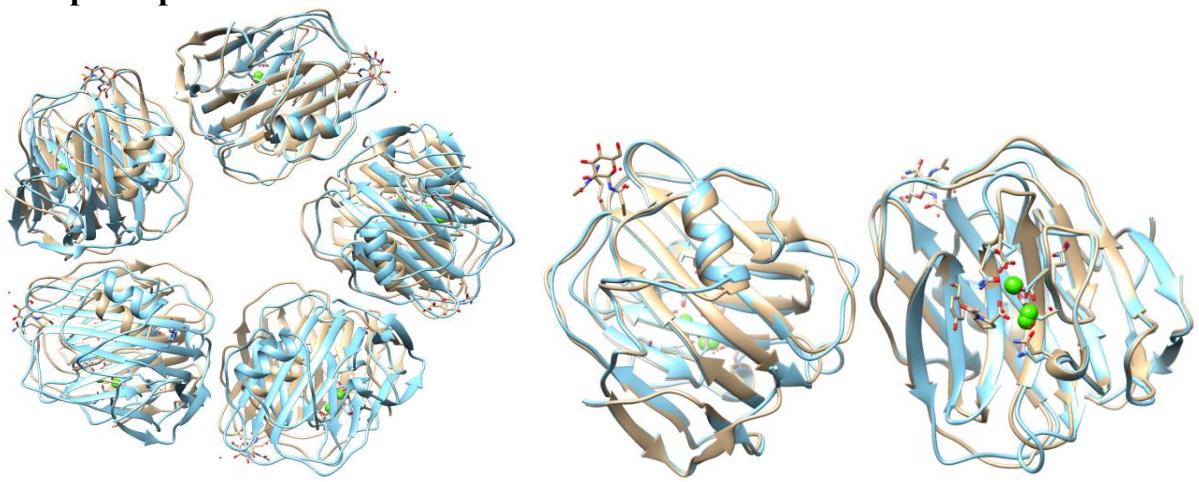
Mammals

Non-mammals

Figure 5-2: Red negative (glutamic acid), blue positive (arginine and histidine), polar and neutral residues in green, hydrophobic residues in black. Every alpha CA has this GSEH/GAEH and EH has highly conserved. In logo here, I got a high letter because 90 or 95% conserved but whereas in cluster W, it has to be 100% exactly, otherwise we don't get it asterisk.

The comparative evaluation of structures is traditionally and effectively used to predict protein functions, make evolutionary inferences, and find functionally important residues. Visual inspection of superimposed structures and associated sequences is necessary to differentiate well-conserved regions of protein structures (Holm L and Sander C 1996). Figure 5-1: both SAP and CRP structures have pentameric interaction, and the protomers are compactly stationed in asymmetry.

5.3 Superimposed molecule between SAP and CRP



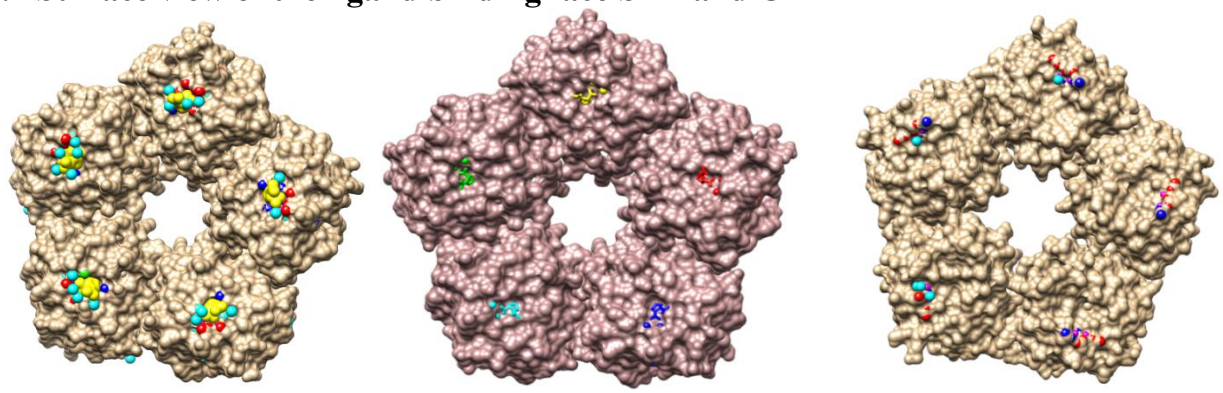
(a) (b) (c)

Figure 5-3: Chimera display [a] the superimposed molecule between 4AVS (Gray) and 3PVN (Cyan) notice that the protein is represents the ribbon representations in both molecule chains; whereas the ligand and some specific residues

are shown in stick representations. [b] The superimposition of 4AVS (Gray) and 3PVN (Cyan) protein structures starting N-terminal and ending at C-terminal of both chains and [c] direct view of active side.

From the visualization of these two structures, the superimposition shows that they are similar to some content. However, it is clear that the structures have differences, as shown in Figure 5-3. Ribbon overlay of SAP protomer (Gray) and CRP protomer (Cyan) (Figure 5-3b) then 22° rotation has been applied to fetch the protomers into identical orientation (Thompson SG *et al.,* 1995). The secondary structures of 4AVS have 5% helical (2 helices; 12 residues), 46% beta-sheet (20 strands; 94 residues), and 3PVN, 7% helical (3 helices; 15 residues) 46% beta-sheet (22 strands; 95 residues) presents.

5.4 Surface view of the ligand binding face SAP and CRP



(a) (b) (c)

Figure 5-4: The surface view of the ligand binding face of serum amyloid P component and C-reactive protein. Both protomer residues interactions between the chains are colored by heteroatom. [a] Every protomer of SAP contains a binding site, which is displayed occupied by N7P (yellow), HOH (cyan), CA (magenta), ASP (red), and TYR (blue). [b] SAP (4AVS) as shown the direct view of active side. [c] Every protomer of CRP contains a binding site, which is displayed occupied by HOH (cyan), CA (magenta), ASP (red), and GLU (blue).

The presence of the calcium-binding site of SAP structure is organized in a similar form to those in CRP. The different number of protein ligands to calcium ions helps explain the distinctions in the known calcium affinity among the proteins. In both sites, phosphate ligands provide one excessive bond per metal atom. In both SAP and CRP (Figure 5-4), protomers have ionic interactions, and hydrogen bonds are present. In addition, both protomer interfaces have an extra salt bridge, and a comparison of interaction between the SAP and CRP shows two conserved salt bridges. In CRP, a similar surface area is buried but variations in sequence relative to SAP. The

subunit rotations lead to small changes in both patterns that must stabilize the rotated state of the subunits. For example, Lys123 and Glu197 residues in CRP form an intermolecular ion pair but not formed by Gln121 and Tyr195 in SAP. The CRP Arg118-Asp155 and Lys201-Glu101 ion pairs, however, correspond to Glu99-Lys199 and Glu153-Arg118 in SAP. Consequently, between the SAP and CRP, protomer interactions have a conservation pattern (Thompson D 1999).

5.5 Hydrophobicity surface of the protein SAP and CRP

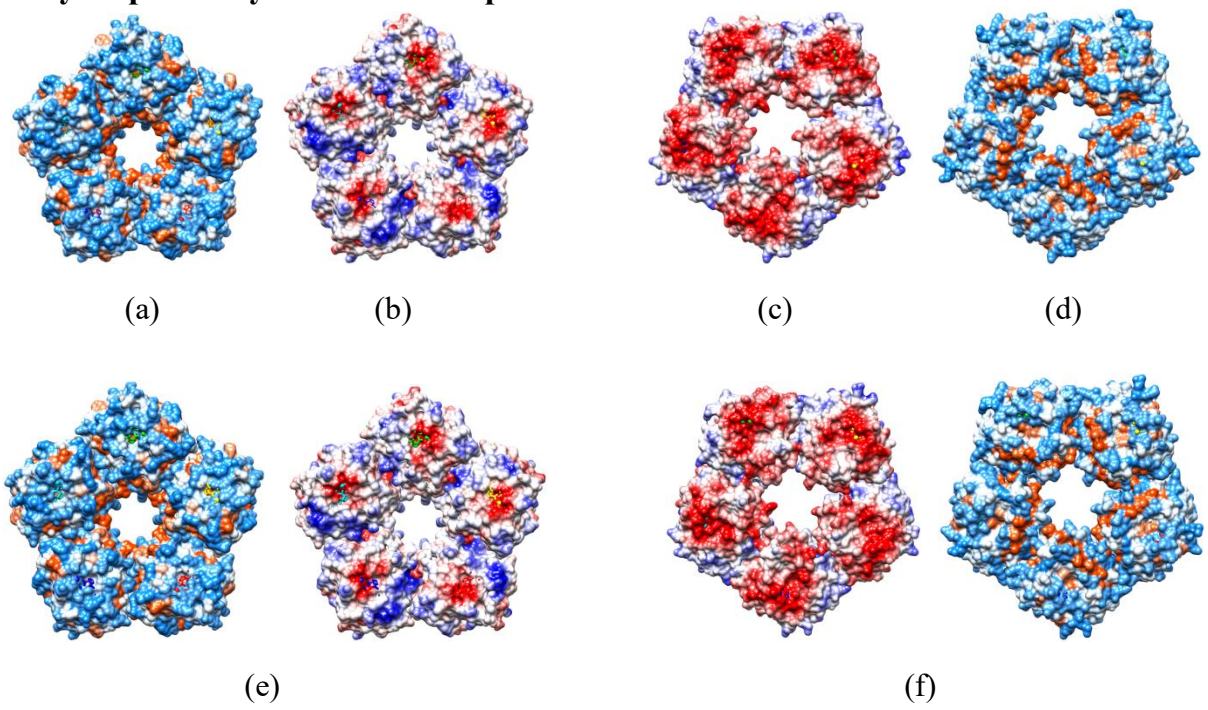


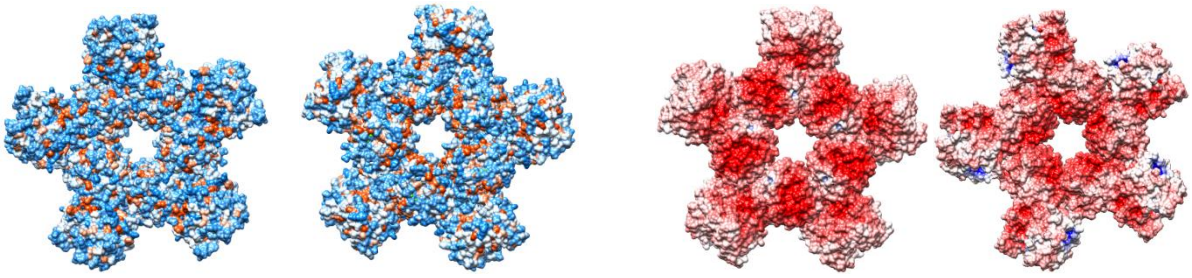
Figure 5-5: The hydrophobicity surface of the protein SAP [a] and CRP [d] as shown blue patches are hydrophilic and orange patches are hydrophobic and white are neutral residues. The representation of the surface charge of calcium depleted human SAP and CRP pentamer. The extremely negative area corresponds to the calcium and ligand-binding region. To visualize the charge distribution, the red color for electronegative potential, blue color electropositive potential, and white for neutral. [e], [f] the back view of SAP and CRP pentamer.

The hydrophobic fragment distribution analysis expresses detachments in surface hydrophobicity at several cavities that exist in SAP and CRP (Figure 5-5a, 5-5d). Both structures showed a large number of hydrophilic and hydrophobic probable patches; however, a significant quantity of hydrophilic patches are existent on the surface areas, and many small and large hydrophilic patches are scattered on the surfaces and also grooves of hydrophilic that might constitute the interaction sites accountable for the stability of SAP and CRP.

In the hydrophobic area, the orientation of residues is indispensable for the native properties and stability of a protein. The folding of protein and stability of hydrophobic interaction is one of the most important factors (Munson M *et al.,* 1996, Dill KA *et al.,* 1995, Giovambattista N *et al.,* 2008). For the stability of SAP and CRP, hydrophobic patches may be the most balancing factor. As opposed to, hydrophobic patches also play a significant role in the Ca^{2+} reliant binding of pentamer to phosphocholine (Shrive AK *et al.,* 1996). The binding site of phosphocholine contains two synchronized calcium ions connected to a hydrophobic pocket. Phe-66 and Glu-81 are two major residues for the binding of phosphocholine. Phe-66 offers hydrophobic interactions with the methyl groups of phosphocholine, while Glu-81, placed on the contrary side of the hydrophobic pocket, interacts with the positively charge choline nitrogen (Agrawal A *et al.,* 2002; Black S *et al.,* 2003).

The electrostatic surfaces of SAP (Figure 5-5b) and CRP (Figure 5-5c) were clearly showing dissimilar, large positive (SAP), and negative potential patches (CRP). The number of positive and negative charged residues provides charge stability, and the charged residues are responsible for ligand binding. The uncharged residues number was comparatively much more than the charged residues. However, protein stability connected to the charge distribution of protein folding and the charge segments (both positively charged and negatively charged) provide the charge stability of SAP and CRP.

5.6 Hydrophobicity surface of the pentameric model



(a) (b) (c) (d)

Figure 5-6: [a], [b] the hydrophobicity surface of the pentameric model of zebrafish CA VI+PTX complex, as constructed by aligning five units with each of the subunit of human CRP structure (3PVN). [c] [d] To visualize the charge distribution, the red color for electronegative potential, blue color electropositive potential, and white for neutral. The font view [a], [d], and the back view [b], [c].

From the hydrophobicity surface and electrostatic view, the two domains of zebrafish CA VI and PTX complex, and the binding cleft of surface area in Figure 5-6a and 5-6c, engulf the two domains and bind the surface in the cleft. The pentameric model of zebrafish CA VI+PTX complex rotates by ~180° vertical axis and visualizes the electronegativity in the binding gap and separating the two domains. Supervisions of the protein binding sites demonstrate that the electrostatic surfaces complementary to the charged substrate. Pentameric model binds virtually inserted C-reactive protein with negatively charge hydrophobic and phosphoryl groups, and the negative electrostatic potential significantly expands over a significant surface area. The pentameric model exhibits a large negative charge and contributes to and around the binding site, Figure 5-6c. In the visualization of Figure 5-6d, the back view of the pentameric model exhibits more neutral residues, and small electronegative potential residues compare with Figure 5-6c, front view. The front view and rear view of the hydrophobicity surface in Figure 5-6a and 5-6b. Generally, there are no visualization differences.

5.7 Pentamer models of zebrafish CA VI based on SAP and CRP pentamers

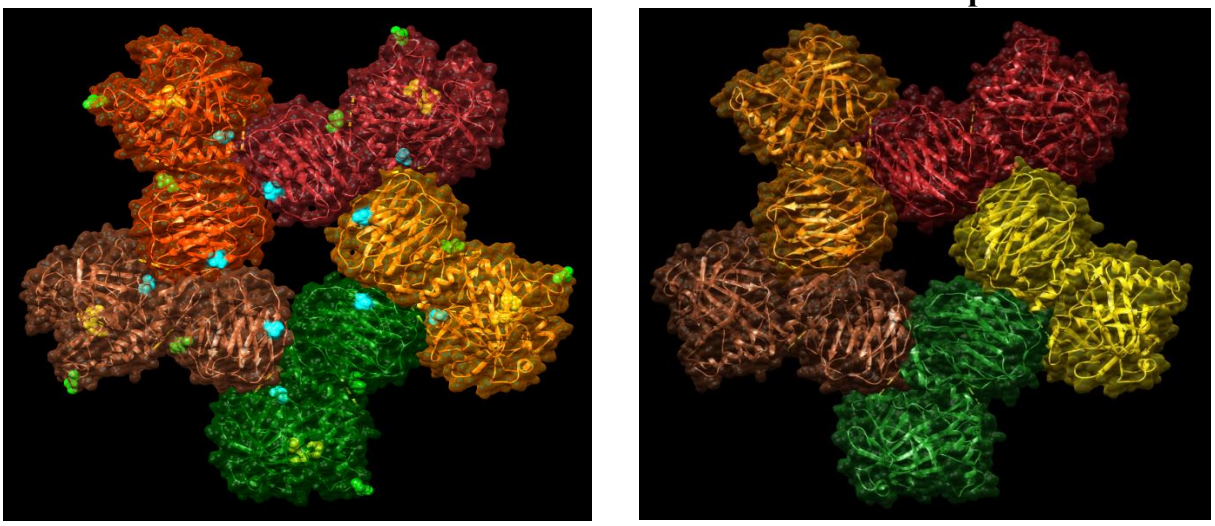


Figure 5-7: The pentamer models of zebrafish CA VI based on SAP & CRP pentamers. In SAP (left) -(yellow)- active side histidine, (Orange)- cysteine in disulphide bond, (hot pink)- amphipathic helix, (green)- occupied N-glycosylation, (cyan)- unoccupied N-glycosylation site.

Our observations, perhaps the idea of compact and the CA domain in CRP (right) are almost flat, but the general shape slightly different in SAP; it is a little bit curved when based on CRP. The number of clashes is higher for both because the model is not optimized as pentamers. The packing of the CA domain and pentraxins domain in the original monomer model might not be accurate.

But instead, the domains could be less compactly packed, allowing the CA domain to be a far way to eliminate the clashes. The other thing is comparing how many clashes are in CRP and SAP; maybe both are many clashes and not showing a big difference.

The CRP structure consists of five promoters, and every protomer has 206 amino acids. They are displayed as a cyclic pentamer. At the same time, the β-fold of antiparallel β-strands is similar to that found in SAP (Emsley J *et al.,* 1994). In the pentamer, promoters arrangement particularly in the tertiary structure and surrounding the CRP region of additional residues Serine and Threonine. A major structural rearrangement, one of the loops firstly involved in calcium coordination, and consequently in the structure of neighboring regions. Opposite the calcium-binding sites on the pentameric face, there is an attractive, extended cleft in CRP but not seen in SAP (Annette K. Shrive *et al.,* 1996).

5.8 Interface residues of SAP and CRP

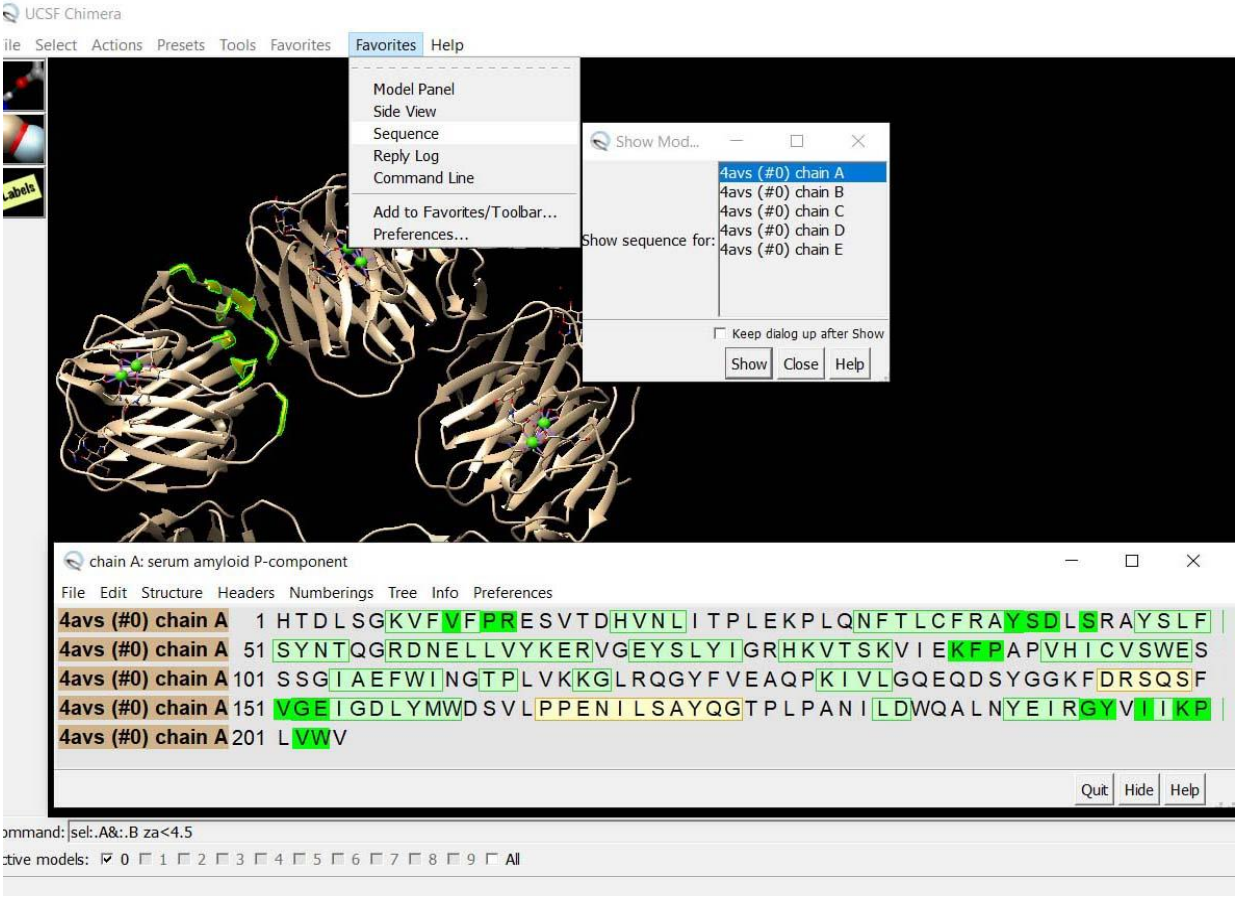
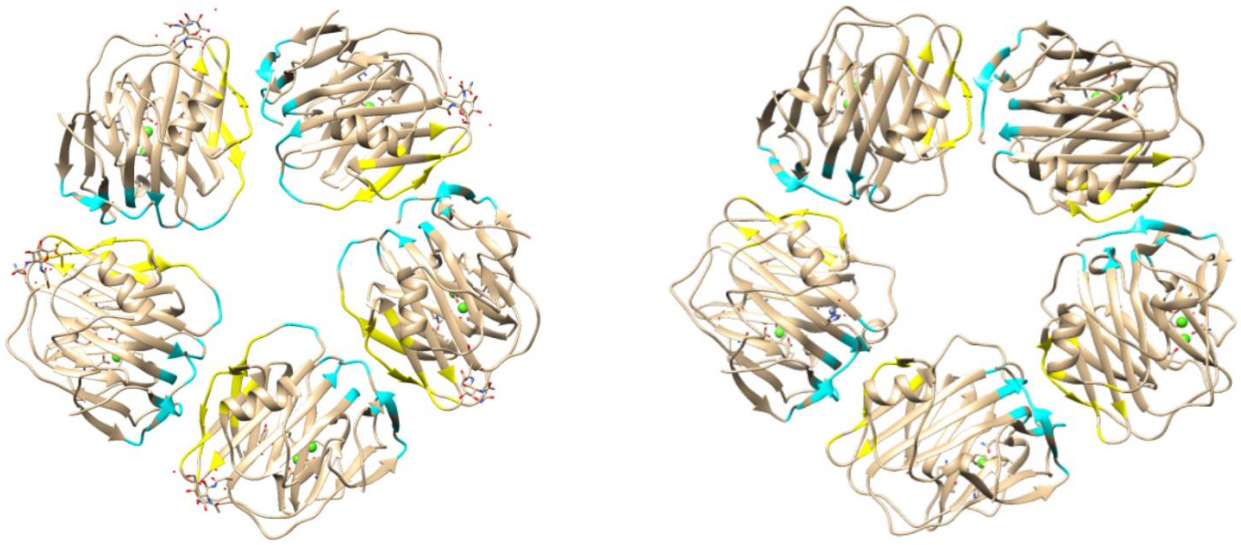


Figure 5-8A: This is the method, how I have extracted by chimera. Which is showing every single pair of atoms.



(a) SAP (b) CRP

Figure 5-8B: This number showing the major conserved residues between the interface residues. (a) SAP:- Right (cyan) :- V10, P12, R13, Y40, S41, D42, S44, K87, F88, P89, V151, G152, E153, G194, Y195, I197, K199, P200, V202, W203 and Left (yellow) :- V80, T81, S82, K83, V84, I85, E99, S102, G103, I104, W104, P113, L114, V115, K116, K117, G118, L119, Q121, P166. (b) CRP: - Right (cyan): - V10, P12, K13, Y40, T41, E42, P93, V153, D155, E197, F199, K201, P202, Q203, L204 and Left (yellow): - E101, S104, I106, W110, P115, R116, V117, R118, K119, S120, V165, P168.

Here, we evaluate whether residues have particular preferences for the interfaces and also comparing with protein-protein interface residues having the same solvent accessibilities. The comparison shows the trend that hydrophobic residues are preferred in surfaces. At the protein's surfaces, hydrophilic residues are also found, within loop or coil, and to the polypeptide chain providing high flexibility at these positions. Proline and glycine (hydrophilic) within a particular protein fold are often highly conserved. In opposite, when we compare interfaces with the overall residues, this trend is not observed. The result shows that more hydrophobic residues have the SAP interfaces are comparatively with CRP and fewer hydrophilic residues of both. The results also show that more conserved are the interfaces and for the maintenance of protein-protein interactions conserved interfaces residues are crucial.

5.9 Interface residues sequence alignment

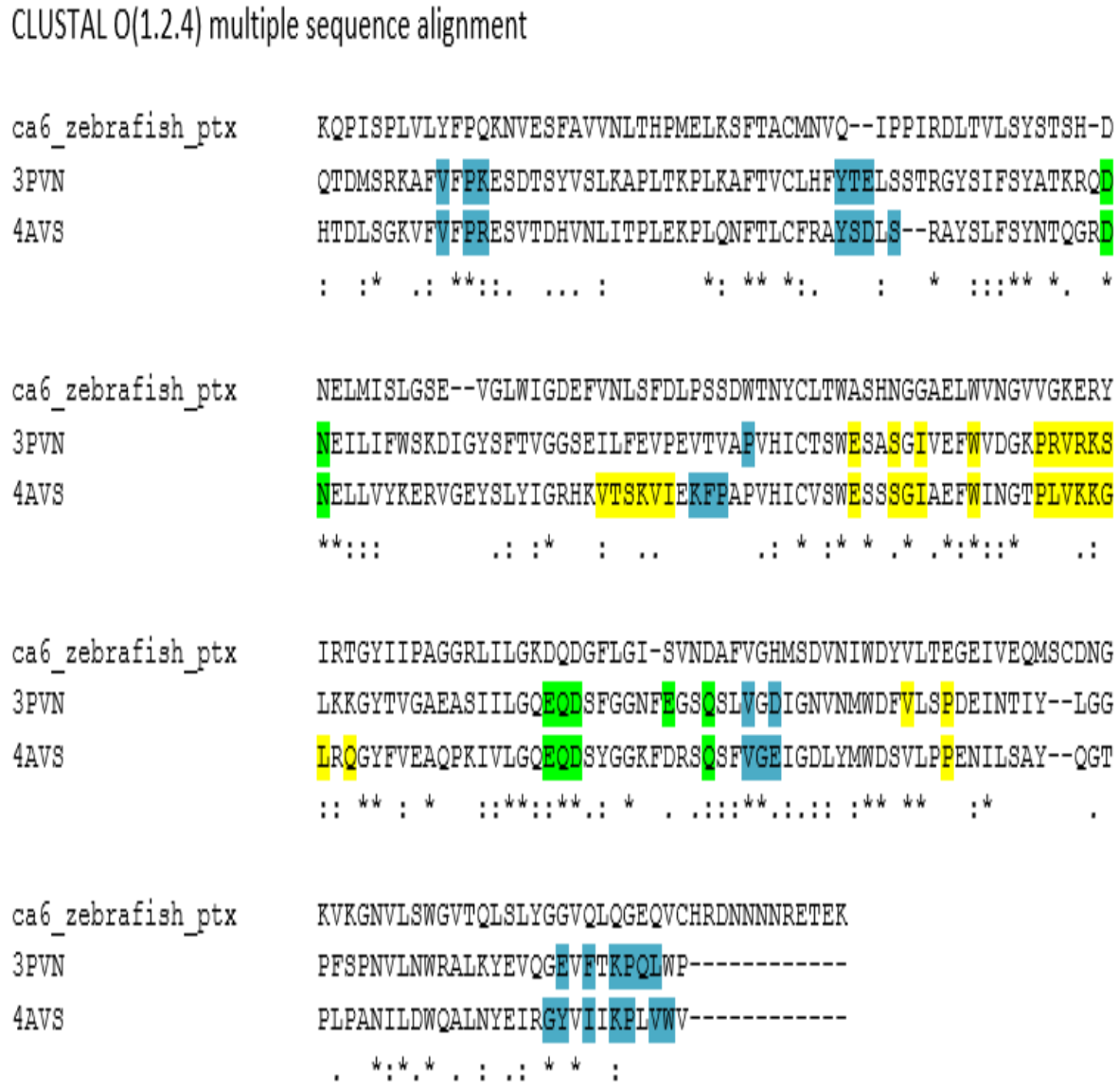


Figure 5-9: Sequence alignment of SAP (4AVS), CRP (3PVN) and ca6_zebrafish_ptx (pentraxins).

Green color showing calcium-binding residues, cyan showing right, and yellow showing left side interface residues (Figure 5-8B). Above the sequence alignment (Figure 5-9), strongly conserved residues are indicated with asterisks, dots indicating the residue variation occurring within the weaker conserved residue groups, and colons indicating the residue variation occurring within the strongly conserved groups.

5.10 Conserved interface residues of SAP and CRP

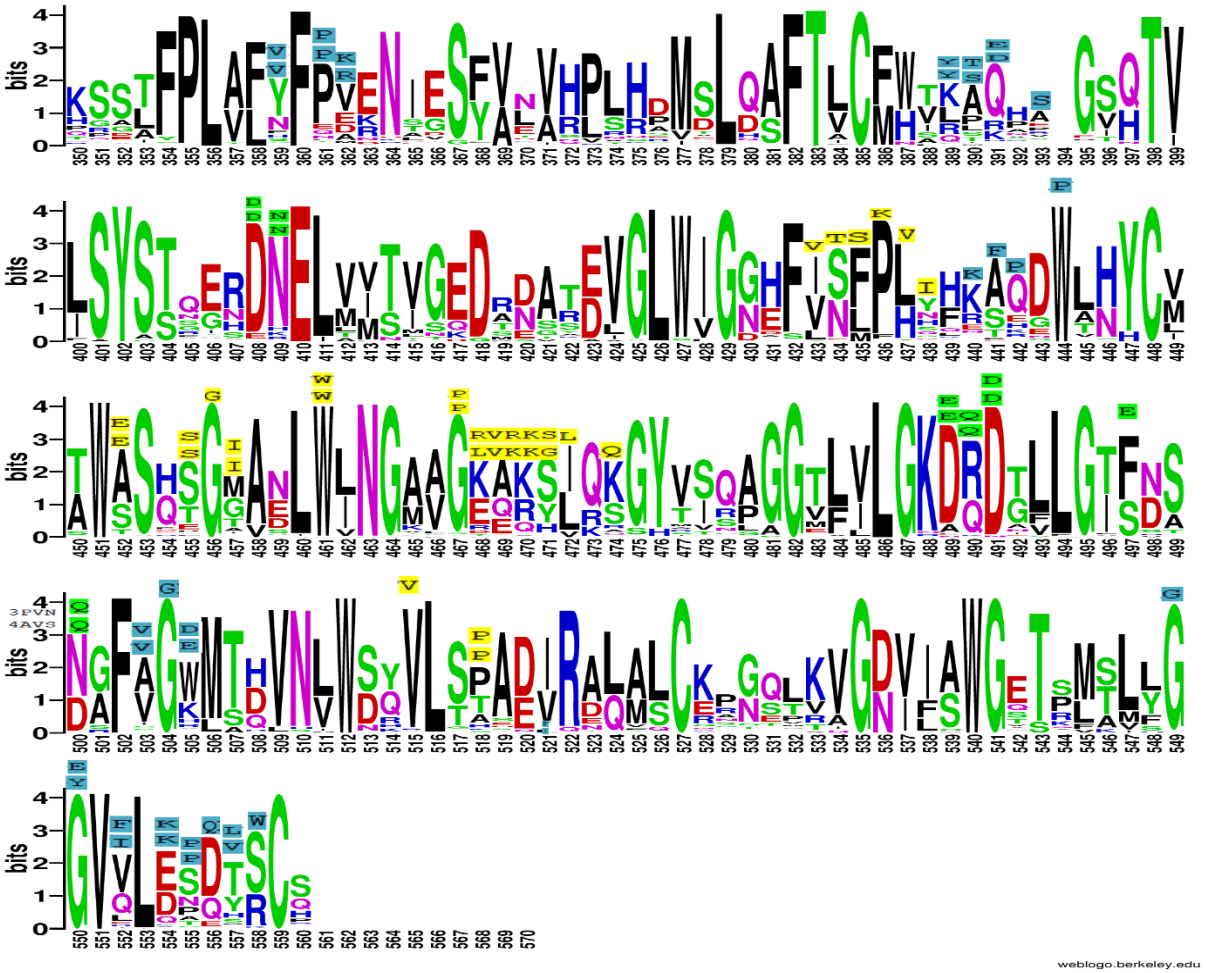
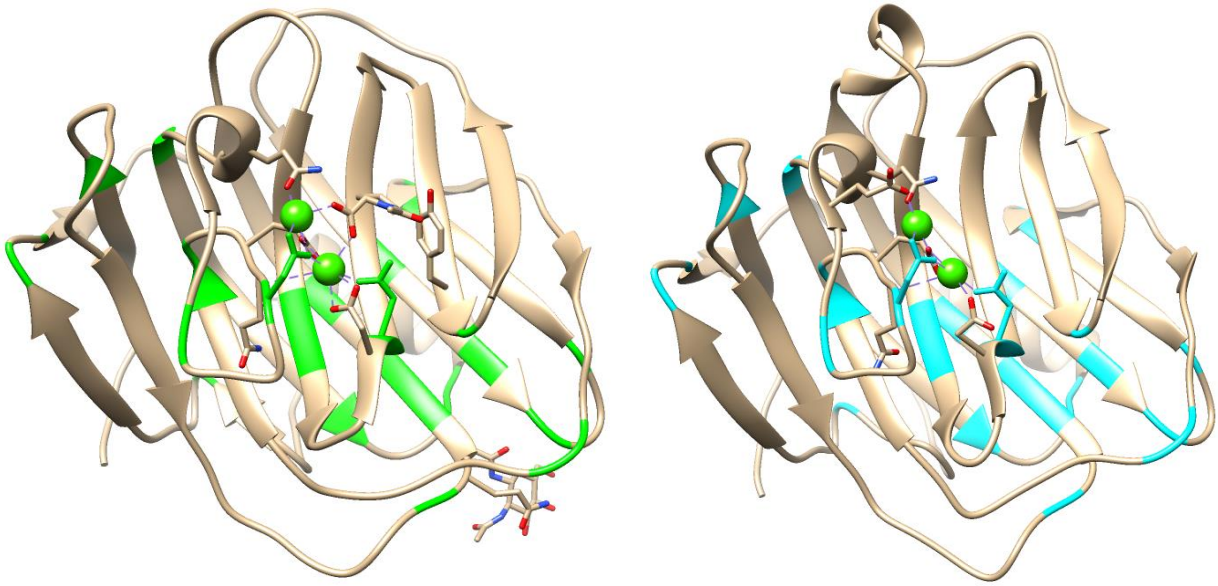


Figure 5-10: Sequence logo of SAP and CRP interface conserved residues. Top on the sequence logo, all the single letters are taken from the sequence alignment of SAP and CRP, which is mention in Figure 5-9. The color of the single letter above the logo and their colors: green; Ca-binding site, yellow and blue two sides of the interface subunit.

Locating the interface residues indicates that all interface residues are not highly conserved, whereas the calcium-binding sites are quite conserved. W (461) is conserved; it is one of the conserved multiple van der Waals; it is the centre of those van der Waals. G (550) shows fully conserved but D (408, 491) is not fully conserved because one tiny letter is down, and it is also showing calcium-binding residues. P (436) shows conserved but is not fully conserved because one tiny letter is down. Many parts of the interface are not strongly conserved. For instance, logo positions 468 to 471 are one of the less conserved regions. When we look at the letters that are in SAP and CRP, they are not necessarily. Even among the most common letters region of this logo.

So, the Ca-binding residues are much more strongly conserved than the interface residues, but there are some like G (456) (tryptophan), which are quite conserved. That interfaces are not similar between SAP and CRP and looking at the sequence alignments, that interface regions sequence is not strongly conserved. Therefore, it could be hard to predict the interface for that ca6-associated pentraxins.

5.11 Comparing single chain of SAP and CRP conserved residues



SAP conserved residue (green) subunits CRP conserved residue (cyan) subunits

Figure 5-11: In SAP: - β sheet of both structure and α helix all are close to the same position and in CRP: - Conserved residues even middle of the β sheet are same.

Normal visualization conserved (mavConservation: the conservation values from Multalign viewer of Chimera will be the residue attribute name mavConservation) residues in both SAP and CRP showing same. The conserved residues were found to be forming hydrogen bonds or hydrophobic interactions with the other conserved residues. Around the calcium-binding site, most structurally conserved residues are sequentially conserved in both structures. In contrast, less sequentially conserved residues are observed in the surface area, and the conserved polar residues are higher than the conserved hydrophobic residues. The majority of conserved residues are polar and hydrophobic. Some of the charged residues also exist. Comparing the beta-sheet of both structure and alpha-helix, all are close to the same position conserved residues even middle of the beta stand. Arrowhead (many of the conserved residues end of the β-strand, which is the

arrowhead). Many of the conserved residue beginning of the β-strand, some of them in loops, and two of them Ca-binding side, which are the same positions in conserved residues.

5.12 Count of the SAP and CRP conserved residues

Res	no	cons	mavCsrv	areaSAS	max ASA	R asa
PHE	11	F	0,995	0,332	240	0,001
LEU	30	L	0,975	3,491	201	0,017
PHE	33	F	0,951	0,000	240	0,000
THR	34	T	0,990	0,299	172	0,002
CYS	36	C	0,995	0,000	167	0,000
SER	51	S	0,975	0,000	155	0,000
TYR	52	Y	0,985	0,011	263	0,000
ASN	59	N	0,921	0,000	193	0,000
GLU	60	E	0,916	0,000	223	0,000
GLY	76	G	0,975	0,000	104	0,000
CYS	95	C	0,995	0,446	167	0,003
TRP	98	W	0,042	0,000	285	0,000
SER	100	S	0,990	0,000	155	0,000
GLY	103	G	0,990	0,000	104	0,000
TRP	108	W	0,980	13,771	285	0,048
GLY	111	G	0,941	0,000	104	0,000
GLY	122	G	0,926	8,853	104	0,085
TYR	123	Y	0,946	8,272	263	0,031
LEU	133	L	0,941	0,307	201	0,002
GLY	134	G	0,985	0,000	104	0,000
ASP	138	D	0,970	24,968	195	0,128
GLY	142	G	0,985	1,012	104	0,010
GLY	152	G	0,042	10,164	104	0,098
TRP	160	W	0,995	0,000	285	0,000
VAL	163	V	0,946	4,507	174	0,026
LEU	164	L	0,911	0,000	201	0,000

Row Labels	Count of Res
ASN	1
ASP	1
CYS	2
GLU	1
GLY	8
LEU	3
PHE	2
SER	2
THR	1
TRP	4
TYR	2
VAL	2
Grand Total	29

Count of the SAP conserved residues.

Res	no	cons	mavCsrv	areaSAS	max ASA	rASA
PHE	11	F	0,995	2,791	240	0,012
LEU	30	L	0,975	10,865	201	0,054
PHE	33	F	0,951	0,000	240	0,000
THR	34	T	0,990	0,478	172	0,003
CYS	36	C	0,995	0,017	167	0,000
SER	53	S	0,975	0,588	155	0,004
TYR	54	Y	0,985	5,892	263	0,022
ASN	61	N	0,921	35,059	193	0,182
GLU	62	E	0,916	5,194	223	0,023
GLY	78	G	0,975	9,328	104	0,090
CYS	97	C	0,995	0,486	167	0,003
TRP	100	W	0,042	0,014	285	0,000
SER	102	S	0,990	10,479	155	0,068
GLY	105	G	0,990	0,000	104	0,000
TRP	110	W	0,980	22,011	285	0,077
GLY	113	G	0,941	21,596	104	0,208
GLY	124	G	0,926	64,171	104	0,617
TYR	125	Y	0,946	63,648	263	0,242
LEU	135	L	0,941	0,090	201	0,000
GLY	136	G	0,985	1,765	104	0,017
ASP	140	D	0,970	64,075	195	0,329
GLY	144	G	0,985	14,164	104	0,136
GLY	154	G	0,042	16,044	104	0,154
TRP	162	W	0,995	6,936	285	0,024
VAL	165	V	0,946	49,634	174	0,285
LEU	166	L	0,911	4,243	201	0,021

Row Labels	Count of Res
ASN	1
ASP	1
CYS	2
GLU	1
GLY	8
LEU	3
PHE	2
SER	2
THR	1
TRP	4
TYR	2
VAL	2
Grand Total	29

Count of the CRP conserved residues.

Figure 5-12: Count of the SAP and CRP conserved residues with crp_sap_ca6 associated pentraxin full MSA.

The counted conserved residues in both 29 but the fact that the conserved residues in a similar position and conserved between SAP and CRP. The color showing the conservation residues, and SAP has a lot more conservation, which is buried, and here we have some that are surface. We have TRP 108 (SAP), which is almost 5% exposed, and here we have TRP 110 (CRP), which is 7.7% exposed, so it is the borderline case here, and ASP 138 (SAP), which is exposed 12% but ASP 140 (CRP) exposed 32% the difference is larger, we have a GLY 111 (SAP) which is not

exposed all and GLY 113, 20% exposed here (CRP). SAP seems to have more contact structure because many of these important residues are totally buried, at least 95% buried.

5.13 SAP, CRP and CRP full MSA conservation

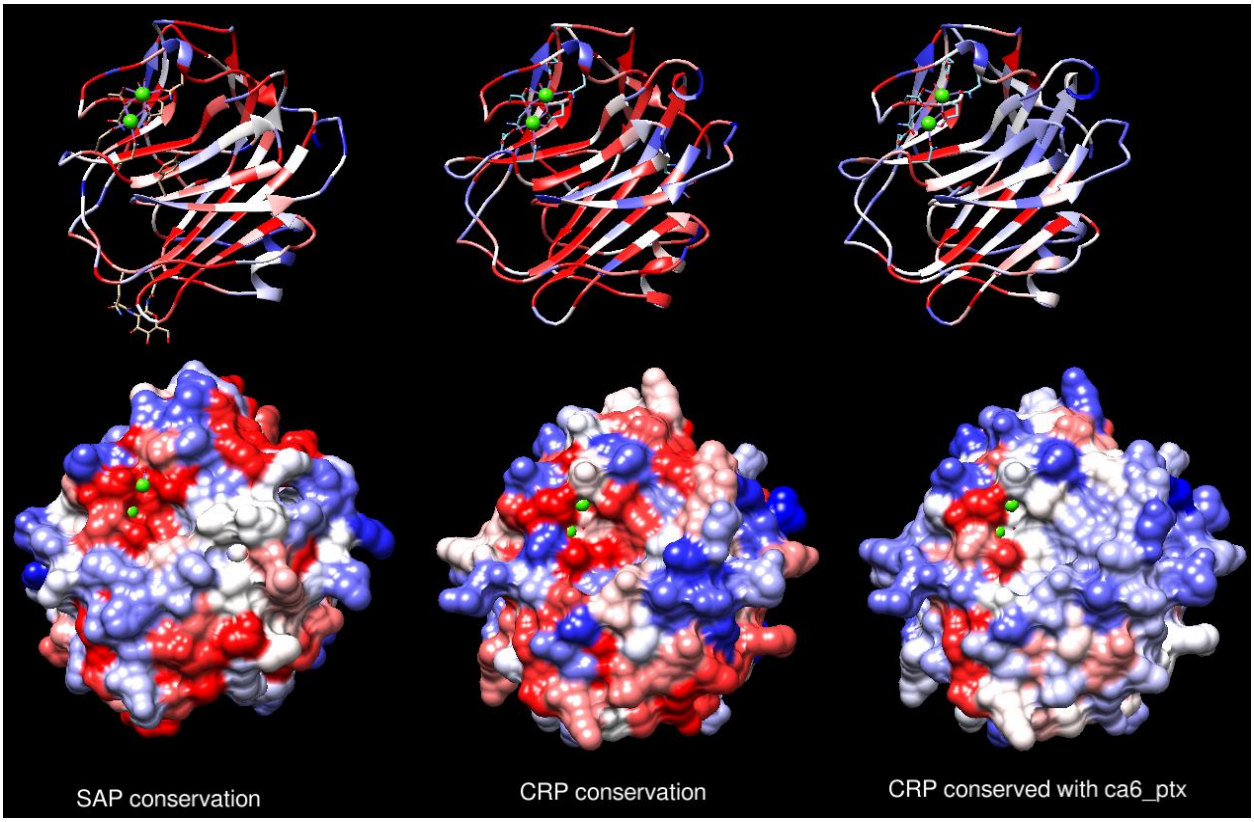


Figure 5-13: Coloring of SAP and CRP with mammals and CRP with non-mammalian-ca6 associated pentraxin domain full MSA conservation.

When we have done CRP conservation and SAP conservation here, they do have quite many conserved residues, but when we make a combined conservation with CRP and SAP and ca6 associated with pentraxins so then there are not many strongly conserved residues. So that the red is strongly conserved residues, blue is poorly conserved, and white is average conserved. The other thing is here which we wanted to see, the surface position that would be like similar in CRP and SAP which might be responsible for the interface interaction, the subunit interfaces but we can do it in details in many ways, and it does the answer is just there maybe two residues or contact with conserved residues.

5.14 SAP and CRP residues mavConservation

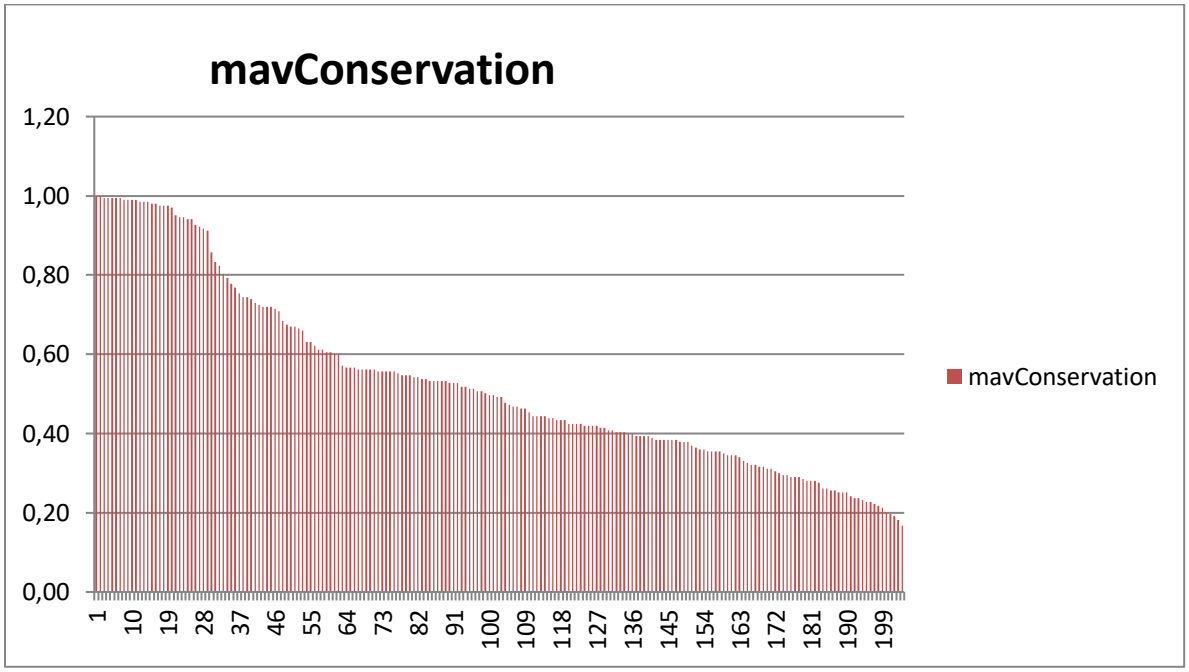


Figure 5-14A: Statistical analysis of SAP is the sorted values of mavConservation from MSA of CRP, SAP and CA6 associated pentraxin.

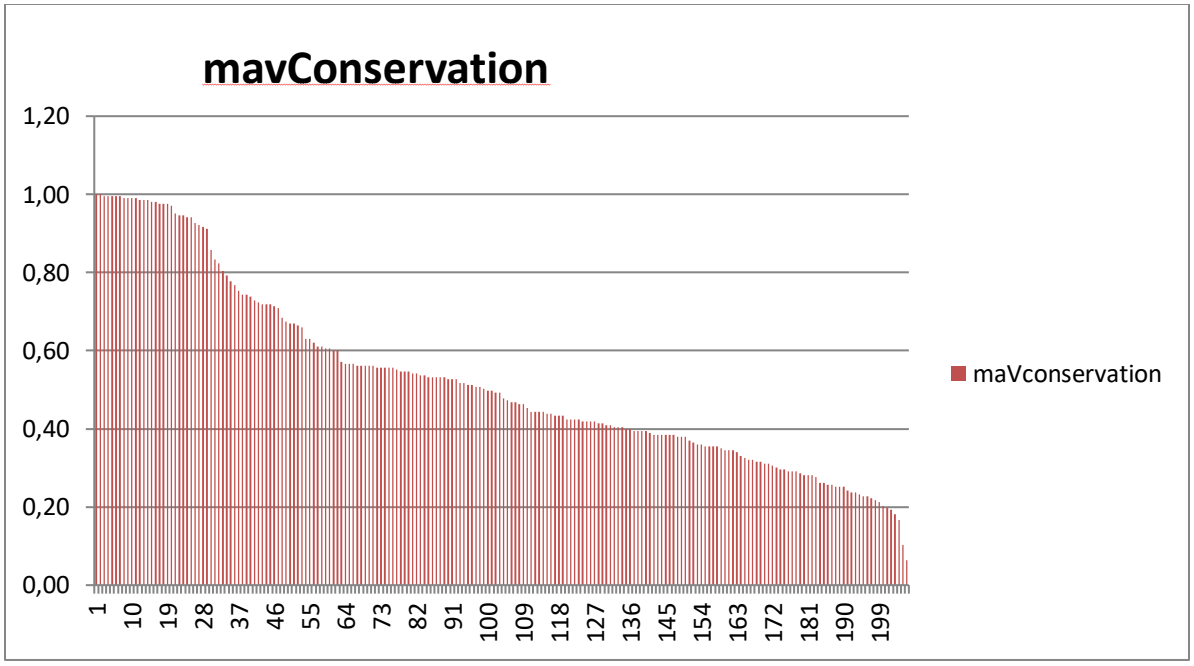


Figure 5-14B: Statistical analysis of CRP is the sorted values of mavConservation from MSA of CRP, SAP and CA6 associated pentraxin.

The specific result of conserved values that I have done by chimera and these values are calculated based on the alignment. The graph (Figure: 5-14A and 5-14B) showing SAP (conserved residues 207) and CRP (conserved residues 205) conservation are higher and which are not many conserved residues share between the two (the higher conservation of SAP and CRP, residues comparing from the sorted values of excel sheets). If we observed most of the highly conserved core residues, they are something that are involved in protein folding and keeping the basic beta-sheet structure together.

5.15 SAP and CRP sequence conservation with pentraxins

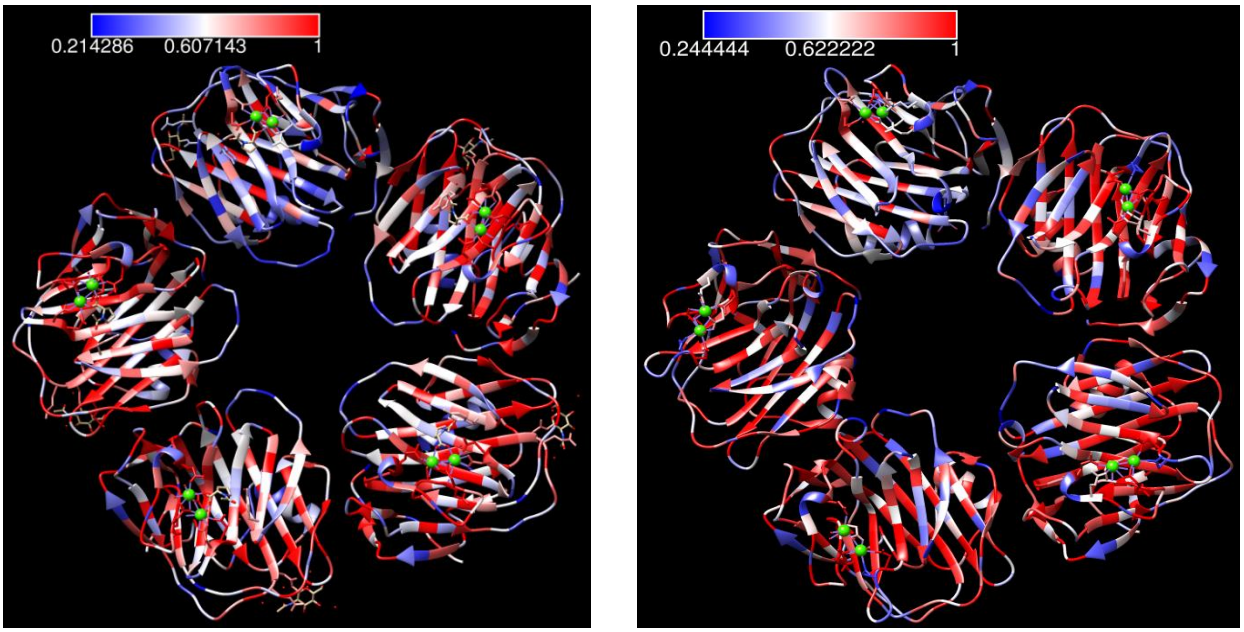


Figure 5-15: Colouring of SAP (left) and CRP (right) sequence conservation with pentraxins. Highly conserved residues are indicated by red color.

SAP- Normal visualization of sequence conservation, some conserved residues near the interface, and most of the red-dark color residues in the β sheets. Another one CRP, is color by combined alignment, and few are highly conserved residues. So that the red is strongly conserved residues, blue is poorly conserved, and white is average conserved.

5.16 Ca^{2+} binding site SAP vs. CRP

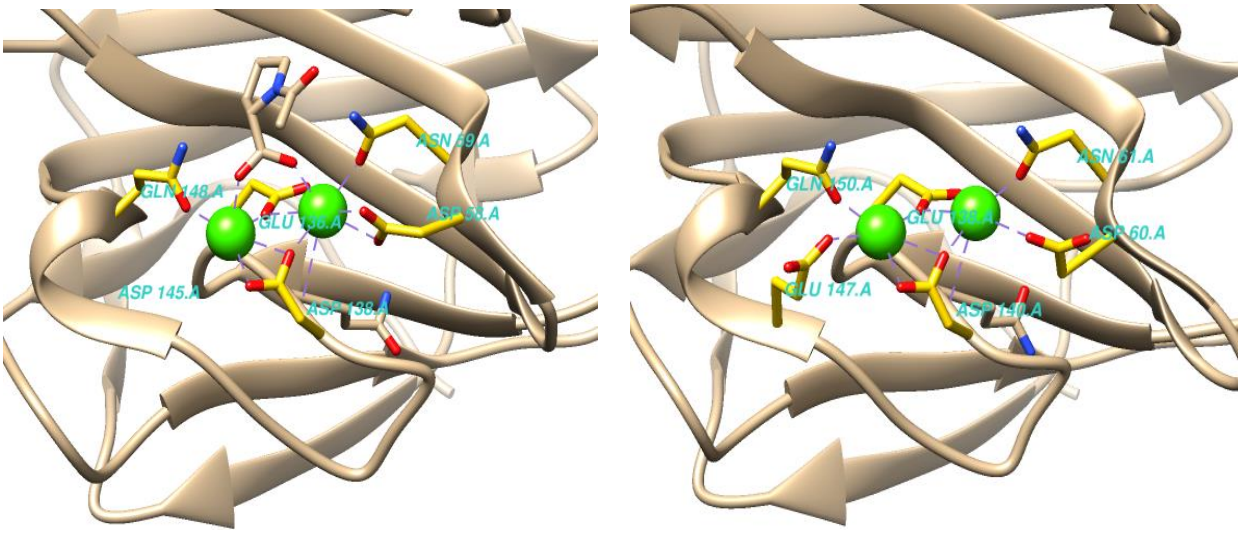


Figure 5-16: Ca^{2+} binding site of SAP Ca^{2+} binding site of CRP

The ribbon diagram of SAP and CRP, in which the two Ca^{2+} atoms are presented (spheres). In SAP structures, the calcium-binding sites are organized in a similar mode to those in CRP. The different number of protein ligands to the Ca^{2+} ions is helping to illustrate known differences in the Ca^{2+} rapport within the proteins. In Protein Data Bank most of the Ca^{2+} binding sites are non-linear, which means the Ca^{2+} coordinating ligands may come from different subunits in a protein or different loops or even several proteins. Ca^{2+} trend to precipitate both organic and inorganic anions. Divination of those Ca^{2+} binding sites requires three-dimensional accumulated in from homology modeling.

The first calcium ion in the SAP structure is coordinated by Asp58, Asn59, Glu136, Asp138, and the second calcium by Glu136, Asp138, Asp145, Gln148 residues (yellow). Similarly, In the first calcium ion in the CRP structure is coordinated by Asp60, Asn61, Glu138, Asp140, and the second calcium by Glu138, Asp140, Glu147, Gln150 residues in a loop. In both structures (Figure 5-16), visible here, Ca^{2+} ions coordinating residues are the same.

Ca^{2+} binding sites do not constantly follow a set of linear amino acid sequences. In Protein Data Bank approximately 90% of known Ca^{2+} binding sites are non-linear, which means the Ca^{2+} ions coordinating ligands may come from several subunits in a protein, or different loops, or even different proteins, and this is also the case in the models studied in this thesis. The Ca^{2+} binding

sites require three-dimensional structural information from either determined structure deposited in PDB or homology modeling. Principal structural factors of Ca^{2+} binding share some ordinary features defined with charge distribution, ligand types, bond angles, Ca^{2+} ligand distances, and geometric configurations (Kirberger M *et al.,* 2008, Yang W *et al.,* 2002, Pidcock E *et al.,* 2001).

5.17 Contact analysis of SAP and CRP

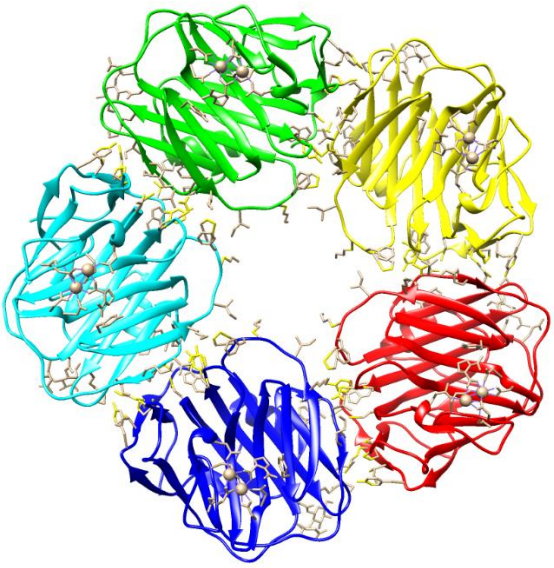
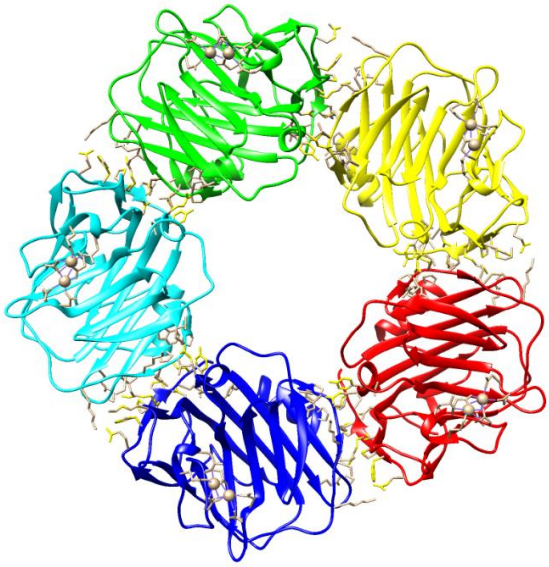


Figure 5-17: Contact analysis of SAP



Contact analysis of CRP

The first look is that the different residues and then the because of these loops here we have additional contact(SAP), which are missing here (CRP) because loops are far away and this is the main difference between the SAP and CRP. That might be related to the building of subunits.

As the detailed view shows, there are some residues that are conserved. Some hydrophobic residues are in the same positions and one of the ionic contacts. The amount of contact or amount of contact residues is similar, only that the binding mode is very different. So one of the results shows that the mode of Ca6 associated pentraxins would be based on the SAP and CRP because of the difference between each other.

5.18 Contact of SAP and CRP in a single interface

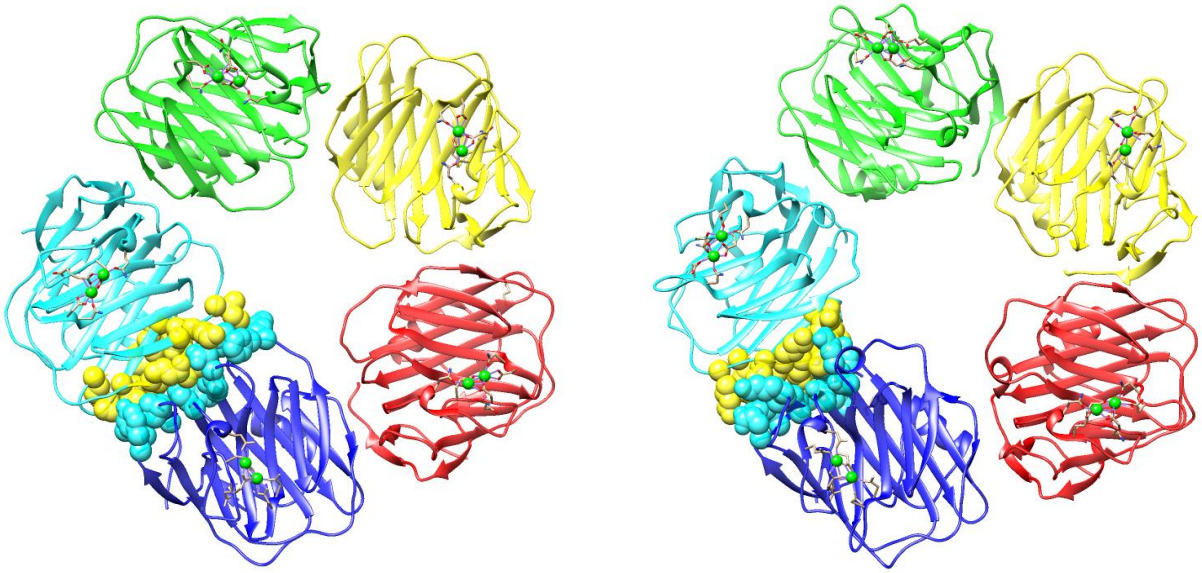


Figure 5-18: Contact analysis of single interface SAP Contact analysis of single interface CRP

The picture showing the interface contact residues sphere is nicely touching both structures, and this is the one with a large surface area because of the loop coming near to the next subunits. It is helpful to see the bigger surface area in the SAP subunit and CRP subunit.

At the interface, residues in the protein-protein complex determine a significant character providing an accurate level of stability and specificity to the association between proteins and their specific 3-D structure (Ryan DP and Matthews JM 2005). For example, Interactions in the protein structure include van der Waals bond, salt bridge, and hydrogen bond into the protein and between the bound proteins (Tna KG *et al.,* 2007). Analyze the protein-protein interface residues for their inter and intra-protein interactions. Two types of interactions were considering every interfacial residue. The first category of interactions over the protein-protein interfaces and the second category mentions two types of interactions made by a side chain; first interaction across the residue with a side chain or main chain atom in the same protein chain, and the second interaction across the protein-protein interfaces (Tna KG *et al.,* 2007).

5.19 Contact analysis of SAP (4AVS)

Analysis of 4pvn contract lists unic values between (ab, bc, cd, de, ea)										
unic values of 4avs (ab)	ab	unic values of 4avs (bc)	bc	unic values of 4avs (cd)	cd	unic values of 4avs (de)	de	unic values of 4avs (ea)	ea	Totall interface residues
ASP 42.A LYS 117.B	4	ASP 42.B LYS 117.C	4	ASP 42.C LYS 117.D	4	ASP 42.D LYS 117.E	4	ASP 42.E LYS 117.A	4	20
ASP 42.A LYS 83.B	2	ASP 42.B LYS 83.C	2	ASP 42.C LYS 83.D	2	ASP 42.D LYS 83.E	2	ASP 42.E LYS 83.A	2	10
ASP 42.A SER 82.B	2	ASP 42.B SER 82.C	2	ASP 42.C SER 82.D	4	ASP 42.D SER 82.E	2	ASP 42.E SER 82.A	3	13
ASP 42.A VAL 115.B	1	ASP 42.B VAL 115.C	1	ASP 42.C VAL 115.D	1	ASP 42.D VAL 115.E	1	ASP 42.E VAL 115.A	1	5
GLU 153.A LYS 116.B	7	GLU 153.B LYS 116.C	7	GLU 153.C LYS 116.D	6	GLU 153.D LYS 116.E	7	GLU 153.E LYS 116.A	7	34
GLY 152.A VAL 115.B	1	GLY 152.B VAL 115.C	1	GLY 152.C VAL 115.D	1	GLY 152.D VAL 115.E	1	GLY 152.E VAL 115.A	1	5
ILE 197.A SER 102.B	1	ILE 197.B SER 102.C	1	ILE 197.C SER 102.D	1	ILE 197.D SER 102.E	1	ILE 197.E SER 102.A	1	5
		ILE 197.B ILE 104.C	1							1
LYS 199.A GLU 99.B	3			LYS 199.C GLU 99.D	3	LYS 199.D GLU 99.E	3	LYS 199.E GLU 99.A	3	12
		LYS 199.B GLU 99.C	3							3
LYS 199.A ILE 104.B	2	LYS 199.B ILE 104.C	1	LYS 199.C ILE 104.D	2	LYS 199.D ILE 104.E	1	LYS 199.E ILE 104.A	1	7
LYS 199.A LYS 116.B	2	LYS 199.B LYS 116.C	2	LYS 199.C LYS 116.D	2	LYS 199.D LYS 116.E	2	LYS 199.E LYS 116.A	2	10
LYS 87.A ILE 85.B	1	LYS 87.B ILE 85.C	1	LYS 87.C ILE 85.D	1	LYS 87.D ILE 85.E	1	LYS 87.E ILE 85.A	1	5
PHE 88.A ILE 85.B	2	PHE 88.B ILE 85.C	2	PHE 88.C ILE 85.D	2	PHE 88.D ILE 85.E	2	PHE 88.E ILE 85.A	2	10
PRO 12.A GLY 118.B	3	PRO 12.B GLY 118.C	3	PRO 12.C GLY 118.D	3	PRO 12.D GLY 118.E	3	PRO 12.E GLY 118.A	3	15
PRO 12.A ILE 104.B	2	PRO 12.B ILE 104.C	2	PRO 12.C ILE 104.D	2	PRO 12.D ILE 104.E	2	PRO 12.E ILE 104.A	2	10
PRO 12.A LYS 117.B	2	PRO 12.B LYS 117.C	2	PRO 12.C LYS 117.D	2	PRO 12.D LYS 117.E	2	PRO 12.E LYS 117.A	2	10
PRO 89.A ILE 85.B	1	PRO 89.B ILE 85.C	1	PRO 89.C ILE 85.D	1	PRO 89.D ILE 85.E	1	PRO 89.E ILE 85.A	1	5
PRO 89.A LYS 83.B	2	PRO 89.B LYS 83.C	3	PRO 89.C LYS 83.D	1	PRO 89.D LYS 83.E	2	PRO 89.E LYS 83.A	2	10
SER 41.A VAL 115.B	2	SER 41.B VAL 115.C	3	SER 41.C VAL 115.D	3	SER 41.D VAL 115.E	3	SER 41.E VAL 115.A	3	14
TRP 203.A PRO 113.B	2	TRP 203.B PRO 113.C	2	TRP 203.C PRO 113.D	2	TRP 203.D PRO 113.E	2	TRP 203.E PRO 113.A	2	10
TYR 195.A GLY 103.B	1	TYR 195.B GLY 103.C	1	TYR 195.C GLY 103.D	1	TYR 195.D GLY 103.E	1	TYR 195.E GLY 103.A	1	5
TYR 195.A GLY 118.B	1	TYR 195.B GLY 118.C	1	TYR 195.C GLY 118.D	1	TYR 195.D GLY 118.E	1	TYR 195.E GLY 118.A	1	5
TYR 195.A SER 102.B	3	TYR 195.B SER 102.C	3	TYR 195.C SER 102.D	3	TYR 195.D SER 102.E	2	TYR 195.E SER 102.A	3	14
TYR 40.A LEU 114.B	2	TYR 40.B LEU 114.C	2	TYR 40.C LEU 114.D	2	TYR 40.D LEU 114.E	2	TYR 40.E LEU 114.A	2	10
TYR 40.A PRO 113.B	2	TYR 40.B PRO 113.C	2	TYR 40.C PRO 113.D	2	TYR 40.D PRO 113.E	2	TYR 40.E PRO 113.A	2	10
TYR 40.A VAL 115.B	8	TYR 40.B VAL 115.C	8	TYR 40.C VAL 115.D	8	TYR 40.D VAL 115.E	8	TYR 40.E VAL 115.A	8	40
VAL 10.A ILE 104.B	1	VAL 10.B ILE 104.C	1	VAL 10.C ILE 104.D	1	VAL 10.D ILE 104.E	1	VAL 10.E ILE 104.A	1	5
VAL 10.A LYS 116.B	2	VAL 10.B LYS 116.C	2	VAL 10.C LYS 116.D	2	VAL 10.D LYS 116.E	2	VAL 10.E LYS 116.A	2	10
		VAL 151.B LYS 117.C	2							2
VAL 202.A LYS 116.B	1	VAL 202.B LYS 116.C	1	VAL 202.C LYS 116.D	1	VAL 202.D LYS 116.E	1	VAL 202.E LYS 116.A	1	5
VAL 202.A PRO 113.B	2	VAL 202.B PRO 113.C	2	VAL 202.C PRO 113.D	2	VAL 202.D PRO 113.E	2	VAL 202.E PRO 113.A	2	10
VAL 202.A PRO 166.B	1	VAL 202.B PRO 166.C	2	VAL 202.C PRO 166.D	1	VAL 202.D PRO 166.E	1	VAL 202.E PRO 166.A	1	6
VAL 202.A TRP 108.B	2	VAL 202.B TRP 108.C	3	VAL 202.C TRP 108.D	3	VAL 202.D TRP 108.E	2	VAL 202.E TRP 108.A	3	13
	68		74		70		67		70	349

Figure 5-19A: Analysis of SAP (4AVS) contact list unique values between the ab, bc, cd, de, ea.

Analysis of 3pvn contract lists unic values between (ab, bc, cd, de, ea)

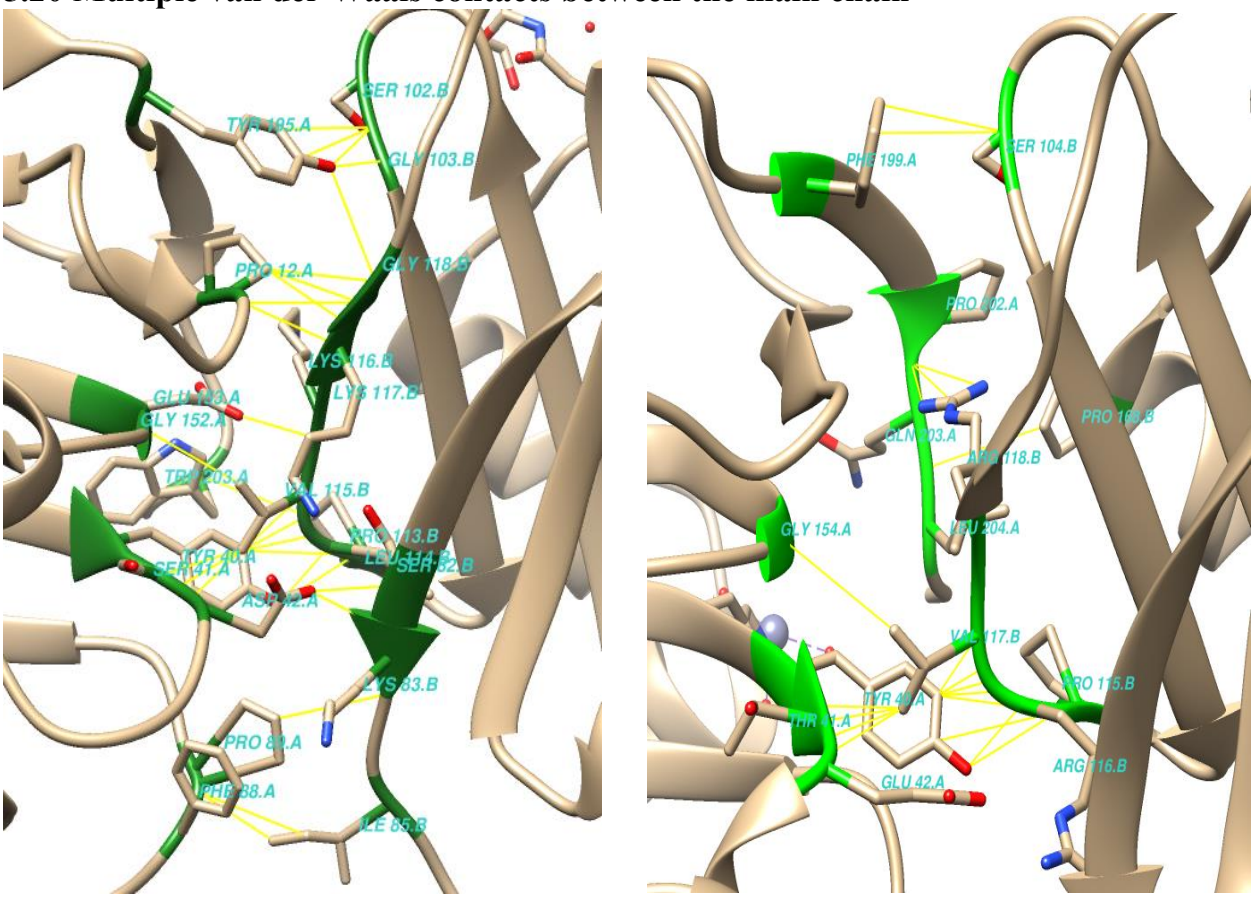
unic values of 3pvn (ab)	ab	unic values of 3pvn (bc)	bc	unic values of 3pvn (cd)	cd	unic values of 3pvn (de)	de	unic values of 3pvn (ea)	ea	Totall interface residues
				ARG 6.C ASP 169.D	2	ARG 6.D ASP 169.E	4			6
ASP 155.A ARG 118.B	5	ASP 155.B ARG 118.C	5	ASP 155.C ARG 118.D	4	ASP 155.D ARG 118.E	5	ASP 155.E ARG 118.A	1	20
GLN 203.A PRO 168.B	1			GLN 203.C PRO 168.D	1	GLN 203.D PRO 168.E	1	GLN 203.E PRO 168.A	1	4
GLU 197.A LYS 123.B	5	GLU 197.B LYS 123.C	3	GLU 197.C LYS 123.D	3	GLU 197.D LYS 123.E	5	GLU 197.E LYS 123.A	6	22
GLU 42.A ARG 116.B	1	GLU 42.B ARG 116.C	8	GLU 42.C ARG 116.D	6	GLU 42.D ARG 116.E	4	GLU 42.E ARG 116.A	3	22
				GLU 42.C GLU 85.D	1					1
GLU 42.A LYS 119.B	4							GLU 42.E LYS 119.A	3	7
GLU 42.A VAL 117.B	3	GLU 42.B VAL 117.C	7	GLU 42.C VAL 117.D	8	GLU 42.D VAL 117.E	10	GLU 42.E VAL 117.A	4	32
GLY 154.A VAL 117.B	1	GLY 154.B VAL 117.C	1	GLY 154.C VAL 117.D	1					3
								LEU 204.E ARG 116.A	1	1
LEU 204.A ARG 118.B	4	LEU 204.B ARG 118.C	5	LEU 204.C ARG 118.D	6	LEU 204.D ARG 118.E	7	LEU 204.E ARG 118.A	4	26
LEU 204.A PRO 115.B	2	LEU 204.B PRO 115.C	2	LEU 204.C PRO 115.D	2	LEU 204.D PRO 115.E	2	LEU 204.E PRO 115.A	3	11
		LEU 204.B PRO 168.C	1							1
LEU 204.A TRP 110.B	1	LEU 204.B TRP 110.C	1	LEU 204.C TRP 110.D	1	LEU 204.D TRP 110.E	1	LEU 204.E TRP 110.A	1	5
LYS 13.A SER 120.B	1					LYS 13.D SER 120.E	1			2
LYS 201.A ARG 118.B	3	LYS 201.B ARG 118.C	3	LYS 201.C ARG 118.D	3	LYS 201.D ARG 118.E	3	LYS 201.E ARG 118.A	3	15
LYS 201.A GLU 101.B	4	LYS 201.B GLU 101.C	4	LYS 201.C GLU 101.D	3	LYS 201.D GLU 101.E	3	LYS 201.E GLU 101.A	3	17
LYS 201.A ILE 106.B	2	LYS 201.B ILE 106.C	2	LYS 201.C ILE 106.D	2	LYS 201.D ILE 106.E	2	LYS 201.E ILE 106.A	2	10
PHE 199.A ILE 106.B	2	PHE 199.B ILE 106.C	3	PHE 199.C ILE 106.D	3	PHE 199.D ILE 106.E	2	PHE 199.E ILE 106.A	3	13
PHE 199.A SER 104.B	5	PHE 199.B SER 104.C	5	PHE 199.C SER 104.D	5	PHE 199.D SER 104.E	5	PHE 199.E SER 104.A	6	26
PRO 12.A ILE 106.B	1									1
PRO 12.A SER 120.B	1	PRO 12.B SER 120.C	1	PRO 12.C SER 120.D	1	PRO 12.D SER 120.E	1	PRO 12.E SER 120.A	1	5
PRO 202.A ARG 118.B	3	PRO 202.B ARG 118.C	2	PRO 202.C ARG 118.D	3	PRO 202.D ARG 118.E	3	PRO 202.E ARG 118.A	3	14
PRO 202.A PRO 168.B	3	PRO 202.B PRO 168.C	3	PRO 202.C PRO 168.D	2	PRO 202.D PRO 168.E	2	PRO 202.E PRO 168.A	3	13
THR 41.A VAL 117.B	3	THR 41.B VAL 117.C	2	THR 41.C VAL 117.D	3	THR 41.D VAL 117.E	2	THR 41.E VAL 117.A	2	12
TYR 40.A ARG 116.B	3	TYR 40.B ARG 116.C	2	TYR 40.C ARG 116.D	3	TYR 40.D ARG 116.E	2	TYR 40.E ARG 116.A	2	12
TYR 40.A PRO 115.B	3	TYR 40.B PRO 115.C	2	TYR 40.C PRO 115.D	2	TYR 40.D PRO 115.E	3	TYR 40.E PRO 115.A	3	13
TYR 40.A VAL 117.B	6	TYR 40.B VAL 117.C	7	TYR 40.C VAL 117.D	7	TYR 40.D VAL 117.E	6	TYR 40.E VAL 117.A	7	33
VAL 10.A ARG 118.B	2	VAL 10.B ARG 118.C	2	VAL 10.C ARG 118.D	4	VAL 10.D ARG 118.E	4	VAL 10.E ARG 118.A	4	16
VAL 10.A ILE 106.B	1	VAL 10.B ILE 106.C	1	VAL 10.C ILE 106.D	1	VAL 10.D ILE 106.E	1	VAL 10.E ILE 106.A	1	5
	70		72		77		79		70	368

Figure 5-19B: Analysis of CRP (3PVN) contact list unique values between the ab, bc, cd, de, ea.

One of the main objectives of the analysis of the human C-reactive protein and serum amyloid P component contact list is to identify the similarities and differences between the chain of ab, bc, cd, de, and ea. The comparisons displaying that the consistent interfaces list have hydrophobic

residues (Leu, Ile, and Val), aromatic residues (Trp, Tyr, Phe), and charged residues are (Arg, Asp, Glu, Lys). The fact that one of the highest preferences Tyr-Val contacts have indicates the important role that type of contacts have in protein-protein interactions. Between the residues of contact with opposite charges, Asp-Arg, Glu-Arg, Lys-Arg, Lys-Glu, and Glu-Lys are also preferred in interfaces. These contact lists are a consistent requirement that salt-bridges, hydrophobic interactions, and disulfide bonds represent the potential forces in protein-protein interactions. The differences (number) due to the loops we look at the contact image these loops have additional contacts (SAP) that increase the surface areas of contact.

5.20 Multiple van der Waals contacts between the main chain



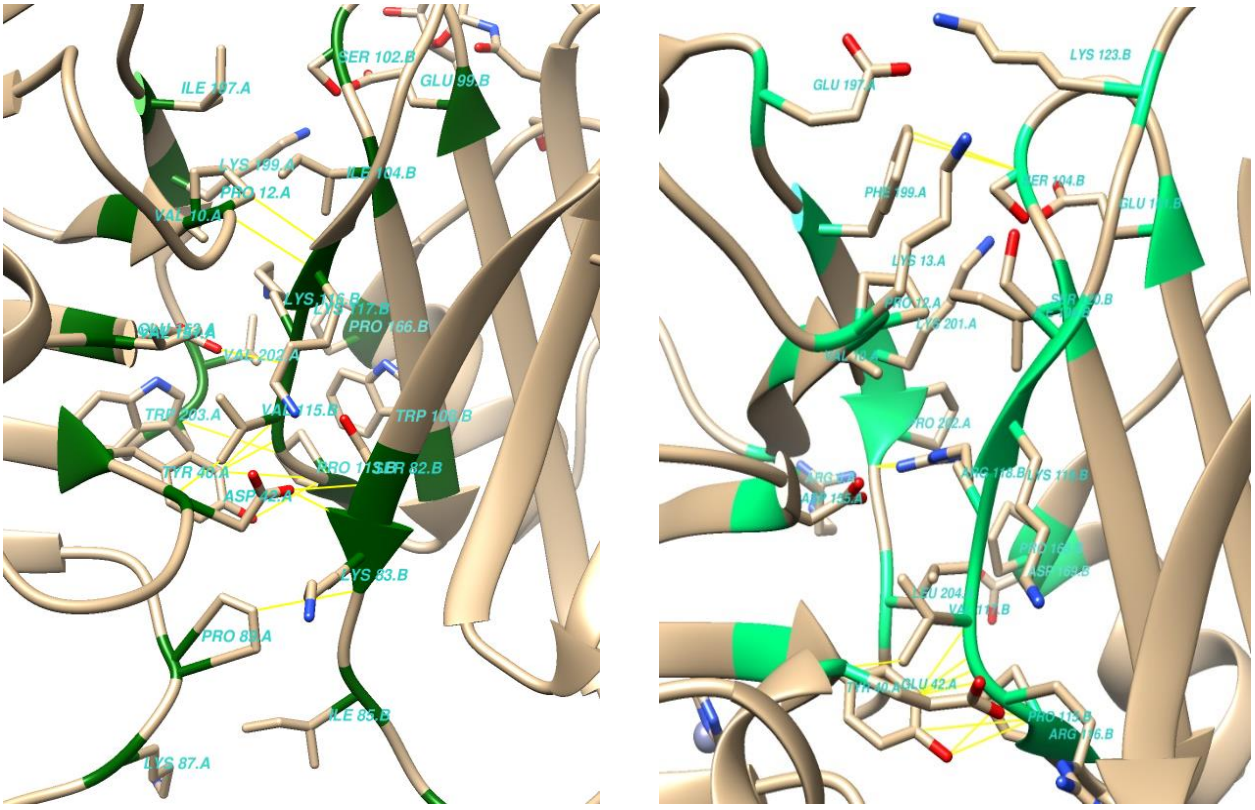
(A) (B)

Figure 5-20: (A) Multiple van der Waals contacts (4AVS) between the main chain of A-B. (B) Multiple van der Waals contacts (3PVN) between the main chain of A-B. The images-based results are comparing from different orientations.

Between the main chain and side chain of (A) SAP 4avs and (B) CRP (3pvn) are showing multiple van der Waals contact residues. The forest green of the SAP and green color of CRP represents the main chain and chain linker yellow color. Multiple van der Waals contacts between the main chain and side chain are shown in the Figure (5-20) SAP and CRP, Tyr40-Pro113 vs. Tyr40-Pro115, and Tyr40-Val115 vs. Tyr40-Val117, Tyr making multiple contacts and Gly154-Val117 vs. Gly152-Val115, Gly making a single contact in CRP and SAP main chain and side chain. In CRP, Phe199-Ser104, and Pro202- Arg118 also shows multiple van der Waals contacts between the main and side chains.

Our observation, shows some of the single contact residues in both CRP and SAP, showing single van der Waals interaction in the main and side chains.

5.21 Multiple van der Waals contacts between the side chain



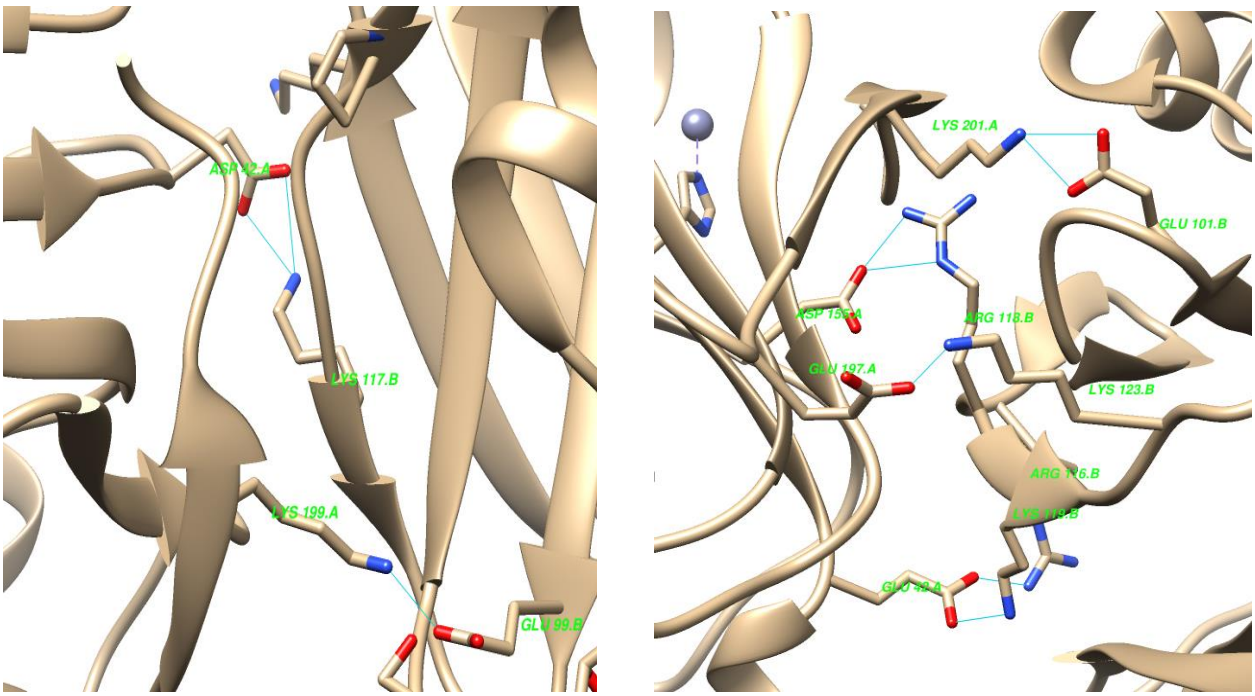
(A) (B)

Figure 5-21: (A) Multiple van der Waals contacts (4AVS) between the side chain of A-B. (B) Multiple van der Waals contacts (3PVN) between the side chain of A-B. The images-based results are comparing from different orientations.

Between the side chain and main chain of (A) SAP 4avs and (B) CRP (3pvn) are showing multiple van der Waals contact residue. The forest green of SAP and spring green of CRP represents the main chain and chain linker yellow color. Multiple van der Waals contacts between the side chain and main chain are shown in Figure (5-21) SAP and CRP, Tyr40-Val115 vs. Tyr40-Val117, Tyr making multiple contacts. But in SAP, Ser82-Asp42, Lys83-Asp42, Pro113-Tyr40 and CRP, Phe199-Ser104, Pro202-Arg118, Tyr40-Arg116, and Tyr40-Pro115, also showing the individual multiple van der Waals contacts. Our observation, shows some of the single contact residues in both CRP and SAP, showing single van der Waals interaction between the side chain and main chain.

Contacts formed between residues in 3-D protein structures are spatially pairs of close residues. Protein can be better re-constructed, with carbon-beta and carbon-alpha being a backbone atom (Duarte JM *et al.,* 2010). A pair of residues is defined as a contact if the distance between their carbon-beta (Cβ) atoms is less than or equal to 8Å (Jones DT *et al.,* 2012). Contacts are extensively categorized as long-range, medium-range, and short-range. Contacts of short-range are those differentiated by 6-11 residues in the sequence; medium-range contacts 12-23 residues and at least 24 residues in long-range contacts (Monastyrskyy B *et al.,* 2011; Monastyrskyy B *et al.,* 2014; Eickholt J *et al.,* 2013). Depending upon the 3-D structure of the protein (fold), some protein has many short-range contacts, while other protein has more long-range contacts, as shown in Figure 5-21 (A and B). Multiple van der Waals contacts (4AVS, 3PVN) between the main chain and side chain of A-B.

5.22 Ionic hydrogen bonds comparison in the pentamer model

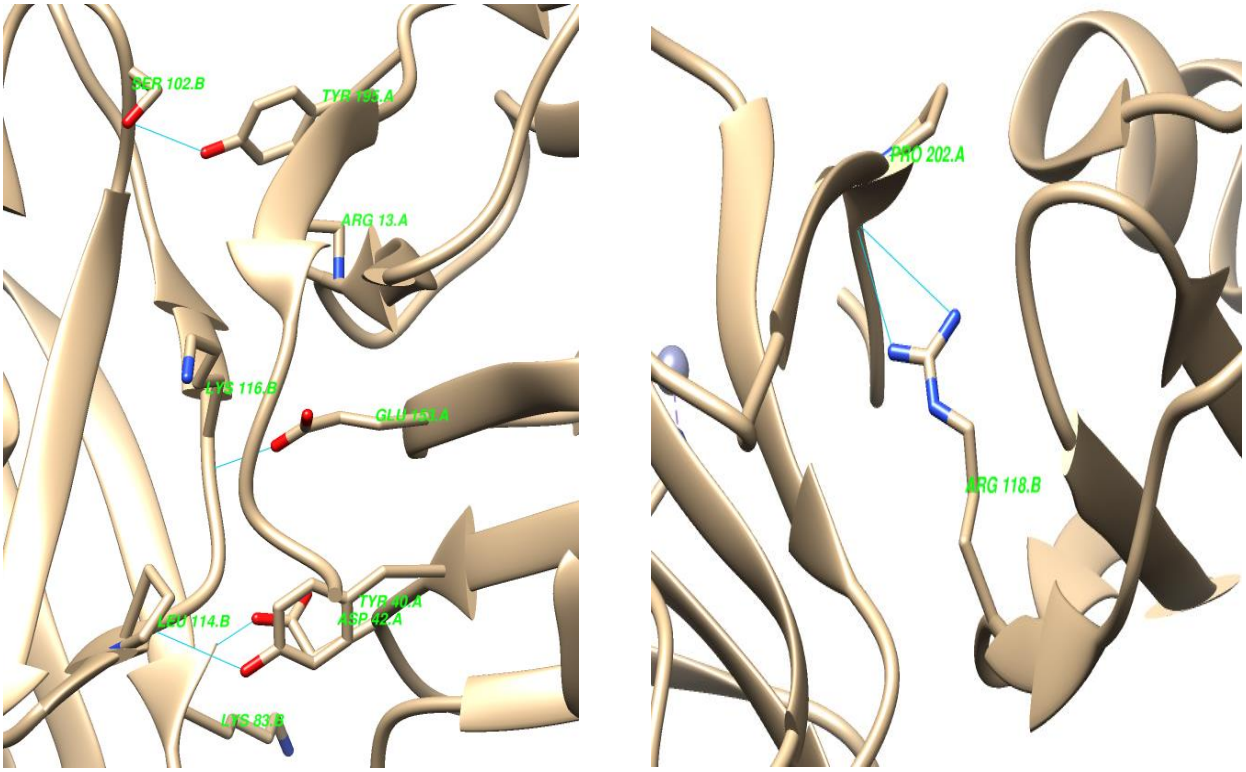


(A) (B)

Figure 5-22: (A) Ionic hydrogen bonds comparison SAP (4avs) in the pentamer model. (B) Ionic hydrogen bonds comparison CRP (3PVN) in the pentamer model.

The important thing is that there are quite differences between SAP and CRP. In CRP, all of the bonds are ionic, but only a few are ionic in SAP because most of the bonds involve the main chain. Ionic hydrogen bonds are stronger because it is not only the hydrogen bonds; there is also an attraction between the positive and negative charge. That is why we wanted to see it separately, and many of the interfaces are complimentary charge is an important part, and there are so many contacts. In Figure (5-22), SAP and CRP, Asp42-Lys117, Glu99- Lys199, and Glu101- Lys201, Asp155-Arg118, Glu197-Lys123, Glu42-Arg116, Glu42-Lys119 are showing hydrogen bonds with different amino acids, for a hydrogen bond to be formed in protein when two electronegative atoms (alpha-helix and carbonyl group) have to interact with the same hydrogen. In proteins, all groups are capable of forming H-bond. Salt bridges are also formed by positively and negatively charged between the amino acids and also important for stabilizing 3-D protein structure.

5.23 Non-ionic hydrogen bonds comparison in the pentamer model



(A) (B)

Figure 5-23: (A) Non-ionic hydrogen bonds comparison SAP (4avs) in the pentamer model. (B) Non-ionic hydrogen bonds comparison CRP (3PVN) in the pentamer model.

Non-ionic hydrogen bond positive/negative binds to the main chain or neutral side chain. Tyr (negative) binds to the main chain carbonyl. The main thing is that the residues are conserved or not. Just for classifying the kind of contacts, we have non-ionic hydrogen bonds that are most important, and even van der Walls contacts also have non-ionic character. The fact that there are few bonds in some of the images. Interestingly, most of the bonds are ionic, and there are not that many hydrogen bonds to the main chains, whereas this has plenty of these (SAP) non-ionic bonds; Ser102-Tyr195, Lys116-Glu153, Leu114-Tyr40, and Lys83-Asp42.

CRP- (multiple contacts) both of the atoms are close enough to something here to the threshold to the contract. Here we look at the residue level if I have the corresponding residues making the contract, so then it is conserved, and this is the contract with the main chain separately (Pro202-Arg118). Both the pentamer model also shows some polar and hydrophobic residues forming a non-ionic hydrogen bond. In van der Waals interactions, hydrophobic residues participate in non-

ionic character and contribute to the protein cores stabilization. My results show that van der Waals contacts are also a part of stabilizing the interfaces in CRP and SAP pentamers.

6 Conclusion

Our results are based on the visualizing comparisons of different species and structural analysis between human SAP and CRP proteins pentamer structures. The experimental analysis demonstrates that SAP and CRP pentamers have a highly diverse structural topology with electrostatic and hydrophobic surfaces. Our study attempts to understand the interaction properties pattern of the pentameric model of zebrafish CA VI+PTX complex and CRP electrostatic surfaces. Significant hydrophilic and hydrophobic differences are observed among the CA VI+PTX complex and CRP. Our counted conserved residues are also the same in both pentraxins. The Ca2+ binding sites are relatively conserved, and Ca2+ ions coordinating residues are the same in both pentraxins. Finally, comparisons were compared to the contact between the SAP and CRP. The amount of contact residues is similar, but only the mode of binding is different. Based on the differences and low conservation of contacting residues, we can conclude that the mode of Ca6 associated PTX can not be fully predicted based on the study.

7 References

Agrawal A, Mitra S, Ghosh N, et al. C-reactive protein in hemolymph of a mollusc, Achatina fulica Bowdich. Indian J Exp Biol. 1990; 28:788–789.

Agrawal A, Simpson MJ, Black S, Carey MP, Samols D. A C-reactive protein mutant that does not bind to phosphocholine and pneumococcal C-polysaccharide. Journal of Immunology. 2002; 169:3217–3222.

Aldred, P., Fu, P., Barrett, G., Penschow, J. D., Wright, R. D., Coghlan, J. P. And Fernley, R. T. (1991) Biochemistry *30*, 569-575

Annette K. Shrive, graham M.T. Cheetham, David Holden, Dean A.A. Myles, William G, Turnell, John E. Volanakis, Mark B. Pepys, Anne C. Bloomer and Trevor J. Greenhough. Three-dimensional structure of human C-reactive protein. Nature (1996).

Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E. The Protein Data Bank. Nucleic Acids Res. 2000; 28:235–242

Björkholm P, Daniluk P, Kryshtafovych A, Fidelis K, Andersson R, Hvidsten TR. Using multi-data hidden Markov models trained on local neighborhoods of protein structure to predict residue–residue contacts. Bioinformatics. 2009;25(10):1264–1270.

Black S, Agrawal A, Samols D. The phosphocholine and the polycation-binding sites on rabbit C-reactive protein are structurally and functionally distinct. Molecular Immunology. 2003; 39:1045–1054.

Campbell SJ, Gold ND, Jackson RM, Westhead DR. Ligand binding: functional site location, similarity and docking. Curr Opin Struct Biol. 2003; 13:389–395.

Chang, Raymond. Physical Chemistry for the Biosciences. Sausalito, CA: Edwards Brothers, Inc. 2005. 508-510.

characteristics of protein Ca2+-binding sites. J Biol Inorg Chem. 2008; 13:1169–81.

Di Lena P, Nagata K, Baldi P. Deep architectures for protein contact map prediction. Bioinformatics. 2012;28(19):2449–2457.

Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, et al. Principles of protein folding, a perspective from simple exact models. Protein Science. 1995; 4:561–602.

Domsic JF, Avvaru BS, Kim CU, Gruner SM, Agbandje-McKenna M, Silverman DN et al. Entrapment of carbon dioxide in the active site of carbonic anhydrase II. J Biol Chem 2008; 283:30766–30771.

Doncheva NT, Klein K, Domingues FS, Albrecht M. Analyzing and visualizing residue networks of protein structures. Trends Biochem Sci. 2011; 36:179–82.

Duarte JM, Sathyapriya R, Stehr H, Filippis I, Lappe M. Optimal contact definition for reconstruction of contact maps. BMC Bioinformatics. 2010; 11(1):283.

Duarte JM, Sathyapriya R, Stehr H, Filippis I, Lappe M. Optimal contact definition for reconstruction of contact maps. BMC Bioinformatics. 2010; 11(1):283. [PubMed: 20507547]

Durbin, R. M., Eddy, S. R., Krogh, A. & Mitchison, G. (1998). *Biological sequence analysis,* Cambridge University Press, Cambridge, UK.

Eickholt J, Cheng J. A study and benchmark of DNcon: a method for protein residue-residue contact prediction using deep networks. BMC Bioinformatics. 2013; 14(Suppl 14): S12.

Emsley J et al. Structure of pentameric human serum amyloid P component. Nature 367, 338-345 (1994)

Fernley, R. T., Wright, R. D. and Coghlan, J. P. (1988) Biochemistry *27*, 2815-2820

Garlanda C, Bottazzi B, Bastone A, et al. Pentraxins at the crossroads between innate immunity, inflammation, matrix deposition, and female fertility. Annu Rev Immunol. 2005; 23:337–366.

Gavin E. Crooks, Gray Hon, John-Marc Chandonia, and Steven E. Brenner. WebLogo: A Sequence Logo Generator. 2004

Garrett, Reginald H. and Charles M. Grisham. Biochemistry. Belmont, CA: Thomas Brooks/Cole. 2005. 15.

Giovambattista N, Lopez CF, Rossky PJ, Debenedetti PG. Hydrophobicity of protein surfaces: Separating geometry from chemistry. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105:2274–2279.

H. Jane Dyson, Peter E. Wright, and Harold A. Scheraga. The role of hydrophobic interactions in initiation and propagation of protein folding**.** PNAS August 29, 2006 103 (35) 13057-1306**1.**

Hewett-Emmett, D. and Tashian, R. E. (1996) Mol. Phylogenet. Evol. *5*, 50-77

Hirschfield GM, Pepys MB. C-reactive protein and cardiovascular disease: New insights from an old molecule. QJM. 2003; 96:793–807.

Holm,L. and Sander,C. (1996) Mapping the Protein Universe. Science, 273, 595–603.

Hurlimann J, Thorbecke GJ, Hochwald GM. The liver as the site of C-reactive protein formation. J Exp Med. 1966; 123:365–378.

Jeffrey, G. A.; *An introduction to hydrogen bonding*; Oxford university press New York, 1997.

Jiang W, Gupta D: Structure of the carbonic anhdrase VI (CA6) gene evidence for two distinct groups within the αCA gene family. Biochem J *344*: 385-390, 1999.

Jiang, W., Woitach, J. T. and Gupta, D. (1996) Biochem. J. *318*, 291-296

Jones DT, Buchan DW, Cozzetto D, Pontil M. PSICOV: precise structural contact prediction using sparse inverse covariance estimation on large multiple sequence alignments. Bioinformatics. 2012; 28(2):184–190. [PubMed: 22101153]

Ken Nishikawa TO, Isogai Yoshinori, Saitô Nobuhiko. Tertiary Structure of Proteins. I. Representation and Computation of the Conformations. J Phys Soc Jpn. 1972; 32:1331–1337.

Kimoto M, Kishino M, Yura Y, Ogawa Y: A role of salivary carbonic anhydrase VI in dental plaque. Arch Oral Biol *51*: 117-122, 2006.

Kirberger M, Wang X, Deng H, Yang W, Chen G, Yang JJ. Statistical analysis of structural

Kivela J, Parkkila S, Parkkila AK, Leinonen J, Rajaniemi H: Salivary carbonic anhydrase isoenzyme VI. J Physiol *520*: 315-320, 1999.

Kolodny R, Koehl P, Levitt M. Comprehensive evaluation of protein structure alignment methods: scoring by geometric measures. J Mol Biol. 2005; 346:1173–1188.

Kornev AP, Haste NM, Taylor SS, Eyck LF. Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. Proc Natl Acad Sci U S A. 2006; 103:17783–8.

Lesk, A. M. Introduction to Protein Architecture. OUP, Oxford, 2001.

Ma, C. D., Wang, C., Acevedo-Vélez, C., Gellman, S. H. & Abbott, N. L. *Nature* **517**, 347–350 (2015).

Martin AJ, et al. RING: networking interacting residues, evolutionary information and energetics in protein structures. Bioinformatics. 2011; 27:2003–5.

Melis Kayikci, A. J. Venkatakrishnan, James Scott-Brown, Charles N. J. Ravarani, Tilman Flock, and M. Madan Babu. Protein contacts atlas: visualization and analysis of non-covalent contacts in biomolecules. Nat Struct Mol Biol. 2018 February; 25(2): 185–194.

Messerchmidt A, Bode W, Cygler M. (2004). Handbook of Metalloproteins. West Sussex, England: John Wiley and Sons Ltd.

Mold, Carolyn; Sun, Peter D; and Du Clos, Terry W (November 2012) Pentraxins. In: eLS. John Wiley & Sons, Ltd: Chichester.

Monastyrskyy B, D'Andrea D, Fidelis K, Tramontano A, Kryshtafovych A. Evaluation of residue–residue contact prediction in CASP10. Protein Struct Funct Bioinform. 2014; 82(S2):138–153.

Monastyrskyy B, Fidelis K, Tramontano A, Kryshtafovych A. Evaluation of residue–residue contact predictions in CASP9. Protein Struct Funct Bioinform. 2011; 79(S10):119–125.

Munson M, Balasubramanian S, Fleming KG, Nagi AD, O'Brien R, et al. What makes a protein a protein? Hydrophobic core designs that specify stability and structural properties. Protein Science. 1996; 5:1584–1593.

Némethy G., Steinberg I. Z., Scheraga H. A. (1963) Biopolymers **1**:43–69.

Orengo CA, Todd AE, Thornton JM. From protein structure to function. Curr Opin Struct Biol. 1999; 9:374–382. doi: 10.1016/S0959-440X (99)80051-7

Osmand AP, Friedenson B, Gewurz H, et al. Characterization of C-reactive protein and the complement subcomponent C1t as homologous proteins displaying cyclic pentameric symmetry (pentraxins). Proc Natl Acad Sci USA. 1977; 74:739–743.

Parkkila, S. and Parkkila, A.-K. (1996) Scand. J. Gastroenterol. *31*, 305-317

Pedro M. Martins, Vinícius D. Mayrink, Sabrina de A. Silveira, Carlos H. da Silveira, Leonardo H. F, de Lima, Raquel C. de Melo-Minardi. Proceedings of the 33rd Annual ACM Symposium on Applied ComputingApril 2018 Pages 60–67.

Pepys MB, Dash AC, Fletcher TC, et al. Analogues in other mammals and in fish of human plasma proteins, C-reactive protein and amyloid P component. Nature. 1978; 273:168–170.

Pidcock E, Moore GR. Structural characteristics of protein binding sites for calcium and lanthanide ions. J Biol Inorg Chem. 2001; 6:479–89. [PubMed: 11472012]

Russell RB, Alber F, Aloy P, Davis FP, Korkin D, Pichaud M, Topf M, Sali A. A structural perspective on protein–protein interactions. Curr Opin Struct Biol. 2004; 14:313–324.

Ryan DP, Matthews JM. Protein-protein interactions in human disease. Curr Opin Struct Biol. 2005; 15:441–6.

Sayers E. (2013). The NCBI handbook, 2nd edition, NCBI Protein Resources

Schneider M, Brock O. Combining physicochemical and evolutionary information for protein contact prediction. PLoS One. 2014;9(10): e108438. doi: 10.1371/journal.pone.0108438.

Scozzafava A, Mastrolorenzo A, Supuran CT. Carbonic anhydrase inhibitors and activators and their use in therapy. Expert Opinion on Therapeutic Patents 2006; 16:1627–1664

Shrive AK, Cheetham GM, Holden D, Myles DA, Turnell WG, Volanakis JE, Pepys MB, Bloomer AC, Greenhough TJ Nat Struct Biol. 1996 Apr; 3(4):346-54.

Sly, W. S. and Hu, P. Y. (1995) Annu. Rev. Biochem. *64*, 375-401

Suel GM, Lockless SW, Wall MA, Ranganathan R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. Nat Struct Biol. 2003; 10:59–69.

Thatcher, B. J., Doherty, A. E., Orvisky, E., Martin, B. M. and Henkin, R. I. (1998) Biochem. Biophys. Res. Commun. *250*, 635-641

Thompson D (1999), The physiological structure of human C-reactive protein and its complex with phosphocholine *structure*

Thompson, S.G., Kienast, J., Pyke, S.D.M., Haverkate, F. & van de Loo, J.C.W. (1995). Hemostatic factors and the risk of myocardial infarction or sudden death in patients with angina pectoris. *N. Engl. J. Med.* **332**, 635-641.

Tillett WS and Francis T Jr. (1930) Serological reactions in pneumonia with a non-protein fraction of pneumococcus. Journal of Experimental Medicine 52: 561–571.

Tillett WS, Francis T Jr. Serological reactions in pneumonia with a nonprotein somatic fraction of pneumococcus. J Exp Med. 1930; 52:561–571.

Tina KG, Bhadra R, Srinivasan N. PIC: protein interactions calculator. Nucl Acids Res. 2007;35: W473–6.

U Hobohm *et al*. *Protein Sci* **1:** 409 (1992).

Xie J., Schultz P. G. An expanding genetic code. Methods. 2005; 36:226–238.

Zacho J, Tybjaerg-Hansen A and Nordestgaard BG (2010) Creactive protein and all-cause mortality – the Copenhagen City Heart Study. European Heart Journal 31(13): 1624–1632.

8 Appendices

8.1 Appendix 1 - Multiple sequence alignment

```
CLUSTAL O(1.2.4) multiple sequence alignment


tr|E9QB97|E9QB97_DANRE      MEQLTLVLLFVTSLNFASAGVDGDYWTYSGELDQKHWAEKYHDCGGQQQSPIDIQRRKVR
pdb|4AVS|A                  ------------------------------------------------------------
pdb|3PVN|A                  ------------------------------------------------------------


tr|E9QB97|E9QB97_DANRE      YSPRMQQLELTGYEDIRGSFLMKNNGHSVEIQLPSTMKITKGFPHQYTAVQMHLHWGGWD
pdb|4AVS|A                  ------------------------------------------------------------
pdb|3PVN|A                  ------------------------------------------------------------


tr|E9QB97|E9QB97_DANRE      LEASGSEHTMDGIRYMAELHVVHYNSEKYPSFEEAKNKPDGLAVLAFFFEDGHFENTYYS
pdb|4AVS|A                  ------------------------------------------------------------
pdb|3PVN|A                  ------------------------------------------------------------


tr|E9QB97|E9QB97_DANRE      DFISNLANIKYVGQSMSISNLNVLSMLSENLSHFYRYKGSLTTPPCFESVMWTVFDTPIT
pdb|4AVS|A                  ------------------------------------------------------------
pdb|3PVN|A                  ------------------------------------------------------------


tr|E9QB97|E9QB97_DANRE      LSHNQIRKLESTLMDHDNKTLWNDYRMAQPLNERVVESTFLPRLSKGGMCRQEEIEAKLK
pdb|4AVS|A                  ------------------------------------------------------------
pdb|3PVN|A                  ------------------------------------------------------------


tr|E9QB97|E9QB97_DANRE      RIESLILSLDKKAVQGKQPISPLVLYFPQKNVESFAVVNLTHPMELKSFTACMNVQ--IP
pdb|4AVS|A                  ----------------HTDLSGKVFVFPRESVTDHVNLITPLEKPLQNFTLCFRAYSDLS
pdb|3PVN|A                  ----------------QTDMSRKAFVFPKESDTSYVSLKAPLTKPLKAFTVCLHFYTELS
                                             :  :*  .: **::.  ... :       *: ** *:.    :

tr|E9QB97|E9QB97_DANRE      PIRDLTVLSYSTSH-DNELMISLGSE--VGLWIGDEFVNLSFDLPSSDWTNYCLTWASHN
pdb|4AVS|A                  --RAYSLFSYNTQGRDNELLVYKERVGEYSLYIGRHKVTSKVIEKFPAPVHICVSWESSS
pdb|3PVN|A                  STRGYSIFSYATKRQDNEILIFWSKDIGYSFTVGGSEILFEVPEVTVAPVHICTSWESAS
                              *  :::** *.  ***:::        .: :*   :  ..       .: * :* * .

tr|E9QB97|E9QB97_DANRE      GGAELWVNGVVGKERYIRTGYIIPAGGRLILGKDQDGFLGI-SVNDAFVGHMSDVNIWDY
pdb|4AVS|A                  GIAEFWINGTPLVKKGLRQGYFVEAQPKIVLGQEQDSYGGKFDRSQSFVGEIGDLYMWDS
pdb|3PVN|A                  GIVEFWVDGKPRVRKSLKKGYTVGAEASIILGQEQDSFGGNFEGSQSLVGDIGNVNMWDF
                            * .*:*::*    .: :: ** : *   ::**::**.: *  . .:::**.:.:: :**

tr|E9QB97|E9QB97_DANRE      VLTEGEIVEQMSCDNGKVKGNVLSWGVTQLSLYGGVQLQGEQVCHRDNNNNRETEK
pdb|4AVS|A                  VLPPENILSAY--QGTPLPANILDWQALNYEIRGYVIIKPLVWV------------
pdb|3PVN|A                  VLSPDEINTIY--LGGPFSPNVLNWRALKYEVQGEVFTKPQLWP------------
                            **   :*       .  .  *:*.* . : .: * *  :
```