Full Paper

www.molinf.com

molecular informatics

DOI: 10.1002/minf.201800145

Computational Analysis of Non-covalent Interactions in Phycocyanin Subunit Interfaces

Luka M. Breberina,[a] Mario V. Zlatović,[b] Milan R. Nikolić,[a] and Srđan Đ. Stojanović*[c]

Abstract: Protein-protein interactions are an important phenomenon in biological processes and functions. We used the manually curated non-redundant dataset of 118 phycocyanin interfaces to gain additional insight into this phenomenon using a robust inter-atomic non-covalent interaction analyzing tool PPCheck. Our observations indicate that there is a relatively high composition of hydrophobic residues at the interfaces. Most of the interface residues are clustered at the middle of the range which we call "standard-size" interfaces. Furthermore, the multiple interaction patterns founded in the present study indicate that more than half of the residues involved in these interactions participate in multiple and water-bridged hydrogen bonds. Thus, hydrogen bonds contribute maximally towards the stability of protein-protein complexes. The analysis shows that hydrogen bond energies contribute to about 88% to the total energy and it also increases with interface size. Van der Waals (vdW) energy contributes to 9.3%±1.7% on average in these complexes. Moreover, there is about 1.9%±1.5% contribution by electrostatic energy. Nevertheless, the role by vdW and electrostatic energy could not be ignored in interface binding. Results show that the total binding energy is more for large phycocyanin interfaces. The normalized energy per residue was less than $-16\ kJ\,mol^{-1}$, while most of them have energy in the range from -6 to $-14\ kJ\,mol^{-1}$. The non-covalent interacting residues in these proteins were found to be highly conserved. Obtained results might contribute to the understanding of structural stability of this class of evolutionary essential proteins with increased practical application and future designs of novel protein-bioactive compound interactions.

Keywords: Phycocyanins · Interface · Hydrogen bonds · Hydrophobic interactions · Salt bridges

1 Introduction

Protein-protein interactions are central to the understanding of molecular mechanisms of biochemical processes and biochemical pathways, varying from enzymatic involvement to signal transduction. Protein-protein interactions are defined as specific physical contacts between protein pairs that occur by selective molecular docking in a particular biological context.[1–3] The study and comparison of protein-protein interfaces are essential for the understanding of the mechanisms of interaction between proteins. Such analysis is expected to have an impact on the prediction of interaction partners, as well as to assist in the design and engineering of protein interactions and interaction inhibitors.[4] Structural aspects, physicochemical properties, affinity and specificity of binding are diverse across different protein-protein interfaces.[5,6] Interactions between proteins have been classified according to different criteria; in a review, Nooren and Thornton use the criteria composition, affinity and lifetime to classify interactions as homo or hetero, obligate or non-obligate and permanent or transient, respectively.[5] Methods have been developed for distinguishing different interaction types based on interface properties.[6–9] Physicochemical properties of protein-protein interfaces include structural and chemical properties.[10–13] These should be examined to understand the nature of the intermolecular interactions. For example, the surface area that is buried by the interacting molecules and the nonpolar fraction, the hydrogen bonds and the salt bridges across the interface, buried water molecules, the charge distribution and the composition of the interface, residue conservation, the strength of the interaction, flexibility of the interface residues and residues that contribute significantly to the free energy of binding (hot spots), the shape of the binding interface, complementarity of two binding sites and the types of secondary structures are some of the properties of binding sites.[10,14] Detailed comparison of protein-protein interfaces is fundamental for their better characterization and for structure-based classification of protein complexes. The explicit comparison of non-covalent interactions provides an intuitive method of comparative analysis and visualization of binding modes and for

[a] *L. M. Breberina, M. R. Nikolić*
University of Belgrade – Faculty of Chemistry, Department of Biochemistry, Belgrade, Serbia
[b] *M. V. Zlatović*
University of Belgrade – Faculty of Chemistry, Center for Computational Chemistry and Bioinformatics, Belgrade, Serbia
[c] *S. Đ. Stojanović*
Institute of Chemistry, Technology and Metallurgy (ICTM) – Department of Chemistry, University of Belgrade, Belgrade, Serbia
E-mail: srdjanst@chem.bg.ac.rs

investigating the degree of conservation between interfaces.[15]

Phycobiliproteins (PBPs) are a family of water-soluble intensely fluorescent holoproteins consisting of apoprotein and covalently bound linear tetrapyrrole chromophores called phycobilins that function as components in the photosynthetic apparatus of cyanobacteria and certain algae.[16] These organisms have been major contributors to the evolution of oxygen and the absorption of carbon dioxide from the atmosphere.[17] Most common PBPs, differing in their protein structure, phycobilin content attached to conserved cysteine residues, absorbance, and fluorescent properties, are phycoerythrins, with phycoerythrobilin as red chromophore and phycocyanins (C-phycocyanin and allophycocyanin) with blue-purple phycocyanobilin chromophore. All PBP self-assembly is initiated by the association of A and B subunits, only fairly homologous on the amino acid sequence level (25–40%) but highly homologous on the three-dimensional level of structure.[18] Their molecular weights differ depending on the organism of origin, ranging 12–20 kDa for the A subunit, and 15–22 kDa for the B subunit. Each A and B subunit contains eight α helices, six of which are folded into a globin-like structure, with (one or two) phycocyanobilins bound in an analogous position to that of porphyrin in hemoglobin.[19] AB heterodimer is traditionally defined as the phycobiliprotein "monomer", reflecting the extreme stability of this complex *in vitro* relative to higher self-association forms. The monomers assemble into trimers (mostly *via* hydrophobic interactions between A and B subunits from different monomers), two of which further assemble into hexamers (mostly by polar interactions between A subunits). To form rods and tube-like structures as a constitutive part of phycobilisomes, hexamers merge together mostly by the face-to-face interaction between B subunits. The open-chain tetrapyrrole structure makes bilin chromophores flexible in responding to changes in their nearest neighbor environment during folding and aggregation processes. All the trimeric and hexameric phycocyanins have similar structures and disk-like shape (Figure 1), with a central channel having a 3.5 to 4.5 nm diameter.[20] Although all listed PBPs have increased practical use in various fields (e.g. medicine, food industry, biotechnology, as a research tool), this particularly applies to C-phycocyanin purified from cyanobacteria *Spirulina*.[21] However, the application of phycocyanins has been hindered by its sensitivity to treatment and storage conditions. In general, the stability of phycocyanin aggregates depends on its origin, amino acid composition, light, pH, temperature, and some exogenous substances.[22] Interestingly, molecular forces (non-covalent interactions) responsible for the observed differences in thermal and chemical stability of different phycocyanin complexes are not completely understood.[18] Understanding the nature of non-covalent interactions is thus extremely important to see what causes these variations in the properties. Therefore, we have studied the role of non-covalent interactions in interfaces of phycocyanin proteins and their environmental preferences. We performed computational analysis of the X-ray structures of proteins containing phycocyanin A or B subunit domain and summarize non-covalent interactions, especially hydrogen bonds, salt bridges and hydrophobic interactions in order to better understand the high stability of phycocyanin oligomers. Also, the relative preference of amino acids participating in interfaces, interface area correlations, energetic contribution and conservation score of amino acid residues were analyzed. Obtained results might contribute to the understanding of structural stability of this class of evolutionary essential proteins with increased practical application and future designs of novel protein-bioactive compound interactions.

Figure 1. Schematic view of the structure of typical phycocyanin hexamer (left) and allophycocyanin trimer (right). The apoprotein and chromophores are shown by ribbon and ball-stick model, respectively. The A and B subunit of each monomer is represented by different shades of the same color. Reproduced with permission[23]; Copyright [2016], [Baishideng Publishing Group].

2 Materials and Methods

2.1 Dataset

To obtain meaningful statistical properties of non-covalent interactions across protein interfaces, one needs high quality, non-redundant experimental dataset. For this study, we used the Protein Data Bank (PDB) October 2nd, 2016 list of 122,832 structures.[24] We then created a non-redundant dataset of 20 proteins such that they satisfy the following conditions. These include: (1) structures of proteins containing phycocyanin alpha or beta subunit domain (SCOP Classification, version 1.75)[25] were accepted; (2) no theoretical model structures and no NMR structures were accepted, these structures were not included since it was difficult to define the accuracy of the ensemble of structures in terms of displacement that was directly comparable to the X-ray diffraction studies; and (3) only crystal structures with the resolution of 3.0 Å or better and a crystallographic R-factor of 25.0% or lower were accepted. In order to have a non-redundant set of native interfaces and avoid ambiguities,

we excluded structures containing ligands and mutant amino acids, thus leaving 20 proteins and 118 interfaces that were actually used as the dataset in our analysis.

Interface area was calculated using the PDBePISA (Proteins, Interfaces, Structures, and Assemblies) service at Protein Data Bank in Europe (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html).[26] Interface area in Å^2, calculated as the difference in total accessible surface areas (ASA) of isolated and interfacing structures divided by two. The program DSSP[27] was used to obtain information about ASA. ASA is typically calculated using the "rolling ball" algorithm developed by Shrake & Rupley in 1973.[28] This algorithm uses a sphere (of solvent) of a particular radius to "probe" the surface of the molecule.

We compared our results for interfaces of phycocyanin proteins with those of a test set formed from the Binding Interface Database (BID).[29] The redundancy in this dataset is removed using PISCES sequence culling server[30] with sequence identity not more than 35% as in the procedure of Darnell et al.[31]

2.2 Identification and Energy Calculation of Non-covalent Interactions

The protein-protein interaction check (PPCheck; http://caps.ncbs.res.in/ppcheck/) server[32] was applied for the identification of interaction strength, interface residues and the normalized energy per residue values for protein-protein complexes. As part of the preparations of the files, the input pdb file is "cleaned", i.e. atoms with multiple occupancies were selected (select the atoms with a maximum value of multiple occupancies), heteroatoms and also other REMARK columns were removed and information about only the amino acid coordinates and those of water atoms (where relevant) was retained. The PPCheck server can be used to identify interactions across two chains, once they are fed as inputs along with the PDB file.

Simple distance criteria are employed for the preliminary identification of non-covalent interactions, such as van der Waals, electrostatic and hydrogen bonding. The respective energies are calculated using standard force fields as described in the following. Relative orientations of interacting entities were not further adjusted in order to change correlations calculated.

All interprotomer interactions between hydrophobic amino acids (Ala, Leu, Ile, Val, Trp, Tyr, Phe) with C^{β}–C^{β} distance equal to or less than 7 Å have been considered to be interacting and contribute to van der Waals interaction energy.

Van der Waals interaction energy is calculated as

$$E_{vdW} = 4.184\times \sqrt{E_i \times E_j}\left[\left(\frac{R_i+R_j}{r}\right)^{12} - 2\left(\frac{R_i+R_j}{r}\right)^{6}\right] kJ\ mol^{-1} \quad (1)$$

E_{vdW} is the van der Waals energy; R_i and R_j are the van der Waals radii for the atoms i and j, respectively; E_1 and E_2 is the van der Waals well depth for the atoms i and j, respectively; and r is the distance between the atoms. The van der Waals interaction also accounts for hydrophobic interactions.[33]

All interprotomer charged amino acid pairs with atomic distance equal to or less than 10 Å are considered and reported. A salt bridge is said to be formed if the side-chain nitrogen and oxygen atoms of two oppositely charged residues are observed closer than 4 Å distance. Electrostatic interactions have been reported, and the corresponding energies are calculated using CHARMM package,[34] if the charged residues are within or equal to an optimum distance cutoff of 10 Å. Coulomb's equation was used to quantify these interactions as follows:

$$E_{el} = 1.184 \times \left(\frac{q_1 \times q_2}{D \times r}\right) \times 332\ kJ\ mol^{-1} \quad (2)$$

E_{el} is the electrostatic energy; q_1 and q_2 are the partial atomic charges; r is the distance between the atoms and D is the dielectric constant of the medium if calculations are performed using dielectric continuum. For this purpose, distance-dependent dielectric (DDD), where $D=2r$, is used for the electrostatic energy. The values for the different parameters are as in CHARMM.[34,35]

The positions of hydrogens with respect to the connecting atoms have been determined/fixed geometrically using standard bond lengths, angles and torsion angles for all types of atoms; methyl, methylene, tertiary groups; considering sp^3, sp^2, and sp atomic states of hybridization as described in the paper of Nardelli.[36] Hydrogen bonds are identified, and the corresponding energy is calculated using Kabsch and Sander's equation as used in the DSSP program as follows:

$$E_{Hb} = 4.184 \times q_1 \times q_2\left(\frac{1}{r_{ON}} + \frac{1}{r_{CH}} - \frac{1}{r_{OH}} - \frac{1}{r_{CN}}\right) \times f\ kJ\ mol^{-1} \quad (3)$$

where $q_1=0.42e$, $q_2=0.20e$, and $f=332$, and partial charges on the C, O ($+q_1$, $-q_1$) and N, H ($-q_2$, q_2) atoms. r is the inter-atomic distance between the corresponding atoms. The values for different parameters are as in DSSP.[27]

Water molecules, when present at the interface, are considered when they form bridging hydrogen bonds with amino acids from two interacting protein chains. The single point charge (SPC) model of water is considered,[37] and

Table 1. The dataset of the proteins used for analyses of interfaces.

Classification	Genetic source	Number of subunits	Number of interfaces[a]	Resolution (Å)	PDB Code
Allophycocyanin					
Light-harvesting protein	*Arthrospira platensis*	2	1	2.3	1all
Photosynthesis	*Mastigocladus laminosus*	14	14	2.3	1b33
Electron transport	*Pyropia yezoensis*	2	1	2.2	1kn1
Photosynthesis	*Gloeobacter violaceus*	2	1	2.5	2vjt
Photosynthesis	*Thermosynechococcus vulcanus*	8	4	2.9	3dbj
Photosynthesis	*Synechococcus elongatus*	6	6	2.5	4f0u
Photosynthesis	*Synechocystis sp. PCC 6803*	6	6	1.75	4po5
Photosynthesis	*Phormidium rubidum*	2	1	2.51	4rmp
C-phycocyanin					
Light-harvesting protein	*Fremyella diplosiphon*	4	2	1.66	1cpc
Photosynthesis	*Arthrospira platensis*	24	30	2.2	1gh0
Photosynthesis	*Polysiphonia urceolata*	6	6	2.4	1f99
Photosynthesis	*Synechococcus elongatus*	2	1	1.45	1jbo
Electron transport	*Cyanidium caldarium*	2	1	1.65	1phn
Photosynthesis	*Gracilaria chilensis*	12	12	2.01	2bv8
Photosynthesis	*Gloeobacter violaceus*	12	12	2.4	2vml
Photosynthesis	*Thermosynechococcus vulcanus*	2	1	1.35	3o18
Photosynthesis	*Leptolyngby asp. N62DM*	12	12	2.61	4 l1e
Photosynthesis	*Hemiselmis virescens*	4	2	1.7	4lm6
Photosynthesis	*Chroomonas sp.CCMP 270*	4	3	1.35	4lms
Photosynthesis	*Phormidium rubidum sp. A09DM*	4	2	2.7	4yjj

[a] The selection criteria for interfaces to be included in the dataset were: (1) without ligands; (2) interface area $\geq$100 Å^2; (3) *x,y,z* symmetry operation; and (4) essential and auxiliary role in complex formation (CSS$\geq$0.1).

therefore, the values of the charges are chosen as follows for the water-amino acid interactions in the Kabsch and Sander's equation:[27]

$q_1 = 0.42e$, $q_2 = 0.41e$,

when water acts as hydrogen bond donor;

$q_1 = 0.82e$, $q_2 = 0.20e$,

when water acts as hydrogen bond acceptor.

In this study, the hydrogen bonding criteria for furcated geometries were set as d$\leq$3.0 Å and $\theta \geq 90°$. These furcated geometries (bifurcated, trifurcated, etc.) constitute independent sets in the sense that the trifurcated geometries do not implicitly include the bifurcated ones and so on.

All the non-covalent interaction energies are summed up to total energy, and the ratio of total energy to the number of interface residues is termed as *normalized energy per residue*.

2.3 Computation of Conservation of Amino Acid Residues

The conservation of amino acid residues in each protein was computed using the ConSurf server (http://consurf.tau.ac.il/2016/).[38] This server computes the conservation based on the comparison of the sequence of a PDB chain with the proteins deposited in Swiss-Prot[39] and finds the ones that are homologous to the PDB sequence. The number of PSI-BLAST iterations and the E-value cutoff used in all similarity searches were 1 and 0.001, respectively. All the sequences that were evolutionary related to each one of the proteins in the data set were used in the subsequent multiple alignments. Based on these protein sequence alignments, the residues are classified into nine categories from highly variable to highly conserved. Residues with a score of 1 are considered to be highly variable and residues with a score of 9 are considered to be highly conserved.

3 Results and Discussion

Various residue interactions like hydrophobic, salt bridges and hydrogen bonds in phycocyanin subunit interfaces were analyzed to understand and predict protein-protein interactions, in particular, interaction sites. Also the relative preference of amino acids participating in interfaces, interface area correlations, energetic contribution and conservation score of amino acid residues were analyzed. Here, we use a selected set of properties to study the interface non-covalent interactions of subunits in phycocyanin proteins (Table 1).

3.1 Interface Area Correlations

Protein interface area and the number of residues varied with different studies and are affected by dataset size and data type. The average size of protein complex interfaces is between 1200 and 2000 Å^2 with an average of 23 residues in each protomer.[40] Richard Bickerton's analysis of the PICCOLO database[41] found that the average size of the protein-protein interfaces is greater than previously reported. Bickerton found that the average of interface size is 2400 ± 1900 Å^2. Interfaces with interface area < 1200 Å^2, were considered as "small" interfaces, while interfaces with interface area > 2000 Å^2 were "large" interfaces.[11,42] Looking at the type of interface, obligate complexes interact on average through larger interfaces than the transient ones.

We estimated the size of interfaces in phycocyanins by measuring the area of the protein surface buried in subunit contacts. Figure 2 is a plot of the interface area observed in the 118 interfaces against the protein size, estimated by the interface residues.

The data indicate a strong linear correlation between the size of the interface and the number of interface residues ($R^2 = 0.9863$). For our dataset, the mean interface area was 1088 ± 489 Å^2 (mean ± SD, here and below) and there were on average 59 ± 23 residues per interface. The interface areas range from 450 Å^2 to above 2,100 Å^2. At the low end of the distribution, about 39% of small proteins have an interface burying about 500 Å^2 per subunit. However, the plot of Figure 2 shows that most values are clustered at the middle of this range (51%), which we call "standard-size" interfaces. Above the limits of standard-size interfaces, we found the four "large" interfaces with interface area about 2,000 Å^2. The data show that the number of interface residues is proportional to the interface area. The smaller proteins obviously cannot form very large interfaces.

3.2 Interface Residue Composition

Interfaces have been shown to be more hydrophobic than the surface of the protein, but are less hydrophobic than the interior of the protein.[42] It is shown that hydrophobic residues dominate large interfaces whereas charged residues dominate small interfaces[43] in the case of dimers. In our study conducted on 118 phycocyanin interfaces, 51% of interface residues were hydrophobic, 27% hydrophilic and 22% charged (Figure 3).

The higher occurrence of hydrophobic residues, when compared to the other residues, is common. The percentage distributions of various residues in the Binding Interface Database (BID) test set[29] and in the phycocyanin dataset are comparable; the amino acid compositions were similar for both sets of proteins. Namely, the interfaces in the BID test set had a high fraction of hydrophobic amino acids (46%), and smaller fraction of hydrophilic (31%) and charged (23%) residues. The difference is particularly large for some amino acids; phycocyanin dataset contains 10.0% of Ala, 7.4% of Tyr, 8.3% of Thr and 9.5% of Lys, while the BID set contained 4.5%, 4.5%, 5.7%, and 5.2%, respectively. Figure 3 describes the contribution of each of the 20 amino acid types to the phycocyanin interfaces and to the BID interfaces.

Only four amino acids appear in phycocyanin with a frequency of more than 8%; 11.4% of Ala, 10.0% of Leu, 8.3 of Thr and 9.5% of Arg residues are in interfaces. However, many amino acids are found in phycocyanin interfaces very rarely. Less than 2% of the Trp, Cys, His and Gln residues in our database are in phycocyanin interfaces. It is interesting to note that Arg is favored charged amino acid, appearing in interfaces. Arg has the ability to form a hydrogen bond network with up to five hydrogen bonds and a salt-bridge with its positively charged guanidinium group. One explan-

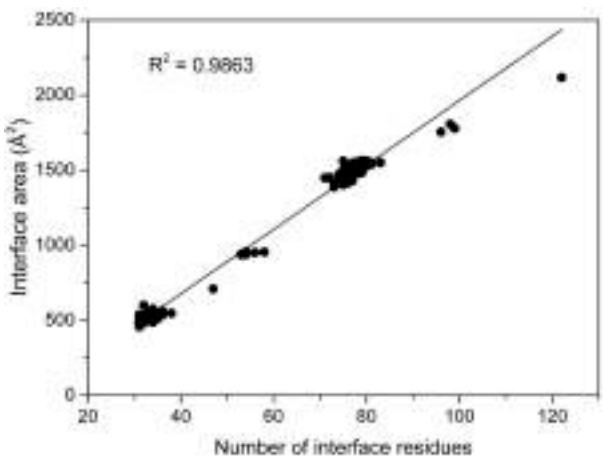

Figure 2. The relationship between interface area and the number of interface residues in phycocyanins.

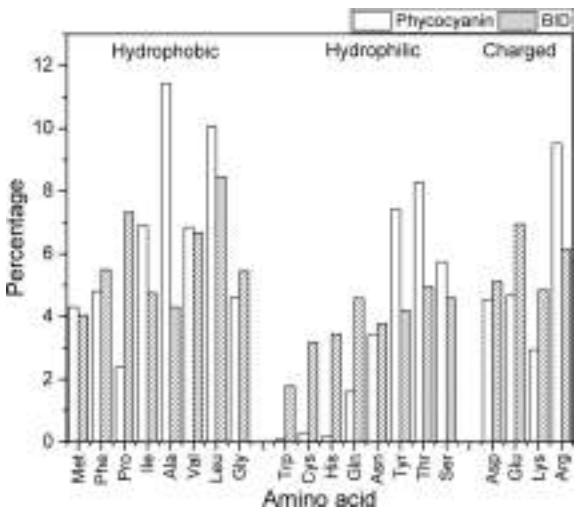

Figure 3. Composition of interface amino acids in phycocyanin and BID datasets.

ation is that amino acids capable of making multiple types of favorable interactions are preferred in the lowered effective dielectric environment of interfaces. The high abundance of Arg at interfaces has also been seen in other protein-protein complexes.[11] Further, Arg is common on all protein surfaces, not only protein-protein interfaces.[40] Among the hydrophilic residues, the most frequently observed in phycocyanin interfaces are Tyr, Thr, and Ser, which are also capable to form multiple types of favorable interactions.

3.3 Interface Hydrogen Bonds (H-bonds)

The networks of hydrogen bonds between subunits are important for protein stability, functionality, and structural integrity.[11,44–46] There are 903 hydrogen bonds across the 118 phycocyanin protein interfaces of our dataset. In the following, we analyzed their distribution, composition, and geometry. Among these interactions, 28% of interactions belong to the hydrophobic, 36% to the hydrophilic and 36% to the charged amino acid residues. On average, there are 0.13 H-bonds per interface residue (7.65 per interface) in our dataset. The maximum number of H-bonds per interface residue is 0.38 (34 per interface). The number of hydrogen bonds is mediocrely correlated with the number of interface residues, with a correlation coefficient of 0.44, as shown in Figure 4.

The dataset used in this study contains structures with resolution $\leq$ 3.0 Å and the data is a mixture of heterodimers, homodimers and other oligomers. In order to have a non-redundant set of interfaces, we divided the dataset into homo and hetero subdatasets. In that case, the correlation coefficient is 0.81 in homooligomers and 0.78 in hetero-oligomers. This is similar to the previous reports in the range of 0.75 and 0.89.[5,11,47] However, there is a subtle difference with the previous studies and the variation is affected by structure resolution, dataset size, and data type. At low resolution, there are fewer H-bonds and the correlation with interface area decreases.[11] This explains why the distribution is much wider (lower correlation) in whole dataset hydrogen bonds than in the subsets (homo- and heterooligomers). Here, we show that the relation between H-bonds and interface residues is highly correlated for both homooligomers and heterooligomers. The strong correlation illustrates a relatively narrow distribution of hydrogen bond density across the protein interfaces. This is useful to evaluate inter-subunit H-bonds prediction and their involvement in interface stability.

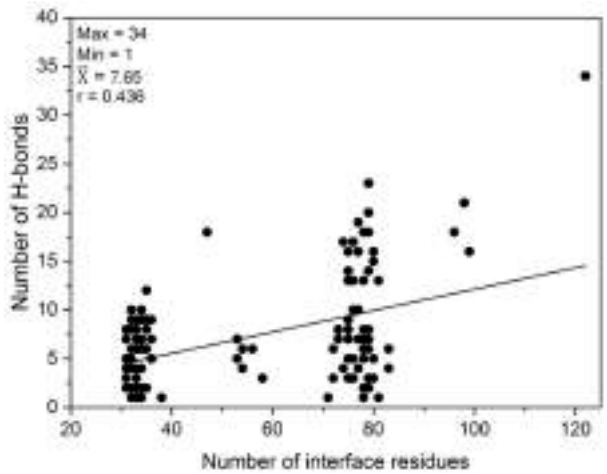

Figure 4. The relation between H-bonds and interface residues in phycocyanins.

We have examined the environment of the hydrogen bonds. We analyzed the types of residue side chains that are present in the vicinity of the donor-acceptor pairs in the hydrogen bonds. The results indicate a high number of hydrophobic residues (55.1%) in the neighborhood of both the donor and acceptors involved in H-bond formation. Both the charged (20.1%) and the polar environment (24.8%) around the hydrogen bonds are considerably less. These results indicate clearly that hydrogen bonds could not only exist between neutral donor-acceptor pairs but also could exist in the absence of a charged environment.

3.4 Multiple and Water-bridged H-bonds

Hydrogen bonds with multiple donors (acceptor furcation) and multiple acceptors (donor furcation) are known to be common in protein structures.[48,49] Strong interactions like O–H···O tend to go towards non-furcated geometries, in contrast to weak interactions, which occur more often in the furcated form due to better interaction geometry and more efficient space filling.[41] In this analysis, we have investigated the cases of furcated multiple hydrogen bonds. The analysis shows that about 42% of the total hydrogen bonds in the dataset are involved in the formation of multiple hydrogen bonds. This conveys that furcation is an inherent characteristic of macromolecular crystal structures. The acceptor furcation (61%) is more predominantly seen as compared to the donor furcation (39%). This could mainly be due to the fact that the geometrical constraints for acceptor furcation are lower, as the two donor atoms are separated in space without sacrificing the hydrogen bond geometry. On the other hand, the geometrical constraints are greater for donor furcation. The bifurcated geometry is suitable for H-bonds. Where it can occur, the bifurcated geometry is energetically favored over single H-bonds. In other words, the energetic benefits of forming two interactions outweigh the cost of nonlinearity[48]

The hydrogen bonding capacity of water makes it easy to interact with protein, ligand or neighboring water molecules. In general, water prefers to accept hydrogen bonds from O–H and N–H donors in macromolecular structures, thus increasing the enthalpy of the final complex

and all this stabilizes the ligand-protein, and protein-protein complex. Because internal water molecules are in mostly apolar environments, their hydrogen bonds are often strong and well defined.[50] Water molecules, with their ability to form multiple bridging hydrogen bonds, have been identified as a key structural factor in mediating these interactions.[51–53] The total number of hydrogen bonds formed by water (as donor or acceptor) in the 118 phycocyanin interfaces under consideration is 477 (52.8%). In other words, these are nearly 4 such hydrogen bonds for each interface, on average, indicating that bound interfacial waters tend to mediate pairs of polar groups which cannot form hydrogen bonds directly with each other. This suggests that the buried water molecules in the interfaces very often participate in hydrogen bonds with the amino acids. Therefore, they are structurally important. The observed occurrence of the donor-H_2O-donor pairs is more frequent than acceptor-H_2O-acceptor pairs. We found that the donor-H_2O-donor pairs constitute as much as 60% of the total number of water-bridged hydrogen bonds. Some water molecules have the potential to form hydrogen bonds with several protein donors/acceptors.

As an illustrative example for multiple and water-bridged hydrogen bonds is shown in Figure 5 (Allophycocyanin from *Mastigocladus laminosus*; PDB ID code 1b33). Examples of acceptor bifurcated hydrogen bonds are atoms OD1 (carboxylate group of Asp11(A)) and OD2 (carboxylate group of Asp3(B)) making hydrogen bonds with Arg90(B), Tyr91(B) and Ser1(A), Thr4(A) respectively. The bottom plot of Figure 5 shows the occurrence of water in the hydrogen bonding network across the K–L interface of Allophycocyanin from *Mastigocladus laminosus*. There are five water molecules that bridge hydrogen bonds between amino acid residues of both subunits. One of them is HOH217(K) bridging hydrogen bond between atom Thr66(K) and Tyr86 (L). The water molecule HOH2043(L) shows furcation bridging three amino acid residues (Tyr62(K) – Glu75(L); Thr79(L)).

3.5 Composition and Geometry of Hydrogen Bonds

The composition of the hydrogen bonds for the types oxygen-oxygen (O–O), nitrogen-nitrogen (N–N) and oxygen-nitrogen (O–N) is shown in Table 2. We compared their properties in phycocyanins with those of interface hydrogen bonds found in the BID test set. The hydrogen bonds across the interfaces are predominantly the O–N type. There are no hydrogen bonds between nitrogen atoms because few types of nitrogens in amino acids (only $N^{\delta 1}$ and $N^{\varepsilon 2}$ of His) can serve as hydrogen bond acceptors. However, mean values and percentage for O–N hydrogen bonds are larger in BID dataset.

The composition of hydrogen bonds formed with main chains and side chains is shown in Table 2.

Table 2. Composition of hydrogen bonds per interface.

Phycocyanin	O–O	N–N	O–N	M–M	M–S	S–S
Mean	0.98	0.00	2.64	0.64	0.96	2.02
Standard deviation	0.93	0.00	2.91	1.30	0.83	2.10
Percentage	26.8	0.0	73.2	17.6	26.5	55.9
BID						
Mean	0.14	0.00	4.41	1.19	1.79	1.58
Standard deviation	0.35	0.00	2.92	1.50	1.41	1.76
Percentage	2.7	0.0	97.3	19.4	44.5	36.1

Interface hydrogen bonds associated with main chains (M) and side chains (S).

The percentage occurrences of the main chain-main chain, main chain-side chain and side chain-side chain hydrogen bonds are 17.6, 26.5 and 55.9%, respectively, within phycocyanin interfaces, but 19.4, 44.5 and 36.1%, respectively, across BID interfaces. Although appreciably fewer hydrogen bonds are formed by main chain atoms across the protein-protein interfaces, there are a considerably more main chain-side chain hydrogen bonds in BID test set. On the other hand, there are substantially larger numbers of side chain-side chain hydrogen bonds (55.9%) between the phycocyanin interfaces. The degree of compactness of the oligomer is probably related to subtle variations of side chain interactions besides the hydrogen bond network. Charged groups are heavily involved in the side chain-side chain packing across the interfaces. From these observations, we consider the contribution of side chain-side chain hydrogen bonds to be significant and may play an important role in determining the structural stability of phycocyanins. The results were consistent with the observations in Sm/LSm oligomeric proteins.[55]

The distribution of hydrogen bonds distances (between the donor and the acceptor atoms) for hydrogen bonds in the interface area of phycocyanins and BID dataset is shown in Figure 6.

The geometry analysis of the phycocyanin hydrogen bonds has revealed that the distance distribution is bimodal with a prominent minimum ~3.4 Å (Figure 6a). Outside the minimum, there are two distinct maxima, corresponding to simple hydrogen bonds (2.6–3.3 Å) and water-bridged hydrogen bonds (3.5–4.2 Å), respectively. In comparison with simple hydrogen bonds, the distances of water-bridged hydrogen bonds display a wider spread and are more difficult to evaluate. This is because there are many weak interactions with long hydrogen bonds, a large proportion of which are water-bridged. The fluctuations are clearly a consequence of their greater flexibility. With water molecules, there are fewer restrictions than in the main-chain to side-chain interactions because the water molecules are free to orient in the most favorable positions without any geometrical constraints. The higher number of hydrogen bonds in the region of 3.5–4.2 Å can be explained

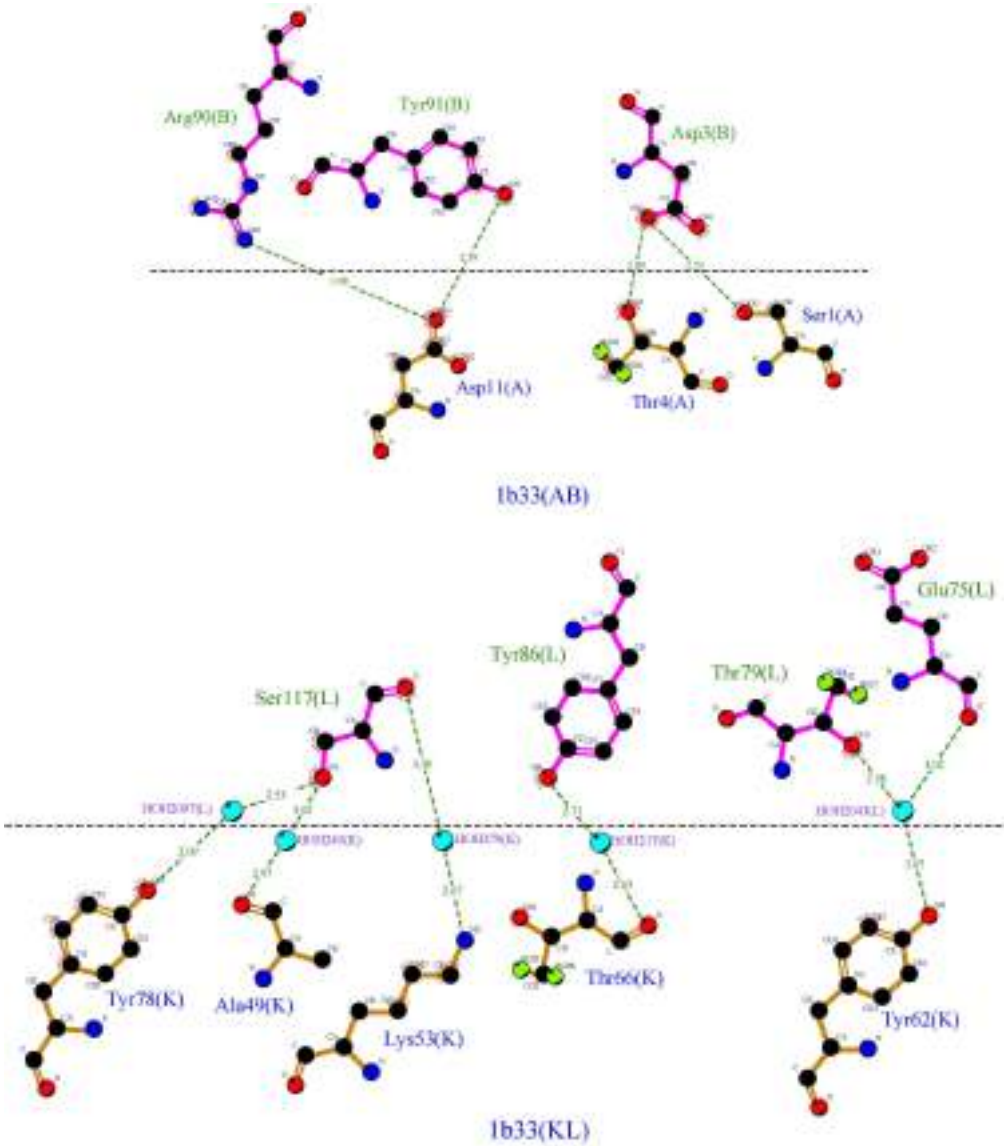

Figure 5. View of the interface multiple (AB interface) and water-bridged hydrogen bonds (KL interface) of Allophycocyanin from *Mastigocladus laminosus* (PDB ID code 1b33). The flattened diagram places atoms and bonds on the 2D page to minimize the overlap of atoms and the crossing of bonds in the final diagram. Hydrogen bonds are indicated by dashed green lines (with the distance between donor and acceptor printed in the middle) between the atoms involved. Because of flattening some of the hydrogen bond lines may be longer or shorter than their true lengths (data in the center of lines). Water molecules are shown with cyan balls. The letters in parentheses in the residue names are the corresponding chain identifiers. The Figure was prepared using the program LigPlot + v.2.1.[54]

by the fact that geometry of the water-bridged hydrogen bonds across protein interfaces is generally less optimal than within the chains. The distribution of BID interface hydrogen bonds is similar to that of the phycocyanin interface hydrogen bonds, with two maxima (Figure 6b).

The donor-acceptor distance is one of the measurements of hydrogen bond strength: they can be categorized as "strong, mostly covalent" with donor-acceptor distances of 2.2–2.5 Å, "moderate, mostly electrostatic" with 2.5–3.2 Å, and 3.2–4.0 Å as "weak, electrostatic".[56] Most of the hydrogen bonds in phycocyanin interfaces possess distances in the region 2.8–4.2 Å what indicate according to the classification[56] that phycocyanin interfacial hydrogen bonds are moderate and weak strength. Only 0.5 % of all hydrogen bonds have a donor-acceptor distance in the range 2.2–2.5 Å, which can be considered as strong hydrogen bonds.[56] The higher number of moderate and weak hydrogen bonds can be explained by the fact that the geometry of the hydrogen bonds across protein interfaces is generally less optimal than within the chains. The difference originates from the more hydrophilic side chains buried in the binding interface than in the folded monomer

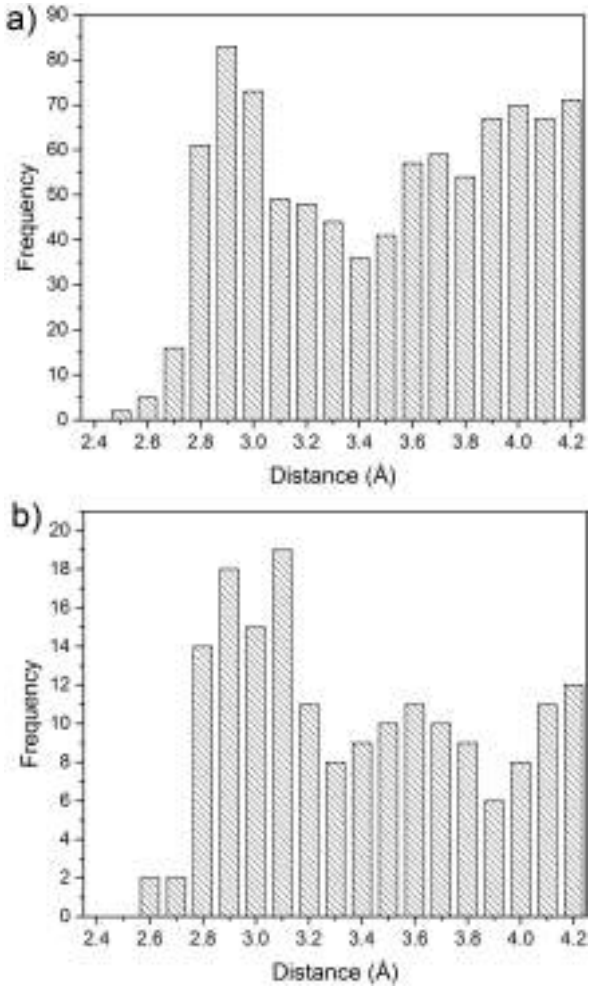

Figure 6. The distribution of the distance between the donor and the acceptor atoms of hydrogen bonds in the interface area (a) of phycocyanins and (b) BID dataset.

interior. Whereas in folding practically all degrees of freedom are available to the chain to obtain optimal configuration in the process of oligomerization this is not the case.[56] Due to the presence of a great number of non-covalent interactions in phycocyanin interfaces, the interplay between interactions may exist. It also should be taken into account in hydrogen bond strength.

3.6 Interface Hydrophobic Interactions

Several criteria reflecting the hydrophobic effect at the interface region have been explored in this study. These include: (1) the number of hydrophobic interactions and (2) the preference of amino acid residues to be involved in hydrophobic interactions.

Figure 7 shows that there is a high correlation between the number of hydrophobic interactions and the number of interface residues for the phycocyanin protein dataset. The

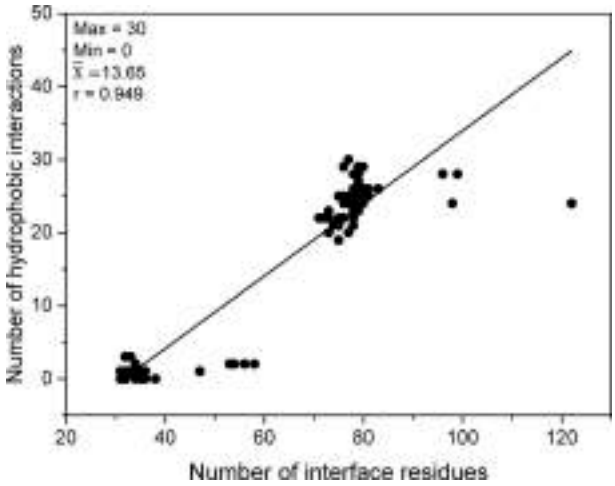

Figure 7. The relation between hydrophobic interactions and interface residues in phycocyanins.

mean number of hydrophobic interactions (per interface) is 13.65. The maximum number of hydrophobic interactions per interface is 30. The strong correlation illustrates a relatively narrow distribution of hydrophobic interactions across the protein interfaces.

The statistics of individual amino acid types were used to demonstrate the hydrophobic effect in protein-protein interfaces compared with that found in the interior of protein monomers. Figure 8 shows the amino acid composition in the interior of the monomeric proteins and in the interfaces for the protein dataset. Clearly, the amino acids in the interface can be considered as being initially on the surface of the protein, and following protein-protein association, they are buried in the interface.

The data in Figure 8 reveals that the charged and polar amino acids are more frequently found buried in the

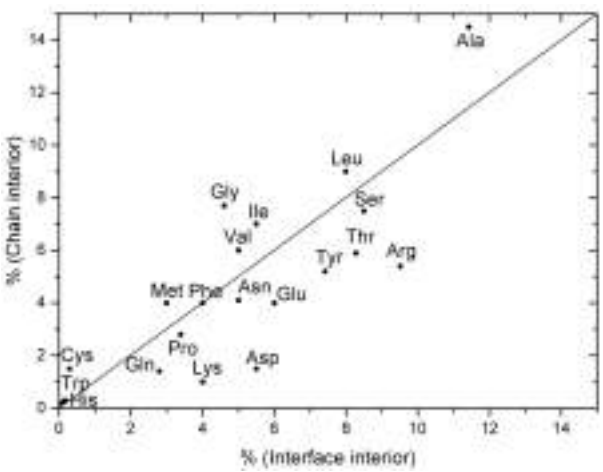

Figure 8. Amino acid composition in the interior of the monomers and the amino acid composition in the interfaces.

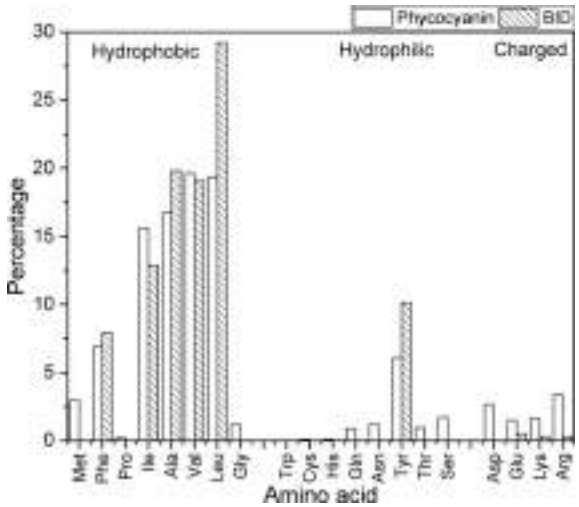

Figure 9. Composition of hydrophobic interaction forming amino acids in phycocyanin and BID datasets.

interior of the interfaces than in the monomers. Of these, the largest differences are observed in Arg, Lys, Asp, Glu, and Tyr, although Asn, Gln, Thr, and Ser are also more prevalent consistently in the interior of the interfaces than in that of the monomers. His, Phe and Trp are distributed equally in both. Pro is slightly more frequent in the interior than on the interface. As expected, the hydrophobic residues, Ala, Val, Leu, Ile, Gly, and Cys are found more frequently in the monomer cores. In particular, Figure 8 demonstrates the very clear difference between the interior of the monomers versus the interior of the interfaces. Clearly, this implies that, although protein-protein interfaces are stabilized by the hydrophobic effect, it is not to the same extent as observed in chains. A high percentage (19%) of ASA buried in the interface is observed in these interfaces. The hydrophobic effect in the stable interfaces is clearly much stronger than that of the overall data set.

We have computed the composition of hydrophobic interaction forming amino acids using protein sequences in phycocyanin and BID protein interfaces (Figure 9).

The results presented in Figure 9 show that the hydrophobic side chains make a larger number of the interactions than side chains of charged and the hydrophilic amino acid. This is not surprising; the charged and hydrophilic residues are typically involved in hydrogen bonds and salt bridges across the interfaces. A small percentage of hydrophilic and charged residues are observed in these interfaces because most of them still possess appreciable hydrophobic portions. A relevant example is Lys, which, although having a charged head, has a hydrophobic tail comprised of four carbon atoms, attaching it to the main chain. The backbone groups are not frequently involved (data not shown), because their atoms are not as accessible as the side-chain atoms. Figure 9 illustrates that the general trend between the phycocyanin and the BID hydrophobic interaction forming residues is similar. However, an inspection of the figure also reveals that, in an appreciable number of representative interfaces, hydrophobicity might play a significant role in protein-protein associations.

3.7 Salt Bridges

Ion pairs play an important role in stabilization of protein structures. Salt bridges are not being frequent among protein interfaces. Many of interfaces did not form salt bridges, while the largest number of salt bridges in an interface was five.[57–59]

The relation between salt bridges and interface residues is highly correlated ($r = 0.826$) for the phycocyanin protein dataset. We have identified 317 salt bridges across the 118 protein interfaces. On average, there are about 3 salt bridges per interface in phycocyanin proteins, while the maximum number of salt bridges per interface is 7. Design of higher-order interactions is an important focus in rational protein design. Interactions with one basic residue and multiple acidic residues are commonly called "complex" or "networked" salt bridges. Less than one-tenth of the salt bridges in our database are networked, to form several triads. The remaining are isolated salt bridges. Most salt bridges (~80%) contain at least one hydrogen bond between the atoms in their side-chain charged groups.

The normalized distribution of different types of salt bridges is shown in Table 3. There is no preferred combination of donors and acceptors. It is in accord with the finding that the count of each salt bridge type indicates a lack of discrimination between specific salt bridge donors and acceptors when they form salt links.[44]

Table 3. Salt bridge distribution.

Residue	Arg	Lys	N-terminal
Asp	1	0.588	0.012
Glu	0.713	0.726	0.005
C-terminal	0.061	0.075	0.004

An illustrative example for non-covalent interaction profile of allophycocyanin from *Arthrospira platensis* with a view of the interface between subunits (PDB ID code 1all) is shown in Figure 10. There are 16 hydrogen bonds (7 of them are water-bridged), 18 hydrophobic interactions and 4 salt bridges between A and B subunits.

3.8 Energetic Contribution

Usually, the stability of a non-covalent complex is related to the complexation energy, which is proportional to the strength of the interactions involved. Therefore, we have

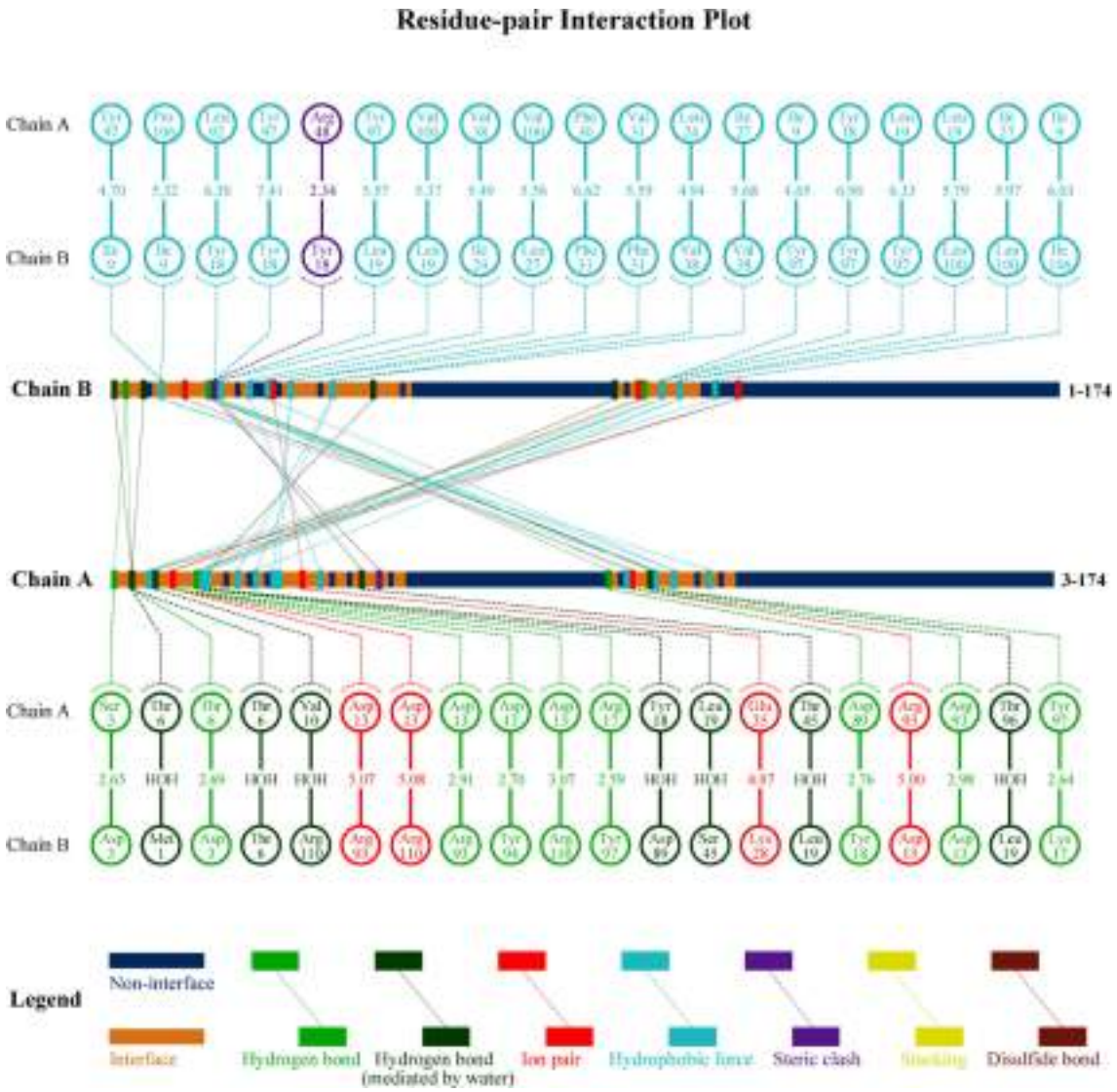

Figure 10. The schematic diagram for non-covalent interaction profile of allophycocyanin from *Arthrospira platensis* with view of the interface between subunits (PDB ID code 1all). The flattened diagram places atoms and bonds on the 2D page to minimize the overlap of atoms and the crossing of bonds in the final diagram. The Figure was prepared using the program 2D-GraLab v.1.0.[61]

computed free energies due to electrostatic interactions, hydrogen bonding and van der Waals contacts for each interacting pair in all the 118 phycocyanin interfaces. Interface binding occurs mostly through hydrogen bonds with 88.5% to the total energy, on average -2722.46 ± 1743.98 kJ mol^{-1} per interface. This finding is quite reasonable according to presence the water-bridged hydrogen bonds described in 3.3.1 section. These water molecules indeed contribute significantly to the total binding energy which imparts additional stability to these complexes. We observed an average energetic contribution of -302.35 ± 155.99 kJ mol^{-1} (9.8%) per interface in the group of van der Waals energy investigated in this work. The contribution of electrostatic energy is -51.93 ± 33.75 kJ mol^{-1} (1.7%) per interface. This observation is inconsistent with the previous analysis that vdW energies provide the major contribution to about 75% on average among all complexes.[60] However, this study did not consider the role of water molecules in the analysis. Water can play a significant role in bridging the residues within an interface pocket and contribute to the free energy of binding.

We performed multiple linear regression analysis of the number of interface residues with hydrogen bond energy, electrostatic energy, van der Waals energy and its total interface energies, and the results are shown in Figure 11.

Figures 11a and 11b show the total energy and H-bond energy in our dataset with a high correlation coefficient of 0.965 and 0.947 respectively. In contrast, we observed a poor correlation between vdW energy and the number of residues. It is interesting to note the correlation of electro-

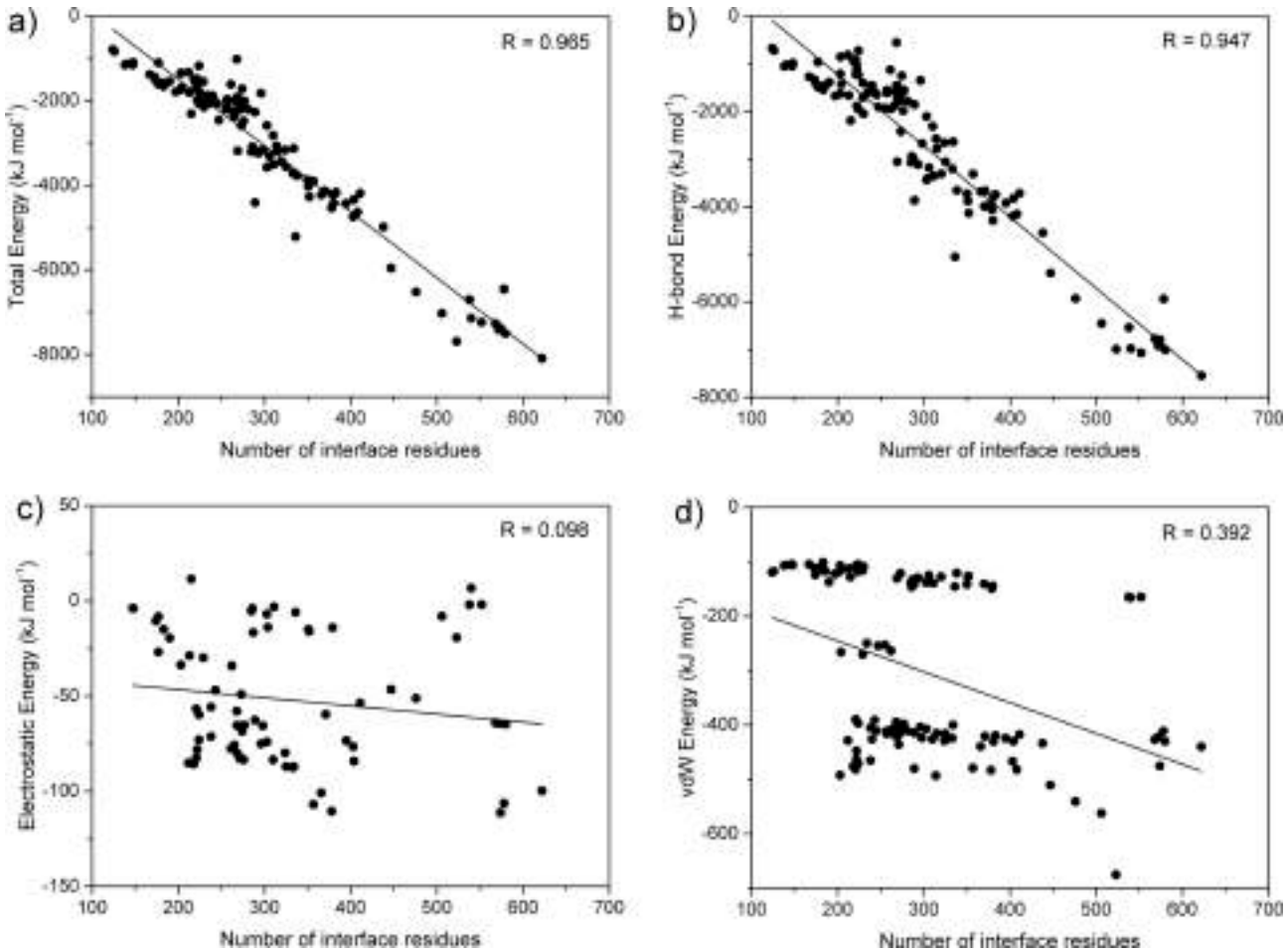

Figure 11. Correlation between the number of residues and interaction energy: a) total energy, b) hydrogen bond energy, c) electrostatic energy and d) van der Waals energy.

static energy is poor (R = 0.098). There are several interacting pairs with unfavorable and positive energy (Figure 11c). This is due to the strong force of repulsion between similarly charged residues at the interface. The repulsive nature of those interactions emerges from the unfavorable geometries of interacting pairs in the crystal structures and usually is counterbalanced by other interactions. Namely, when examined under isolated conditions, this type of interaction is considered unfavorable, but similar to other potentially unfavorable interactions, their influence can be compensated by other interactions from the rest of the polypeptide chain.[62] This observation indicates that hydrogen bonding free energy is important for maintaining the stability of phycocyanin proteins. However, the role of vdW and electrostatic energy could not be ignored. It has been reported that the vdW free energy due to carbon atoms and hydrogen-bonding energy play important roles to determine the folding rate in combination with other free energies.[63,64] The tendency of hydrophobic side chains to be buried away from water is commonly accepted as the main driving force involved in stabilizing the native state. Hydrophobic residues are among the most conserved amino acids in protein sequences and their solubility correlates with protein stability. In particular, increased solubilities of these residues at high and low temperatures as well as high pressure have been associated with heat, cold and pressure denaturations in proteins, respectively.[65,66] The special role of electrostatic energy between protein domains has been noticed. Hydrogen bonds and salt bridges are particularly essential in determining binding specificity.[59,67] Salt bridges occur frequently in proteins, providing conformational specificity and contributing to molecular recognition and catalysis.[59]

We have also analyzed the role of normalization for relating energy terms with interface size of phycocyanin proteins. The energy terms normalized by dividing with the number of residues in each protein are displayed in Figure 12.

We observed that the normalized energy range per residue was between −6 to −14 kJ mol^{-1}. There are few residues which have energies in the range −4 to −6 and −14 to −16 kJ mol^{-1} as well. The binding of two proteins is

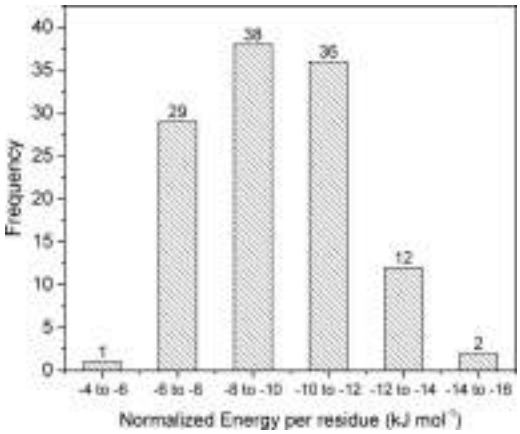

Figure 12. The distribution of the normalized energy per residue in phycocyanins.

related to interface size (number of interface residues involved in binding) and its corresponding interface area related to total interface energy.[32,60] The corresponding strength of non-covalent interactions to interface size of phycocyanins is of significance (Figure 12). The different trend was observed in the number of interface residues and normalized energy per residue value for obligatory as well as non-obligatory protein complexes as obtained by an earlier study.[68] An earlier analysis of the effect of chain length also demonstrates the importance of normalization for understanding the protein-folding rates.[32,64]

3.9 Conservation Score of Amino Acid Residues

Analysis of conservation patterns of protein-protein interfaces with respect to the protein surface has shown that the interfaces have been conserved more than the protein surfaces during the course of evolution,[69,70] and it is considered that structurally conserved residues are important in protein stability and folding.[71]

We next examined the evolutionary conservation of positions of interacting pairs in phycocyanin interfaces using the empirical Bayesian method. Our findings revealed that salt bridge forming residues are highly conserved with

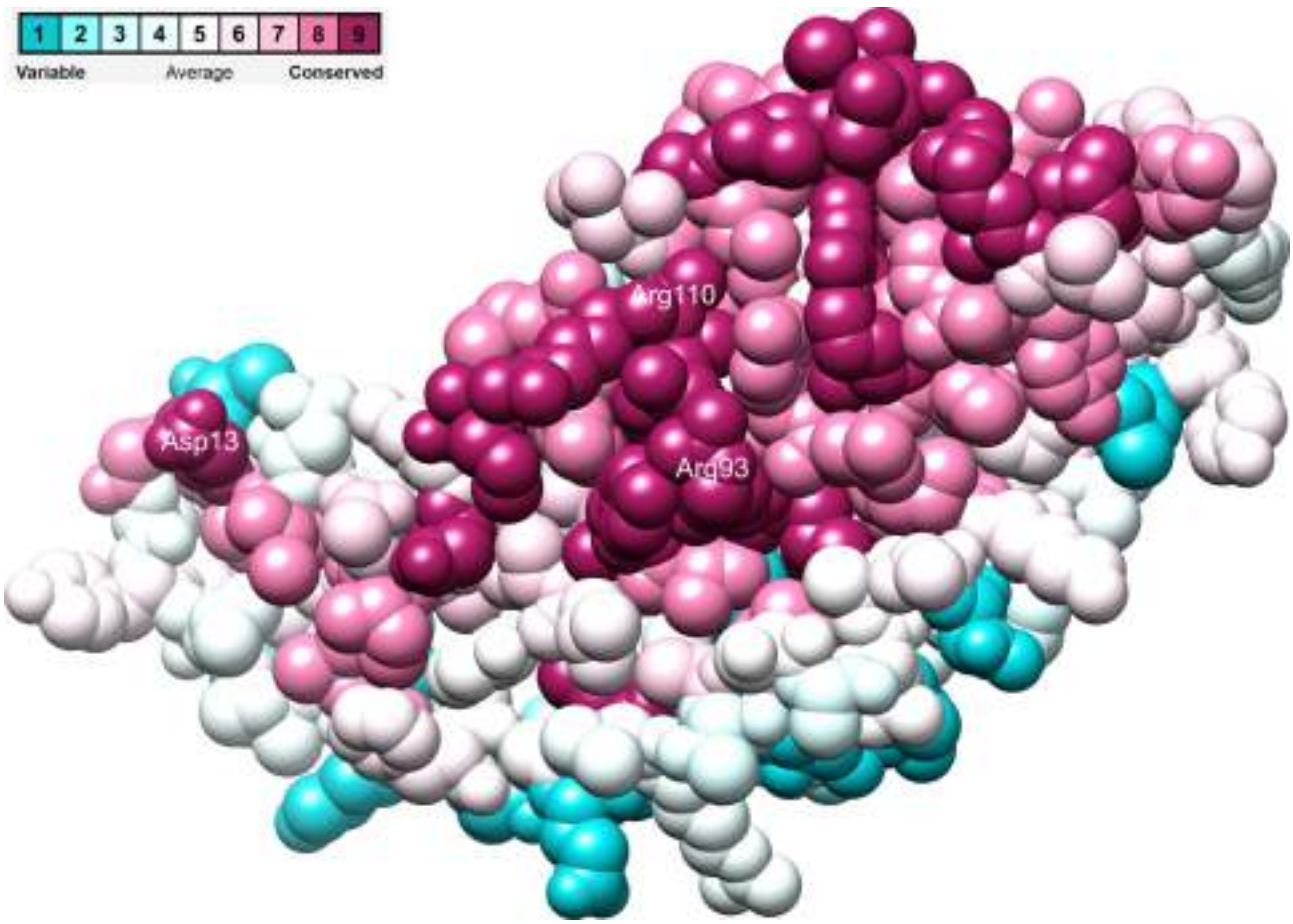

Figure 13. Conservation pattern of allophycocyanin from *Arthrospira platensis* (PDB ID code 1all; Chain B) using Chimera. Conservation score of B:Asp13, B:Arg93 and B:Arg110 residues is 9.

average conservation scores 7.3 ± 2.7. The calculated average conservation score for the amino acid residues involved in hydrogen bonds is 7.0 ± 2.5. The amino acid residues forming hydrophobic interactions and water-bridged hydrogen bonds have average conservation scores; 5.9 ± 2.1 and 5.9 ± 2.5, respectively. The average conservation score for the other interface residues was 5.1 ± 1.3, which is significantly lower than the value for the amino acid forming non-covalent interactions. The higher conservation of salt bridge and hydrogen bond forming residues compared to other non-covalent interaction residues is due to bond directionality, illuminating the importance of these residues in the functional and conformational preference.[59,72] From these observations, we were able to infer that the majority of interacting residues is evolutionary conserved and they might be important in maintaining the structural stability through non-covalent interactions in phycocyanins.

As a representative picture, the conservation grade of amino acid residues in allophycocyanin from *Arthrospira platensis* (PDB ID code 1all; Chain B) using Chimera[73] is shown in Figure 13. Conservation score of salt bridge interacting residues (B:Asp13, B:Arg93 and B:Arg110) is 9.

4 Conclusions

Protein-protein interactions are important in carrying out many biological processes and functions. Therefore, it is interesting to understand its molecular principles using known structural complexes with defined molecular functions. In our investigation, we have studied the role of non-covalent interactions in interfaces of phycocyanin proteins and their environmental preferences. The data show that the number of interface residues is clustered at the middle of the range which we call "standard-size" interfaces. The higher occurrence of hydrophobic residues, when compared to the other residues, is common, which is also the case for the BID test set. Although hydrophobic component of overall non-covalent interactions in protein interface contribute to the overall strength and rigidity of the assembly, interface binding occurs mostly through hydrogen bonds with 88.5 % to the total energy. This finding is quite reasonable according to the presence of multiple and water-bridged hydrogen bonds. These water molecules indeed contribute significantly to the total binding energy which imparts additional stability to these complexes. We observed a high correlation between total and H-bond energy with a number of interface residues. However, the role by vdW and electrostatic energy in interface binding could not be ignored. We found that the normalized interface energy per residue ranged from −6 to −14 kJ mol^{-1}. The corresponding strength of non-covalent interactions to interface size of phycocyanins is of significance. The high conservation score of amino acids that are involved in non-covalent interactions in phycocyanin interfaces is an additional strong argument for their importance. The overriding conclusion from this study is that the non-covalent interactions in phycocyanin interfaces considerably contribute to their high stability. This study would also be helpful in testing the efficacy of the existing biologically active molecules for the novel protein-protein interactions.

Conflict of Interest

None declared.

Acknowledgements

This work was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia (Grants Nos. 172001, 172035).

References

[1] J. De Las Rivas, C. Fontanillo, *PLoS Comput. Biol.* **2010**, *6*, e1000807.
[2] O. K. Mathew, R. Sowdhamini, *Bioinformatics* **2016**, *10*, 105–109.
[3] B. Skrlj, J. Konc, T. Kunej, *Mol. Inf.* **2017**, *36*, 1700017.
[4] O. Keskin, N. Tuncbag, A. Gursoy, *Chem. Rev.* **2016**, *116*, 4884–4909.
[5] I. M. Nooren, J. M. Thornton, *EMBO J.* **2003**, *22*, 3486–3492.
[6] R. P. Bahadur, P. Chakrabarti, F. Rodier, J. Janin, *J. Mol. Biol.* **2004**, *336*, 943–955.
[7] J. Mintseris, Z. Weng, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 10930–10935.
[8] H. Zhu, F. S. Domingues, I. Sommer, T. Lengauer, *BMC Bioinf.* **2006**, *7*, 27.
[9] P. Lavanya, S. Ramaiah, H. Singh, R. Bahadur, A. Anbarasu, *Comput. Biol. Med.* **2015**, *65*, 85–92.
[10] S. Jones, J. M. Thornton, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 13–20.
[11] C. L. Lo, C. Chothia, J. Janin, *J. Mol. Biol.* **1999**, *285*, 2177–2198.
[12] F. B. Sheinerman, R. Norel, B. Honig, *Curr. Opin. Struct. Biol.* **2000**, *10*, 153–159.
[13] F. Rodier, R. P. Bahadur, P. Chakrabarti, J. Janin, *Protein J.* **2005**, *60*, 36–45.
[14] O. Keskin, A. Gursoy, B. Ma, R. Nussinov, *Chem. Rev.* **2008**, *108*, 1225–1244.
[15] K. E. Riley, P. Hobza, *WIREs Comput. Mol. Sci.* **2011**, *1*, 3–17.
[16] M. N. de Tandeau, *Photosynth. Res.* **2003**, *76*, 193–205.
[17] P. Falkowski, R. J. Scholes, E. Boyle, J. Canadell, D. Canfield, J. Elser, N. Gruber, K. Hibbard, P. Hogberg, S. Linder, F. T. Mackenzie, B. Moore, III, T. Pedersen, Y. Rosenthal, S. Seitzinger, V. Smetacek, W. Steffen, *Science* **2000**, *290*, 291–296.
[18] A. McGregor, M. Klartag, L. David, N. Adir, *J. Mol. Biol.* **2008**, *384*, 406–421.
[19] F. E. Silva, F. D. S. Figueira, A. P. Lettnin, M. Carrett-Dias, D. M. V. B. Filgueira, S. Kalil, G. S. Trindade, A. P. S. Votto, *Pharmacol. Rep.* **2018**, *70*, 75–80.
[20] R. MacColl, *J. Struct. Biol.* **1998**, *124*, 311–134.

[21] D. Stanic-Vucinic, S. Minic, M. R. Nikolic, T. Cirkovic Velickovic in *Microalgal Biotechnology*, (Ed. J. L. Eduardo), IntechOpen, Rijeka, **2018**, pp. 129–149.
[22] M. G. de Morais, D. da Fontoura Prates, J. B. Moreira, J. H. Duarte, J. A. V. Costa, *Ind. Biotechnol.* **2018**, *14*, 30–37.
[23] R. R. Sonani, R. P. Rastogi, D. Madamwar, *World J. Biol. Chem.* **2016**, *7*, 100–109.
[24] P. W. Rose, B. Beran, C. Bi, W. F. Bluhm, D. Dimitropoulos, D. S. Goodsell, A. Prlic, M. Quesada, G. B. Quinn, J. D. Westbrook, J. Young, B. Yukich, C. Zardecki, H. M. Berman, P. E. Bourne, *Nucleic Acids Res.* **2011**, *39*, D392–D401.
[25] A. G. Murzin, S. E. Brenner, T. Hubbard, C. Chothia, *J. Mol. Biol.* **1995**, *247*, 536–540.
[26] E. Krissinel, K. Henrick, *J. Mol. Biol.* **2007**, *372*, 774–797.
[27] W. Kabsch, C. Sander, *Biopolymers* **1983**, *22*, 2577–2637.
[28] A. Shrake, J. A. Rupley, *J. Mol. Biol.* **1973**, *79*, 351–371.
[29] T. B. Fischer, K. V. Arunachalam, D. Bailey, V. Mangual, S. Bakhru, R. Russo, D. Huang, M. Paczkowski, V. Lalchandani, C. Ramachandra, B. Ellison, S. Galer, J. Shapley, E. Fuentes, J. Tsai, *Bioinformatics* **2003**, *19*, 1453–1454.
[30] G. Wang, R. L. Dunbrack Jr., *Bioinformatics* **2003**, *19*, 1589–1591.
[31] S. J. Darnell, D. Page, J. C. Mitchell, *Protein J.* **2007**, *68*, 813–823.
[32] A. Sukhwal, R. Sowdhamini, *Mol. BioSyst.* **2013**, *9*, 1652–1661.
[33] J. Novotny, R. E. Bruccoleri, M. Davis, K. A. Sharp, *J. Mol. Biol.* **1997**, *268*, 401–411.
[34] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, *J. Comput. Chem.* **1983**, *4*, 187–217.
[35] B. R. Brooks, C. L. Brooks, III, A. D. Mackerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, M. Karplus, *J. Comput. Chem.* **2009**, *30*, 1545–1614.
[36] M. Nardelli, *Comput. Chem.* **1982**, *6*, 139–152.
[37] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, J. Hermans in *Intermolecular Forces*, (Ed. B. Pullman), Springer, Dordrecht, **1981**, pp 331–342.
[38] H. Ashkenazy, E. Erez, E. Martz, T. Pupko, N. Ben-Tal, *Nucleic Acids Res.* **2010**, *38*, W529–W533.
[39] B. Boeckmann, A. Bairoch, R. Apweiler, M. C. Blatter, A. Estreicher, E. Gasteiger, M. J. Martin, K. Michoud, C. O'Donovan, I. Phan, S. Pilbout, M. Schneider, *Nucleic Acids Res.* **2003**, *31*, 365–370.
[40] J. Janin, F. Rodier, P. Chakrabarti, R. P. Bahadur, *Acta Crystallogr. Sect. D* **2007**, *63*, 1–8.
[41] G. R. Bickerton, *PhD thesis*, University of Cambridge, (UK), **2009**.
[42] S. Đ. Stojanović, B. L. Zarić, S. D. Zarić, *J. Mol. Model.* **2010**, *16*, 1743–51.
[43] F. Glaser, D. M. Steinberg, I. A. Vakser, N. Ben-Tal, *Protein J.* **2001**, *43*, 89–102.
[44] D. Xu, C. J. Tsai, R. Nussinov, *Protein Eng.* **1997**, *10*, 999–1012.
[45] S. Đ. Stojanović, E. R. Isenović, B. L. Zarić, *Mol. Inf.* **2011**, *30*, 430–442.
[46] A. S. Mahadevi, G. N. Sastry, *Chem. Rev.* **2016**, *116*, 2775–2825.
[47] R. P. Bahadur, P. Chakrabarti, F. Rodier, J. I. Janin, *Protein J.* **2003**, *53*, 708–719.
[48] S. K. Panigrahi, G. R. Desiraju, *Protein J.* **2007**, *67*, 128–141.
[49] N. Fox, I. Streinu, *BMC Bioinf.* **2013**, *14 Suppl 18*, S3.
[50] G. A. Jeffrey, W. Saenger in *Hydrogen bonding in biological structures*, Springer-Verlag, Berlin, **1991**.
[51] T. Okada, Y. Fujiyoshi, M. Silow, J. Navarro, E. M. Landau, Y. Shichida, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 5982–5987.
[52] S. Amiri, M. S. Sansom, P. C. Biggin, *Protein Eng. Des. Sel.* **2007**, *20*, 353–359.
[53] R. Kadirvelraj, B. L. Foley, J. D. Dyekjaer, R. J. Woods, *J. Am. Chem. Soc.* **2008**, *130*, 16933–16942.
[54] R. A. Laskowski, M. B. Swindells, *J. Chem. Inf. Model.* **2011**, *51*, 2778–2786.
[55] B. L. Zarić, V. B. Jovanović, S. Đ. Stojanović, *J. Theor. Biol.* **2011**, *271*, 18–26.
[56] G. A. Jeffrey in *An introduction to hydrogen bonding*, Oxford University Press, New York and Oxford, **1997**.
[57] B. Musafia, V. Buchner, D. Arad, *J. Mol. Biol.* **1995**, *254*, 761–770.
[58] J. N. Sarakatsannis, Y. Duan, *Protein J.* **2005**, *60*, 732–739.
[59] J. E. Donald, D. W. Kulp, W. F. DeGrado, *Protein J.* **2011**, *79*, 898–915.
[60] C. Nilofer, A. Sukhwal, A. Mohanapriya, P. Kangueane, *Bioinformation* **2017**, *13*, 164–173.
[61] P. Zhou, F. Tian, Z. Shang, *J. Comput. Chem.* **2009**, *30*, 940–951.
[62] V. R. Ribić, S. Đ. Stojanović, M. V. Zlatović, *Int. J. Biol. Macromol.* **2018**, *106*, 559–568.
[63] G. Calloni, N. Taddei, K. W. Plaxco, G. Ramponi, M. Stefani, F. Chiti, *J. Mol. Biol.* **2003**, *330*, 577–591.
[64] M. M. Gromiha, K. Saraboji, S. Ahmad, M. N. Ponnuswamy, M. Suwa, *Biophys. Chem.* **2004**, *107*, 263–272.
[65] C. L. Dias, T. Ala-Nissila, M. Karttunen, I. Vattulainen, M. Grant, *Phys. Rev. Lett.* **2008**, *100*, 118101.
[66] C. L. Dias, *Phys. Rev. Lett.* **2012**, *109*, 048104.
[67] B. Honig, A. S. Yang, *Advances in Protein Chemistry Protein Stability*; Academic Press: **1995**.
[68] S. De, O. Krishnadev, N. Srinivasan, N. Rekha, *BMC Struct. Biol.* **2005**, *5*, 15.
[69] O. Lichtarge, H. R. Bourne, F. E. Cohen, *J. Mol. Biol.* **1996**, *257*, 342–358.
[70] W. S. Valdar, J. M. Thornton, *Protein J.* **2001**, *42*, 108–124.
[71] W. L. DeLano, *Curr. Opin. Struct. Biol.* **2002**, *12*, 14–20.
[72] A. Shahi, E. Arunan, *J. Chem. Sci.* **2016**, *128*, 1571–1577.
[73] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605–1612.

Received: November 8, 2018
Accepted: August 26, 2019
Published online on September 18, 2019

