

Correlation of the Side-Chain Hubs with the Functional Residues in DNA Binding Protein Structures

R. Sathyapriya, K. V. Brinda, and Saraswathi Vishveshwara*

Molecular Biophysics Unit, Indian Institute of Science, Bangalore, Karnataka 560012

Received April 12, 2005

The structure networks of DNA-binding proteins have been constructed and analyzed. The detailed analysis of the networks indicates a strong relation between the positions of the residues interacting with DNA and those that form extensive interactions within the protein structure (called hubs). This study shows that the functional residues in these proteins are held in place by efficient scaffolding of the structure using side-chain interactions, thus highlighting the role of these side-chain hubs with respect to the functional residues in the protein structure.

1. INTRODUCTION

The precise location of critical functional residues in proteins can be upheld by providing a suitable structural framework that facilitates its function. The possible existence of only a limited number of folds that protein structures can take,¹ compared to a very large number of proteins with varied functions, suggests that the biological functions can extend beyond the fold level. The functions must have been encoded at the level of specific side-chain interactions. Further, recent investigations have shown that any polypeptide chain would adopt helical or sheet-like structures for the sake of optimal packing.² Thus, it reinforces the idea that higher specificity comes through noncovalent side-chain interactions within proteins. These noncovalent interactions have been known to play important roles in protein folding,^{3,4} stability,⁵ protein assembly, and in protein–protein interactions.⁶

The concept of protein structure networks is an attractive model to investigate the overall topology of proteins, which explicitly takes into account the amino acid side-chain interactions. Such a network representation allows us to explore the inter-relation between interacting residues in protein structure. Earlier, we had constructed protein structure graphs to identify a variety of clusters.^{7,8} Recently, we have investigated the clustering properties of residues interacting with specific bases of the DNA in DNA binding proteins.⁹ The study investigated the cluster forming properties of residues involved in a cation- π /Hbond stair interaction. In the present study, we investigate the relationship of the functional residues with respect to the overall topology of these proteins.

The present work involves the construction and analysis of the Protein Structure Graphs (PSG) of a set of 52 proteins engaged in interacting with DNA through the cation- π /Hbond stair motifs. This interaction is defined as an interaction of an amino acid residue (such as Arg) with two successive

bases of DNA (e.g. Gua) that are stacked in the double helix. This amino acid simultaneously forms a cation- π interaction with one of the bases of the DNA and a hydrogen bond interaction with the other to form the stair motif. This interaction had been characterized in a data set of protein–DNA complexes earlier¹⁰ to address the specificity of interactions of the amino acids (Arg, Lys, Asn, and Gln) with bases of DNA. The contribution of different types of interactions constituting the cation- π /Hbond stair motifs have been analyzed, and their relative contributions to the energetics in vacuum and in different solvents have been estimated by ab initio calculations.^{11,12} In the present study, the presence of hubs (an amino acid residue making connections with a large number of other side-chain residues) as well as their relation to that of amino acids residues involved in formation of stair motifs with the DNA is analyzed. The analysis of the locations of the hubs with respect to the functional residues has revealed an extremely important feature, which is related to the anchoring of such functional residues to the protein structure through these side-chain hubs. We believe that the present study will significantly enhance the understanding of the role of amino acid side-chain hubs in the proper positioning of functional groups in the structural networks of DNA binding proteins.

2. METHODOLOGY

The algorithm presented here consists of representing protein structures as graphs comprising a set of nodes and edges, where the amino acid residues are nodes and the strength of the noncovalent interactions between them determine the edges.⁸ Such a graph can then be analyzed in various ways to obtain information regarding clusters of amino acid residues as well as about the highly connected amino acid residues (known as hubs) involved in these protein structure networks as explained below.

2.1. Construction and Analysis of Protein Structure Graphs. Each amino acid in the protein structure is represented as a node, and the noncovalent interactions

* Corresponding author phone: 91 80 22932611; fax: 91 80 23600535; e-mail: sv@mbu.iisc.ernet.in.

existing between their side chains are evaluated for edge-formation as follows

$$I_{ij} = (n_{ij} \div (\min(N_i, N_j))) \times 100 \quad (1)$$

where n_{ij} is the number of distinct side-chain atom pairs of residues i and j coming within a distance of 4.5 Å, evaluated from the crystal structures. N_i and N_j are the normalization values for residue type i and j , which have been evaluated earlier from a nonredundant set of proteins.⁸ A cutoff value of interaction, I_{\min} (user defined), is then considered, and any ij residue pair, which has I_{ij} greater than I_{\min} , is connected in the graph with unit weight. For example, when an I_{\min} of 6% is used, all the connected residues in the graph interact with a value more than 6%. The protein structure graph (constructed based on the strength of a noncovalent interaction existing between the side chains) obtained for a given I_{\min} is represented mathematically as an unweighted (0,1) adjacency matrix. Distinct clusters of interacting residues are in turn obtained from this adjacency matrix by employing the Depth First Search (DFS)¹³ algorithm. The side-chain clusters can be obtained and analyzed at varying interaction cutoffs (I_{\min}). A higher I_{\min} indicates stronger interactions among the residues forming the cluster, whereas a lower I_{\min} indicated weaker interactions among the same. Clusters of biological significance, like the clusters at the protein–protein¹⁴ and protein–DNA interface,⁹ have been identified at the interaction cutoffs varying from 4 to 12%. A working cutoff of I_{\min} 6% was earlier used in the study for obtaining amino acid side-chain clusters at the protein DNA interface as this provided a trade off between the strength of interaction observed between the residues and the size of the clusters obtained.⁸

2.2. Identification of Hubs. Hubs are defined as highly connected nodes in a graph. The presence of hubs is a characteristic feature of many real-world networks, where they have been found to provide robustness to the networks from random attacks.¹⁵ Applying a similar principle to the protein structure networks, one would expect the hubs (amino acid residues) in these residue-based protein structure networks to play an important role in stabilizing the folded structure of the protein, and hence a targeted mutation of the hub residue may destabilize the protein structure.

The identification of hubs in a protein structure is carried out using a slightly different definition of protein structure graph, described as follows. In each ij residue pair, the number of distinct pairs of side-chain atoms coming within a distance of 4.5 Å is identified. Any two amino acid residues with at least one side-chain atom pair (contacts evaluated at I_{\min} 0%) coming within this distance are considered to be interacting in the protein structure in the present study. The contact number of a residue ‘ i ’ is defined as the total number of residue–residue interactions which it makes in the protein structure according to this criterion. Any residue, with a contact number of four or more is identified as a hub. The amino acid composition of hubs as well as the residues to which these hubs interact to in the data set is analyzed. The hydrogen bonds made by these hubs are also characterized with HBPLUS¹⁶ using standard geometries.

A data set of 52 DNA binding proteins and a list of amino acid residues that interact specifically with the bases of the DNA in all these proteins (77 in number) is obtained from

Rooman et al.¹⁰ for the present study [Table 1]. A few additional interactions identified from our previous study⁹ [9 in number] have also been considered. The coordinates of the protein–DNA complexes were obtained from the PDB.¹⁷ The figures of proteins in the manuscript are prepared using MOLSCRIPT.¹⁸

3. RESULTS AND DISCUSSION

3.1. Cluster Analysis. We had earlier constructed⁹ the structure graphs of proteins binding to DNA through cation- π /Hbond stair motifs¹⁰ and analyzed the cluster-forming properties of these proteins at the DNA binding interface at I_{\min} 6%. Important conclusions regarding the residues interacting with DNA, forming clusters with other aromatic and cationic residues of the protein, were derived from the earlier analysis. Additionally, a few new interactions with DNA were also detected through such analysis. However, no pattern was found relating the type of interacting clusters with the fold or of the function of proteins.⁹ The present study focuses on the identification of side-chain hubs in the vicinity of the residues interacting with DNA, with the idea of exploring the position of the interacting residue with respect to the tertiary structure of the protein.

3.2. Identification of Hubs in DNA-Binding Proteins. The hubs are defined as those amino acid residues, which have the contact number greater than or equal to four at $I_{\min} = 0\%$. We identify the hubs from all the proteins in this data set. Evaluation of the number of hubs in proteins revealed that about 38% of the amino acids in proteins from the data set are hubs. The ratio of the number of hubs in a protein to that of the size of the protein is presented in Figure 1.

Though the sizes of the proteins in the data set vary widely (from 50 to 765 residues), it is evident from the figure that approximately one-third of the residues are hubs in most proteins. Thus it is likely that these hub residues are significant from the perspective of the structure or the function of the protein when compared to residues that are not hubs.

3.3. Composition of Amino Acid Residues as Hubs. The amino acid composition of the hub forming residues in the complete data set is presented in Figure 2a. The frequency of a particular amino acid ‘ i ’ is normalized with respect to the total number of occurrences of that amino acid in the data set.

It can be seen from the figure that the hydrophobic residues such as Ile, Leu, and Val along with the aromatic amino acids, Phe, Tyr, Trp, and positively charged Arg predominantly form hubs in the data set. Lys is less predominantly seen as a hub compared to Arg though the occurrence of Lys is greater than that of Arg in the present data set. On the other hand, more than 90% Trp in the data set form hubs. The frequency of Cys and Met being hubs is also high. It is natural to expect that the hydrophobic hubs might contribute to the formation of the hydrophobic core in the protein structure, while the polar and the aromatic hubs are likely to contribute to the general stability as well as the specific functions of the protein. However, the relationship of these hubs to the DNA-binding interfaces of proteins is elaborated in the following section.

3.4. Relation of Hubs to the Functional Residues. A list of amino acid residues, which interact with specific bases

Table 1. List of Interacting Residues and the Hubs in Proteins^a

protein	residues (<i>i</i>) interacting with base of DNA	hubs (<i>i</i> ± 1), (<i>i</i> ± 2)
	ETS Domain	
1awcA	Arg 376	<i>Leu 374, Leu 378</i>
	Arg 379	<i>Leu 378</i>
1bc8C	Arg 61	<i>Leu 59, Leu 63, Ser 60</i>
	Arg 64	<i>Leu 63, Tyr 66</i>
1pueE	Arg 232	<i>Met 230, Leu 234</i>
	Arg 235	<i>Leu 234, Tyr 237</i>
	Phage Cro and Repressor	
1lmb3	Gln 44	<i>Met 42</i>
1lmb4	Asn 55	<i>Leu 57</i>
1rpeL	Gln 28	<i>Thr 26, Ser 30</i>
3croL	Gln 28^	<i>Val 26</i>
	Homeodomains	
1akhA	Asn 120	<i>Phe 118, Lys 121, Arg 122</i>
1akhB	Arg 185	<i>Arg 183, Arg 184</i>
1au7A		
	<i>Arg 49 ^</i>	<i>Ile 47, Phe 50, Glu 51</i>
1b72A	Gln 44	<i>Phe 42, Thr 46</i>
	<i>Asn 253</i>	<i>Phe 251, Arg 254, Arg 255</i>
1b72B	Asn 286	<i>Phe 284, Lys 287, Arg 288</i>
	Arg 290	<i>Arg 288, Tyr 291</i>
1fjIA	<i>Asn 51</i>	<i>Phe 49, Arg 52, Arg 53</i>
1mnmD	<i>Arg 185</i>	<i>Arg 183, Arg 184</i>
	<i>Asn 182</i>	<i>Arg 184, Arg 183</i>
2hddA	<i>Asn 51</i>	<i>Phe 49, Lys 52, Arg 53</i>
	<i>Lys 50</i>	<i>Trp 48, Phe 49, Lys 52</i>
9antA	<i>Arg 5</i>	<i>Gln 6</i>
	<i>Asn 51</i>	<i>Phe 49c, Arg 52, Arg 53</i>
	Rap1 DNA Binding Domain	
1ignA	Arg 404	<i>Ile 403, Arg 406</i>
	Arg 542	<i>Trp 541, Arg 544</i>
	Arg 546	<i>Arg 544, Phe 545, Phe 548</i>
	<i>Asn 401</i>	<i>Ile 403</i>
	REL Homology Domain	
1a3qA		
	<i>Arg 52^</i>	<i>Phe 51, Phe 53</i>
1a3qB	<i>Arg 54</i>	<i>Arg 52, Phe 53, Tyr 55</i>
	<i>Lys 221</i>	<i>Asp 219</i>
	<i>Lys 120</i>	<i>Val 122</i>
	<i>Arg 280^</i>	<i>Pro 278, Arg 282</i>
1tsrB	Arg 33^	<i>Phe 34, Arg 35</i>
	Methyl Transferase	
6mhtA	Arg 240^	<i>Glu 239, Ile 241, Tyr 242</i>
	Zinc Fingers	
1a1gA	Asn 121^	<i>Asn 121</i>
	Arg 124	<i>His 125, Ile 126</i>
	Arg 146^	<i>Phe 144</i>
	<i>Arg 174</i>	<i>Phe 172</i>
	<i>Arg 180</i>	<i>Arg 178, His 181</i>
1ubdC	Arg 342^	<i>Leu 340, His 343</i>
	<i>Gln 396</i>	<i>Phe 394</i>
	Lys 339	<i>Leu 340</i>
1meyC	Asn 19	<i>Leu 20</i>
	Gln 16	<i>Phe 14</i>
	Lys 22	<i>Leu 20, His 23</i>
	Lys 50	<i>His 51, Arg 53</i>
	Gln 44	<i>Phe 42</i>
	Arg 72^	<i>Phe 70</i>
	<i>Arg 78</i>	<i>Leu 76, His 79</i>
	Hormone Receptor	
1hcqA	<i>Lys 28^</i>	<i>Cys 27, Phe 30</i>
	Lys 32	<i>Phe 30, Phe 31, Ser 34</i>
1latA	<i>Arg 466</i>	<i>Phe 464</i>
	<i>Lys 461^</i>	<i>Cys 460, Phe 463</i>
2nllB	Arg 328^	<i>Phe 326, Phe 327, Arg 329, Thr 330</i>
	Endonuclease	
1a73A	Arg 74	<i>His 73, Trp 75</i>
	Gln 63	<i>Tyr 64, Trp 62</i>
	<i>Lys 65</i>	<i>Tyr 64, Arg 66</i>
3pviA	<i>Asn 140</i>	<i>Ile 139, Pro 142</i>
	<i>Asn 141</i>	<i>Ile 139, Pro 142</i>
1bhmA	<i>Arg 155</i>	<i>Thr 157</i>

Table 1 (Continued)

protein	residues (<i>i</i>) interacting with base of DNA	hubs ($i \pm 1$), ($i \pm 2$)
	Lambda Intergrase-like N Terminal Domain	
lcrxA	Arg 259 Lys 86	Leu 261 Val 85, <i>Ile 88</i>
	SRF-like Protein	
lmmmA	Lys 38	<i>Phe 36, Lys 40</i>
	TC3 Transposase	
ltc3C	Arg 236 [^]	Arg 234
	Trp Repressor	
lтроA	Arg 69	Gln 68, Leu 71
	Interferon Regulatory Factorlike	
2irfL	Lys 2075	Pro 2074, Trp 2077

^a [^]Interacting hubs. The noncluster forming residues interacting with DNA in column 2 and the hubs in column 3 that do not form connections with other residues at high I_{\min} are given in *italics*.

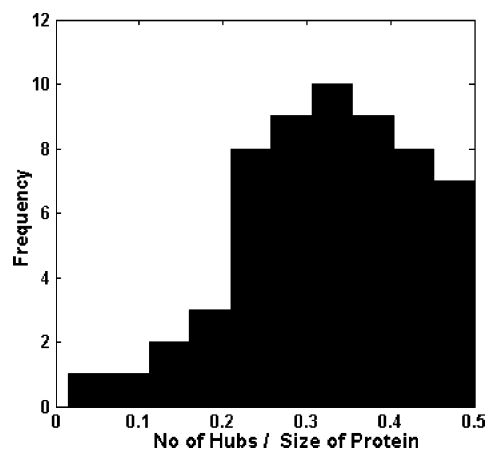


Figure 1. Histogram of the ratio of the number of amino acid hubs to the size of protein.

of the DNA through cation- π /Hbond stair motif interaction in the whole data set (86 in number) is obtained from the papers by Rooman¹⁰ and Sathyapriya.⁹ This stair motif is defined as the interaction of an amino acid (e.g. Arg), with a specific base of the DNA (e.g. G_i) through the cation- π interaction and with a neighboring base (e.g. G_{i+1}) through the hydrogen bond interaction when the bases G_i and G_{i+1} are themselves stacked in the DNA. The amino acid residues that can interact with a specific base of the DNA are henceforth called interacting residues. The ability of these interacting amino acid residues to form clusters has been analyzed earlier.⁹ The analysis had revealed that half the number of the residues (40/86) participating in the stair motif formation with bases formed clusters at the DNA binding region, whereas the remaining half did not. These were called as cluster forming and noncluster forming residues, respectively. Further, the cluster forming residues belonged to two different types, namely types I and II depending upon the presence of other aromatic amino acid residues in these clusters. In the present section, we correlate all the interacting residues in the proteins with the hubs identified in these proteins, and the results are consolidated in Table 1.

3.4.1. Interacting Residues as Hubs. The evaluation of the network parameters for the interacting residues has revealed that 14 (out of 86) amino acid residues that interact with a specific base of the DNA through the stair motif interaction are also hubs within the protein structure. These hubs are called as ‘interacting hubs’. They are mainly Arg

and Lys residues and are found in the following families of proteins: zinc fingers (1a1gA, 1meyC), REL homology proteins (1a3qA, 1tsrB, 2ramA), hormone receptors (1hcqA, 1latA, 2nllB), homeodomain (1au7A), TC3 transposase (1tc3C), Cro repressor (3croL), and methyl transferase (6mhtA). It is interesting to see that the amino acids that carry out the task of recognition of specific bases of DNA mainly Arg, Lys, Asn, and Gln are highly connected to other amino acid residues in these proteins. This might impart additional stability to these functionally important residues to carry out their biological function of DNA recognition and binding.

3.4.2. Hubs Flanking the Interacting Residues. In most cases, even if the interacting residues themselves are not hubs, the hubs are found flanking the residues that specifically interact with the base of the DNA. Such hubs are called as ‘flanking hubs’. Specifically, hubs are detected at positions $i \pm 1$ and $i \pm 2$ of an interacting residue which binds to the base of DNA. The list of the interacting residues in proteins and the hubs that are present flanking these interacting residues is given in Table 1. From the table it can be seen that about 60% of the interacting residues have more than one hub flanking them [residues in ($i \pm 1$ and/or $i \pm 2$) positions]. This is significant from the perspective that the functionally important residues are adjacent to the structural hubs and are thus anchored by residues having highly connected side-chain interactions.

The interacting residues are given in column 2 of Table 1, the interacting hubs being represented with a ‘[^]’. The hubs flanking these residues are given in column 3 of the table. We have also investigated whether the interacting residues and the flanking hubs form connections with other amino acid residues at a higher I_{\min} of 6%. We find that about 80% of the hubs that flank the interacting residues form connections with at least one other amino acid at high I_{\min} (form clusters at high I_{\min}). The noncluster forming residues (at high I_{\min}) in both interacting and flanking hubs are given in *italics* in Table 1. This is only a small fraction of the total number of the interacting and flanking hubs. This implies that the hubs that are flanking the interacting residues, anchoring them to perform their biological function, not only connect to several other amino acid residues in the vicinity but also interact strongly with at least a few of these residues.

The amino acid composition of the hubs flanking the interacting residues is given in Figure 2b. It can be seen from

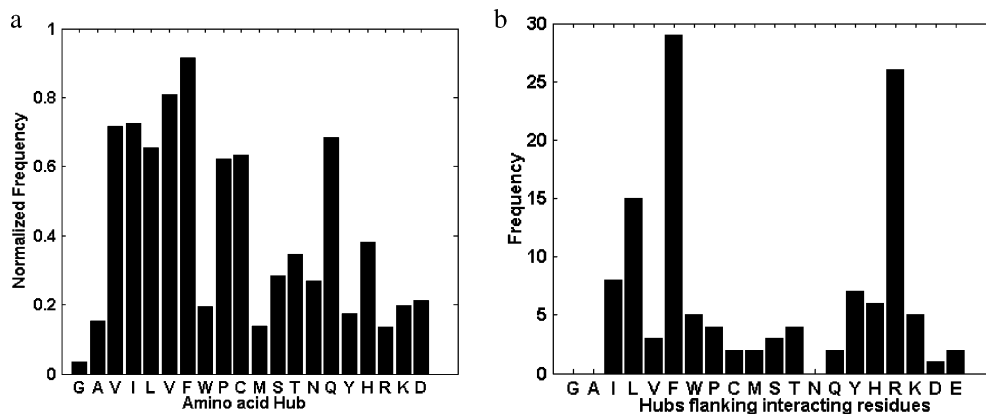


Figure 2. Composition of hubs in the DNA binding proteins. a. The normalized frequency for all the hubs is presented for all 20 amino acids. b. Composition of the hubs flanking the residues interacting with the DNA.

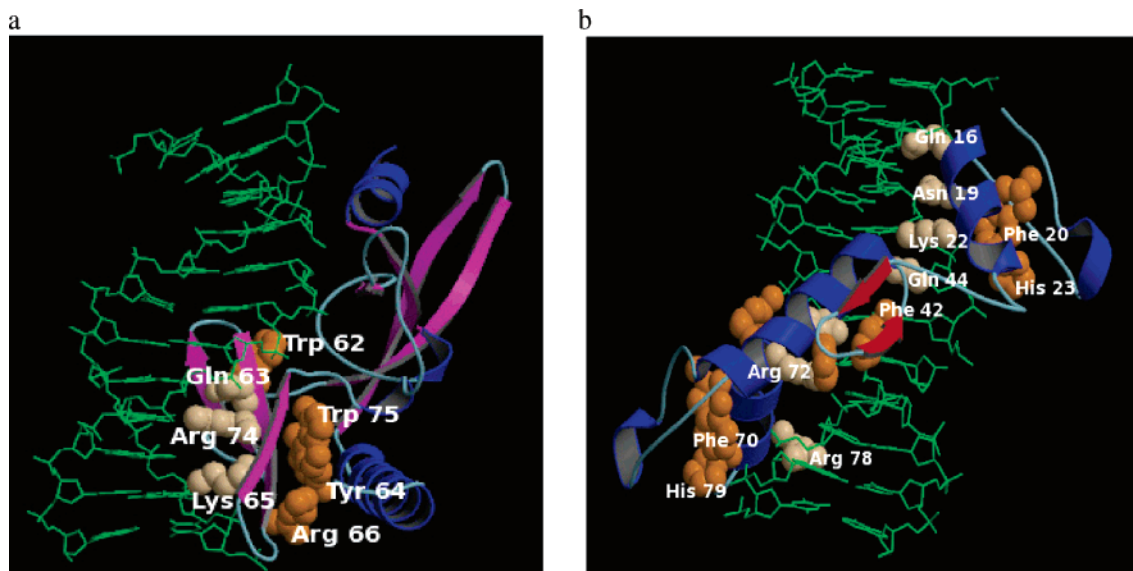


Figure 3. a. Hubs and interacting residues in homing endonuclease A(1a73A). The residues interacting with the bases of DNA (Gln 63, Arg 74, and Lys 65) are shown as white vdw spheres. The hubs flanking these interacting residues (Trp 62, Trp 75, Tyr 64, Arg 66) are shown as orange vdw spheres. The hubs are present at positions $i \pm 1$ of the interacting residue ' i '. b. Hubs and interacting residues in synthetic zinc finger construct (1meyC). The interacting residues are shown as white vdw spheres and the hubs in orange. The hubs are present at both $i-2$ and the $i+1$ positions of the interacting residue ' i '.

Figure 2b that there is an increased occurrence of Phe, Arg, Leu, and Ile in the hubs present at the vicinity of the interacting residue. Though Tyr, Trp, Val, Cys, and Met have high propensity to form hubs in the protein structure (Figure 2a), they are not very frequently present as the hubs flanking the interacting residue at the DNA binding interface of the protein (Figure 2b).

An example of interacting residues and the hubs flanking them is shown in Figure 3. As can be seen from the figure, in the case of endonuclease [(1a73A) Figure 3a], the interacting residues (white) are from an extended strand region facing the groove of the DNA, whereas the flanking hubs (orange) are facing the interior of the protein. The hubs are constituted from different secondary structures. It is interesting to note that Trp, which is never found as a part of an interacting cluster in all DNA binding proteins,⁹ is now detected as a hub in this example as well as in a few other cases [Table 1]. In Figure 3b, in case of the synthetic zinc finger (1meyC), the interacting residues as well as the hubs are from helical regions of the protein and face away from each other.

Thus side-chain hubs can either interact with a specific base of the DNA or occur flanking the residues that interact with the bases of DNA and impart structural stability to them.

3.5. Comparison of Interface and Noninterface Hubs.

The nature of hubs, its composition, and its relation to the residue interacting with a specific base of DNA has been analyzed in the previous sections. In this section, we have compared the features of flanking hubs with the other nonflanking hubs in the present data set. Specifically, the details of the residues with which the hubs interact at $I_{\min} = 0\%$ are characterized. This characterization has been carried out for the two cases of hubs separately (i) the hubs that either interact directly with specific bases of the DNA or occur in close proximity to these interacting residues, i.e., the interacting and the flanking hubs (Henceforth, both these categories have been referred to as interface hubs as they are found at the protein–DNA interface regions.). Though the flanking hubs are not exactly at the DNA binding interface, for the sake of convenience we refer to both interacting and flanking hubs as 'interface hubs'. (ii) The other hubs in the protein structure that neither interact with

Table 2. Comparison of Interface and Noninterface Hubs^a

property	interacting hubs	noninteracting hubs
total no.	138	1784
predominant hubs (%)	Arg (26.0) Phe (21.0) Leu (10.9), Val (2.4), Ile (6.4) Tyr (5.1) Lys (5.0) Glu (1.5)	Arg (8.1) Phe (7.6) Leu (18.1), Val (9.1), Ile (11.3) Tyr (7.1) Lys (2.2) Glu (4.4)
preference of residues interacting to hubs	Arg:- Phe \gg Glu \approx Tyr $>$ Pro Phe:- Leu $>$ Arg $>$ Phe \approx Tyr \approx Val Leu:- Leu \gg Tyr $>$ Phe \approx Val	Arg:- Glu $>$ Leu $>$ Arg \approx Asp \approx Ile Phe:- Leu \gg Ile \approx Phe $>$ Trp $>$ Arg \approx Tyr Leu:- Leu $>$ Ile $>$ Val \approx Phe $>$ Arg
hydrogen bonds average hbond/hub	1.22	1.03

^a All hubs and their interactions are identified at $I_{\min} = 0\%$.

a specific base of DNA nor ‘flank’ an interacting residue, are referred to as the ‘noninterface hubs’ (hubs not present at the protein–DNA binding region). The details of the composition of the residues that interact with the aforementioned types of hubs are analyzed in detail. Significant results of the analysis are given in Table 2. It can be seen from the table that the total number of interface and noninterface hubs are 138 and 1784, respectively. The amino acids Arg, Phe, Leu, Tyr, and Lys constitute more than 60% of the interface hubs. The amino acids Phe and to a lesser extent Glu, Tyr, and Pro prefer to interact with the interface Arg hubs. However, the interactions of the interface Phe hubs are more with Leu, Arg, and aromatic residues (Phe, Tyr). These preferences suggest that cation- π , hydrophobic, and aromatic stacking interactions are predominant in the interface hubs. Noninterface hubs in contrast are predominantly hydrophobic residues (Val, Leu, Ile) along with Arg, Phe, and Tyr (Table 2). The preference of amino acids that interact with noninterface Arg hubs is also quite different when compared to the case of interface Arg hubs. Glu interacts frequently with the noninterface Arg hubs followed by Leu, Arg, Asp, and Ile. It is clear that while the interface Arg hubs favor cation- π interactions, the same hubs in the noninterface regions prefer salt bridges and hydrophobic interactions. Noninterface Phe hubs prefer hydrophobic and aromatic stacking interactions (Leu, Ile, Phe, Trp, Tyr) than cation- π interactions with Arg, which was highly preferred by the interface Phe hubs. Thus the interface hubs exhibit significant differences in their interactions when compared to the noninterface hubs.

Apart from the noncovalent interactions, hydrogen bonds are also analyzed in the two categories of hubs. There are no significant differences observed between the interface and the noninterface hubs either in the hydrogen bond geometry, the type of atoms of the amino acid residue (such as backbone and side chain), or in the distribution of hydrogen bonds. It can be seen from Table 2 that an average of 1.22 and 1.03 hydrogen bonds is present in the interface and the noninterface hubs, respectively. A large proportion ($>50\%$) of hydrogen bonds is found involving backbone atoms in both the interface as well as the noninterface hub residues.

3.6. Similar Patterns of Hubs in a Given Family. The analysis of the amino acid composition of the hubs within the DNA-binding protein families has revealed that most of the members in some families exhibit a particular pattern of hub composition. For example, in ETS protein family, an interacting residue i has an $i-2$ Leu/Met and an $i+2$ Leu or

Tyr as a hub. Similarly, in zinc fingers, an interacting residue i has an $i+1$ His and $i-2$ Phe/Leu as a hub. Such patterns can be observed for other families such as homeodomains, repressors, and hormone receptors. Though it appears that the hubs flanking the interacting residues are conserved within a family, a thorough analysis of all the families including other homologous sequences is required to validate this observation.

4. CONCLUSIONS

Protein structure graphs are constructed for 52 DNA binding proteins based on the strength of noncovalent interactions between the amino acid side chains. These graphs are analyzed to detect correlations between the residues that interact with DNA and those that form highly connected nodes in the structure graphs (hubs). The details of the hub-forming residues, composition of the hubs, the preferences of residues interacting with these hubs, and the location of the hubs with respect to the DNA binding residues are presented in this paper.

Interestingly, we see that residues, which interact with a specific base of the DNA, are either hubs themselves or flanked by hubs in $i \pm 1$ and $i \pm 2$ positions. Further, many of these hubs have strong interactions with at least one of the residues with which they interact. The hubs interacting with the DNA and the flanking hubs have different preferences of residues interacting with them as compared to other hubs in the protein structure. Such hubs can be of biological significance since they show common patterns of occurrence in families of DNA-binding proteins.

The present analysis on the chosen data set elucidates that the amino acid hubs formed from side-chain interactions provide a structural scaffold to anchor the functional residues to the protein structure. This highlights the specific role played by side-chain interactions in scaffolding the DNA-binding residues at appropriate positions in the three-dimensional structure.

REFERENCES AND NOTES

- (1) Chothia, C. Proteins: One thousand families for the molecular biologist. *Nature* **1992**, *357*, 543–544.
- (2) Hoang, T. X.; Trovato, A.; Seno, F.; Banavar, J. R.; Maritan, A. Geometry and symmetry presculpt the free-energy landscape of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7960–4.
- (3) Onuchic, J. N.; Wolynes, P. G. Theory of protein folding. *Curr. Opin. Struct. Biol.* **2004**, *14*, 70–75.

- (4) Fersht, A. R.; Daggett, V. Protein folding and unfolding at atomic resolution. *Cell* **2002**, *108*, 573–582.
- (5) Shaw, A.; Bott, R. Engineering enzymes for stability. Engineering enzymes for stability. *Curr. Opin. Struct. Biol.* **1996**, *4*, 546–50.
- (6) Janin, J.; Wodak, S. Protein Modules and Protein–protein interactions. *Advances in protein chemistry*; Academic Press: 2003; Vol. 61.
- (7) Patra, S. M.; Vishveshwara, S. Backbone cluster identification in proteins by a graph theoretical method. *Biophys. Chem.* **2000**, *84*, 13–25.
- (8) Kannan, N.; Vishveshwara, S. Identification of side-chain clusters in protein structures by a graph spectral method. *J. Mol. Biol.* **1999**, *292*, 441–464.
- (9) Sathyapriya, R.; Vishveshwara, S. Interaction of DNA with clusters of amino acids in proteins. *Nucleic Acids Res.* **2004**, *32*, 4109–4118.
- (10) Rooman, M.; Lievin, J.; Buisine, E.; Wintjens, R. Cation– π /H bond stair motifs at protein–DNA interfaces. *J. Mol. Biol.* **2002**, *319*, 67–76.
- (11) Biot, C.; Wintjens, R.; Rooman, M. Stair motifs at protein–DNA interfaces: Nonadditivity of Hydrogen bonds, stacking and Cation– π interactions. *J. Am. Chem. Soc.* **2004**, *126*, 6220–6221.
- (12) Wintjens, R.; Biot, C.; Rooman, M.; Lievin, J. Basis Set and Electron Correlation Effects on ab initio calculations of Cation– π /H–Bond Stair Motifs. *J. Phy. Chem. A* **2003**, *107*, 6249–6258.
- (13) West, D. B. *Introduction to Graph Theory*; Prentice Hall of India Private Limited: 2000.
- (14) Brinda, K. V.; Kannan, N.; Vishveshwara, S. Analysis of homodimeric protein interfaces by graph-spectral methods. *Protein. Eng.* **2002**, *4*, 265–77.
- (15) Barabasi, A. L. *Linked: The new science of networks*; Persues Publishing: Cambridge, MA, 2002.
- (16) 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.
- (17) McDonald, I. K.; Thornton, J. M. Satisfying Hydrogen Bonding Potential in Proteins. *J. Mol. Biol.* **1994**, *238*, 777–793.
- (18) Kraulis, P. J. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **1991**, *24*, 946–950.

CI050122G