http://www.jstage.jst.go.jp/browse/biophysics

BIOPHYSICS

Vol. 3, pp. 27–35 (2007)
doi: 10.2142/biophysics.3.27

The role of charged surface residues in the binding ability of small hubs in protein-protein interaction networks

Ashwini Patil[1,2] and Haruki Nakamura[1]

[1]*Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan*
[2]*Bioinformatics Centre, University of Pune, Ganeshkhind Road, Pune 411007, India*

Received 13 March, 2007; accepted 1 June, 2007

Hubs are highly connected proteins in a protein-protein interaction network. Previous work has implicated disordered domains and high surface charge as the properties significant in the ability of hubs to bind multiple proteins. While conformational flexibility of disordered domains plays an important role in the binding ability of large hubs, high surface charge is the dominant property in small hubs. In this study, we further investigate the role of the high surface charge in the binding ability of small hubs in the absence of disordered domains. Using multipole expansion, we find that the charges are highly distributed over the hub surfaces. Residue enrichment studies show that the charged residues in hubs are more prevalent on the exposed surface, with the exception of Arg, which is predominantly found at the interface, as compared to non-hubs. This suggests that the charged residues act primarily from the exposed surface rather than the interface to affect the binding ability of small hubs. They do this through (i) enhanced intra-molecular electrostatic interactions to lower the desolvation penalty, (ii) indirect long – range intermolecular interactions with charged residues on the partner proteins for better complementarity and electrostatic steering, and (iii) increased solubility for enhanced diffusion-controlled rate of binding. Along with Arg, we also find a high prevalence of polar residues Tyr, Gln and His and the hydrophobic residue Met at the interfaces of hubs, all of which have the ability to form multiple types of interactions, indicating that the interfaces of hubs are optimized to participate in multiple interactions.

Key words: protein-protein interaction networks, hubs, electrostatic potential, multipole expansion, surface charge

Hubs are important proteins in protein-protein interaction networks given their ability to bind multiple partners. Their binding promiscuity has been attributed to the presence of disordered domains and highly charged surfaces[1]. Conformational flexibility from disordered domains is seen as the major contributor to the binding ability of large hubs (average length greater than 300 residues), while high surface charge is the predominant factor affecting the binding ability in small hubs[1]. Examples of such small hubs include Ubiquitin and Ferredoxin, which bind several proteins and have highly charged surfaces but no known disordered domains.

Previous studies on hubs have concentrated on large hubs with disordered regions[1–3]. Here, we focus on the small hubs (average length of 231 residues) without known disordered regions and extend our previous work to identify the role of charged residues in their binding promiscuity. Though the presence of high surface charges is known to promote binding promiscuity in small hubs, its reason is not yet clear. For instance, it is not known whether these charges have a specific distribution on the surface of the small hubs that allows them to bind multiple proteins, or if the charges are concentrated either at the binding interfaces or on the exposed surface or both, and how different these characteristics are in hubs as compared to non-hubs.

Corresponding author: Haruki Nakamura, Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan. e-mail: harukin@protein.osaka-u.ac.jp

Charged residues on protein surfaces have been known to contribute to binding through either long-range or short-range electrostatic interactions with residues on the partner proteins[4]. Along with polar residues, they are also known to contribute to binding specificity and complex stabilization[5], and have been recognized as hotspots on protein interfaces[6,7]. Charged residues are also more commonly found at the binding interfaces of proteins participating in transient interactions[8]. Given these properties of charged residues, we attempt to find out more about their role on the surfaces of small hubs.

In this study, we first find the distribution of charges on hub surfaces by mapping the surface electrostatic potential onto a sphere and performing multipole expansion to identify the dominant multipole moments. We calculate the relative enrichment of conserved residues, at the interface and on the exposed surface, of small hubs as compared to those in the entire proteins, to find the region where the charged residues are more prevalent. We also study the hydropathy of the interfaces and exposed surfaces of small hubs. We then compare the above characteristics of hubs with those of non-hubs. Using these analyses, we try to elucidate the role of charged surface residues in the binding ability of small hubs in the absence of disordered regions. There may be other factors that differentiate hubs from non-hubs, such as, their molecular function, levels of expression and degradation, and cellular localization. However, we limit our study to the structural features of hubs and non-hubs.

Materials and methods

Data set

We used the data set of hubs and non-hubs with structures in Protein Data Bank (PDB)[9] identified in Patil and Nakamura[1] as our base. The interactions were taken from the protein-protein interactions database, IntAct[10]. Interactions from high-throughput datasets were filtered to remove false positives[11]. For the purposes of this study, hubs were defined as proteins with 5 or more known interactions, while non-hubs were those with only 1 known interaction. We selected 50 hubs and 131 non-hubs, which satisfy the following criteria:

- protein has 10 or less disordered residues i.e. residues with missing electron density or missing structure in PDB,
- surface electrostatic potential of the protein was available through eF-site[12].

Of these, 35 hubs and 27 non-hubs had at least one complex structure with another protein. The selected set of 35 hubs had an average length of 231 residues and the set of 27 non-hubs had an average length of 273.

Multipole expansion of surface electrostatic potential

The multipole expansion of surface electrostatic potential was calculated for 50 hubs and 131 non-hubs. For each protein, the surface electrostatic potential was obtained from eF-site[12] as the potential at several points on the surface of the protein. This potential was then mapped onto a sphere of radius equal to the radius of gyration of the protein. Thus, for every point (x, y, z) on a protein with surface potential V, we calculated the spherical coordinates (R, θ, ϕ). Then, for each point (R_0, θ, ϕ), where R_0 is the radius of gyration of the protein plus 1.5 Å, we set the surface potential of the point to V. This gave us a sphere of radius R_0 mapped with the surface potential of the protein at all θ and ϕ. Using this sphere with the representative surface potential, we performed the multipole expansion and calculated the first 4 poles (monopole, dipole, quadrupole, octupole) on the surface of the sphere at 60 points which were randomly selected from all possible points on the sphere. The multipole expansion at any point (R_0, θ, ϕ), was performed as follows:

$$\begin{aligned} V(R_0) &= \sum_{l=0}^{\infty} Y_l(\theta,\phi)/R_0^{l+1} \\ &= Z/R_0 \\ &\quad + m_i R_{0i}/R_0^3 \\ &\quad + Q_{ij} R_{0i} R_{0j}/R_0^5 \\ &\quad + O_{ijk} R_{0i} R_{0j} R_{0k}/R_0^7 \\ &\quad + \ldots \end{aligned}$$

where $V(R_0)$ = surface potential at (R_0, θ, ϕ),
Z = monopole,
m_i = dipole moment,
Q_{ij} = quadrupole tensor,
O_{ijk} = octupole tensor,
R_{0i}, R_{0j}, R_{0k} = vectors of magnitude R_0, from the centre of the sphere to the points i, j and k on the sphere.

The spherical harmonics, $Y_l(\theta,\phi)/R_0^{l+1}$ are given by,

$$\begin{aligned} Y_l(\theta,\phi)/R_0^{l+1} &= a_l^0 P_l^0(\cos\theta) \\ &\quad + \sum_{m=1}^{l} [a_l^m P_l^m(\cos\theta)\cos m\phi + b_l^m P_l^m(\cos\theta)\sin m\phi] \end{aligned}$$

$$a_l^0 = (2l+1)\sum_{i=1}^{n} [V_i(R_0) P_l^0(\cos\theta_i)(\Delta S_i / \sum_{i=1}^{n} \Delta S_i)]$$

$$a_l^m = 2(2l+1)\frac{(l-m)!}{(l+m)!}\sum_{i=1}^{n} [V_i(R_0) P_l^m(\cos\theta_i)\cos m\varphi_i(\Delta S_i / \sum_{i=1}^{n} \Delta S_i)]$$

$$b_l^m = 2(2l+1)\frac{(l-m)!}{(l+m)!}\sum_{i=1}^{n} [V_i(R_0) P_l^m(\cos\theta_i)\sin m\varphi_i(\Delta S_i / \sum_{i=1}^{n} \Delta S_i)]$$

where P_l^m = associated Legendre function,
$V_i(R_0)$ = surface potential at point i, (R_0, θ, ϕ), on the sphere,
ΔS_i = surface area of a triangular region on the sphere surface centered at i,
$\sum_{i=1}^{n} \Delta S_i$ = total surface area of the sphere.
n = total number of points on the sphere as extrapolated from eF-site

We calculated the magnitude of the dipole contribution by adding the squares of the dipole moments, $Y_l(\theta,\phi)/R_0^2$, calculated at 60 random points on the sphere and then taking its square root. The quadrupole and octupole tensors were calculated in a similar manner. In order to check the validity of our method, we predicted the actual potential at each point on the surface of the sphere as the sum of the monopole, dipole, quadrupole and octupole tensors at that point and manually compared it with that on the original sphere. The number of positive charges on the surfaces of hubs and non-hubs, shown in Supplementary Tables S1 and S2, were calculated as the sum of the residues, Arg and Lys. The number of negative surface charges was calculated as the sum of Asp and Glu residues.

Conserved interface and exposed surface residues

An interface residue was identified as a residue on a chain with at least one atom at a distance of 3 Å or less from an atom of another residue on another chain. Exposed surface residues were those that were not interface residues and were identified by DSSP[13] as surface residues with a solvent accessible surface area greater than 60 Å^2. Conserved interface and exposed surface residues were those that had a conservation score of less than –0.5 as determined by Consurf-HSSP[14]. The correlation of the enrichment of conserved interface residues with the experimentally determined interface hotspot residues was calculated using the Pearson's correlation coefficient, r. A t-test with 49 degrees of freedom (for a sample size of 51 proteins) was used to determine the statistical significance, or p-value, of the correlation coefficient.

Enrichment of residues

The enrichment of conserved interface residues for 31 of 35 hubs and 20 of 27 non-hubs was calculated as:

$$\text{Conserved interface residue enrichment} = \frac{\text{propensity of conserved residue i at interfaces}}{\text{propensity of conserved residue i in proteins}} = \frac{\text{Nci(int)/Nc(int)}}{\text{Ni(tot)/N(tot)}}$$

where Nci(int) = number of conserved residues of type i at the interfaces of all proteins,
Nc(int) = total number of conserved residues at the interfaces of all proteins,
Ni(tot) = number of residues of type i in all proteins,
N(tot) = total number of residues in all proteins.

The conserved exposed residue enrichment was calculated in a similar manner over all conserved exposed residues.

Hydropathy

The average hydropathy of the interfaces of 35 hubs and 27 non-hubs was calculated as:

$$\text{Hydropathy} = \text{Hi} * \text{propensity of residue i at interfaces} = \text{Hi} * \text{Ni(int)/N(int)},$$

where Hi = hydropathy of residue i in the Kyte-Doolittle hydropathy index[15],
Ni(int) = number of residues of type i at the interfaces of all proteins,
N(int) = total number of residues at the interfaces of all proteins.

The hydropathy of the exposed surfaces of hubs and non-hubs was calculated in a similar manner using the residue propensity over all exposed residues. Hydropathy was calculated for all residues and not just conserved ones to take into account the entire protein surface. 95% confidence intervals were calculated for the mean hydropathy values at a significance of 0.05.

Results

Higher surface charge in hubs is known to affect their binding ability resulting in promiscuity[1]. We study the distribution of the charged residues on the surfaces of 50 hubs and 131 non-hubs. Of these, we use 35 hubs and 27 non-hubs (Supplementary Tables S1 and S2) to identify the role of charged residues on the surfaces of hubs (see Methods). For the purposes of this study, we define a hub as a protein with 5 or more known interactions, and a non-hub as a protein with 1 known interaction.

Charge distribution on hub surfaces

In order to determine the distribution of charges on the surfaces of hubs as compared to non-hubs, we performed multipole expansion of the surface electrostatic potential on 50 hubs and 131 non-hubs. As indicated in the Methods section, we calculated the monopole, dipole, quadrupole and octupole moments of the surface electrostatic potential of the sphere using multipole expansion. To validate our calculations, we predicted the electrostatic potential on the surface of the sphere using the values of the multipole moments obtained and checked this prediction against the actual values on the protein surface. Figure 1 shows the surface electrostatic potential of the protein, the potential mapped onto a sphere, the calculated multipole moments and the predicted surface potential from the multipole moments, for two proteins — Ubiquitin and Ubiquitin-like SMT3 precursor. It can be seen that predicted surface electrostatic potential roughly follows the trends of the actual potential, though sharp changes in the potential are lost. The values for individual multipole moments, for some of the hubs and non-hubs, are shown in Supplementary Tables S1 and S2.

Figure 2 shows the percentage of 50 hubs and 131 non-hubs with a dominant dipole, quadrupole and octupole moments. The combination of quadrupole and octupole moments is dominant in 64% of hubs and 63% non-hubs. Thus, on average, small hubs and non-hubs show a domi-

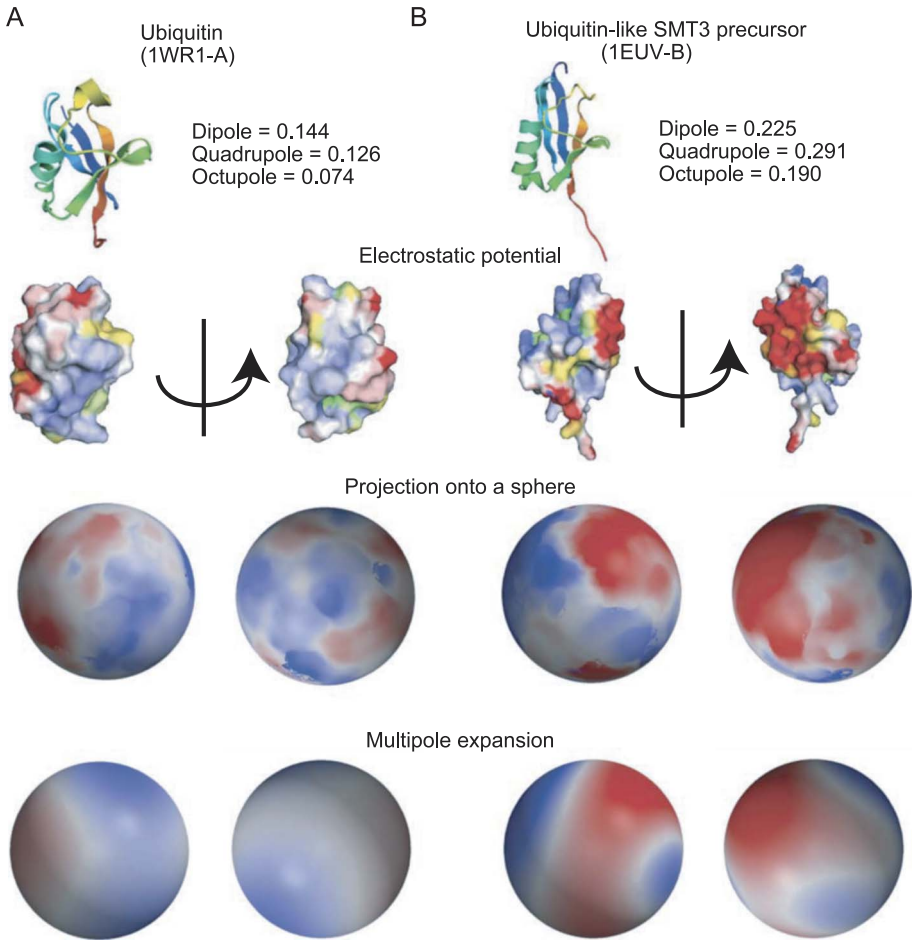

Figure 1 Projection of surface electrostatic potential onto a sphere, calculation of multipole moments by multipole expansion and prediction of the surface potential for (A) Ubiquitin and (B) Ubiquitin-like SMT3 precursor. Negative potential is indicated in red, positive potential in blue and hydropathy in yellow. Electrostatic potential on the protein surface was obtained from eF-site[12]. The electrostatic potential on the surface of the sphere were visualized using Molscript[33] and Raster3D[34]. Dipole, Quadrupole and Octupole values indicated are those calculated for 60 random points on the surface of the sphere.

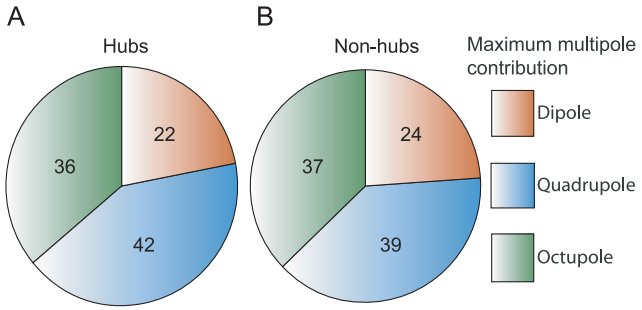

Figure 2 Percentage of hubs and non-hubs with a dominant dipole (green), quadrupole (blue) or octupole (orange) moment.

nant quadrupolar or octupolar nature in their surface charge distribution. However, we do not find any significant difference in the distribution of charges on the surfaces of hubs and non-hubs. A comparison of the monopole moments of hubs and non-hubs shows that 51.43% hubs have a positive monopole moment as compared to 40.74% non-hubs, indicating a net positive charge on the hub surface. The dominance of quadrupolar or octupolar moments in hubs can be interpreted by understanding the physical manifestation of each moment in the multipole expansion. The monopole moment corresponds to the net charge on the surface of the protein while the dipole moment corresponds to the average position of the positive and negative charges along each co-ordinate axis. The quadrupole and octupole moments represent the spread of charge from the co-ordinate axes. Thus a dominant quadrupolar or octupolar nature corresponds to a greater spread of charge over the surface of the sphere, and hence the surface of the protein. This suggests that the charged residues on the surfaces of hubs are well distributed over the surface. Since the surface charges are implicated in promiscuous binding of hubs[1] and they are spread out over the surface, it leads to two possible roles for the charged residues: (i) the charged residues act directly through short-range electrostatic interactions with the residues of the part-

Table 1 Relative enrichment of conserved residues at the interface and on the exposed surface (non-interface) of hubs and non-hubs

Residues	Hubs		Non-hubs	
	Interface	Exposed surface	Interface	Exposed surface
D	1.14	1.74	1.61	1.24
E	1.29	2.08	0.34	1.41
K	0.53	1.90	1.58	2.02
R	3.55	2.43	2.33	2.07
S	0.81	0.41	1.27	0.64
T	1.02	0.58	1.67	0.77
Y	2.27	1.04	1.15	0.67
H	1.39	1.24	0.98	1.43
C	0.79	0.19	0.91	0.33
N	0.91	1.50	1.85	0.90
Q	1.67	0.82	1.42	0.93
W	0.93	2.27	0.61	1.57
A	0.22	0.35	0.58	0.64
V	0.30	0.26	0.23	0.42
F	1.14	0.70	0.29	0.96
P	0.64	1.38	0.00	1.90
M	1.52	1.39	1.00	0.44
I	0.43	0.38	0.39	0.61
L	0.57	0.50	0.65	0.62
G	1.09	0.48	1.42	0.49

ner proteins through several different interfaces distributed over the surface of the hub to bind multiple partners, or (ii) the charged surface residues participate in long-range electrostatic interactions, from the exposed surface, with the residues of the partner proteins in the form of electrostatic steering[4], while using only a few specific interfaces for binding. In the following sections, we examine which of these roles is predominantly observed in small hubs.

Residue enrichment on hub surfaces

In order to determine if the charged residues on hub surfaces act, primarily, through multiple interfaces or through the exposed surface (non-interface), we studied the enrichment of conserved interface and exposed surface residues in hubs with respect to its total residue propensity. We looked for the conserved residues on the surface because we wanted to identify those that were important to protein function. We were able to identify conserved interface and exposed surface residues for 31 hubs and 20 non-hubs (Supplementary Tables S1 and S2). Table 1 and Figure 3 give the enrichment of the conserved residues at the interface and exposed surface of hubs and non-hubs. We also checked the correlation of the enrichment of conserved interface residues in all 51 proteins (31 hubs and 20 non-hubs) with the interface hotspots identified by Bogan and Thorn[6]. Our data correlates reasonably well with the experimentally determined hotspots at $r=0.60$ ($t=5.27$, $p=0.001$), when the outlier, Trp, is removed (Supplementary Table S3). The correlation coefficients for 31 hubs and 20 non-hubs with the interface hotspot data, independently, are 0.60 and 0.30 respectively.

Figure 3 shows that charged residues, Asp, Glu and Lys, are more prevalent on the exposed surface than at the interface of hubs. This suggests that the charged residues on hub surfaces are more likely to participate indirectly in binding,

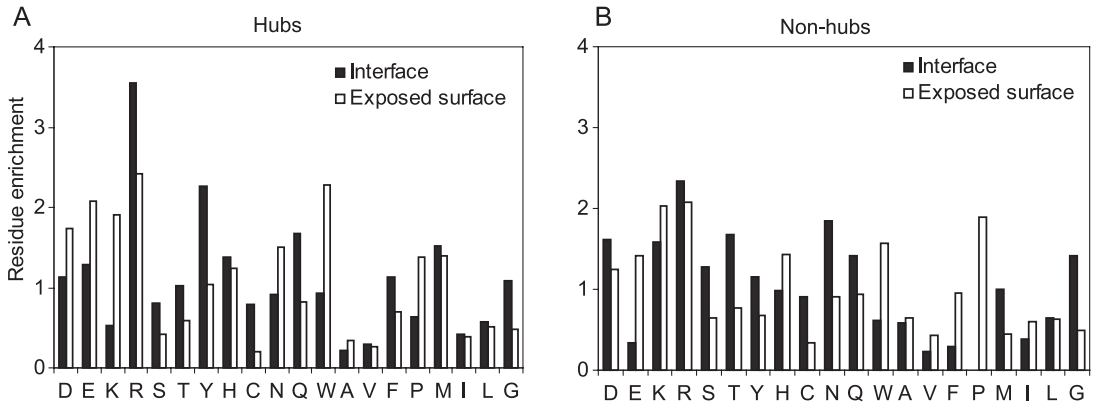

Figure 3 Relative enrichment of conserved interface (black) and exposed surface (white) residues in (A) hubs and (B) non-hubs.

Table 2 Average hydropathy of the interfaces and exposed surfaces of hubs and non-hubs. Negative values indicate hydrophilic nature, while positive values indicate hydrophobic nature. Values in brackets indicate the 95% confidence intervals at a significance of 0.05

Surface region	Hydropathy	
	Hubs	Non-hubs
Interface	−1.44 (±0.31)	−1.31 (±0.45)
Exposed surface	−1.86 (±0.28)	−1.46 (±0.27)

acting from the exposed surface. Arg appears to be a special case. It is the most abundant residue at the interfaces of both hubs and non-hubs. This is in agreement with previous studies which indicate Arg as the most common residue at protein interfaces[16]. However its enrichment at the interface is almost 1.5 times greater in hubs as compared to non-hubs. We also find an increased enrichment of Tyr, Gln, His and Met at the hub interface as compared to the non-hub interface. All of these have been previously identified as interface hotspot residues[6,7].

Hydropathy

We calculated the average hydropathy of the interfaces and exposed surfaces of 35 hubs and 27 non-hubs, using the amino acid propensity and the Kyte-Doolittle hydropathy scale[15]. Table 2 shows that both the interfaces and exposed surfaces are hydrophilic in hubs and non-hubs, as indicated by negative hydropathy values. As expected, the exposed surfaces are more hydrophilic than the interfaces. But the interfaces and exposed surfaces of hubs are more hydrophilic than those of non-hubs, further supporting the interpretation that the charges on the hub surfaces, though present at the interface, are more likely to act from the exposed surface.

Discussion

We used multipole expansion and residue enrichment studies to identify the role of the excess charged residues on the surfaces of hubs in their binding ability, as compared to non-hubs.

Electrostatics of binding

Using multipole expansion, we find that the charges are widely distributed over the surfaces of hubs. The residue enrichment results clearly imply that more charged residues are present on the exposed surfaces of hubs as compared to the interfaces. This indicates that charged residues on hub surfaces contribute to the promiscuity of hubs while primarily acting from the exposed surface. Clearly, this does not mean that hubs do not use multiple interfaces to bind their targets. But the results indicate that irrespective of the number of interfaces used, the charged residues, except Arg, are more likely to be found on the exposed surface than at the interface. For instance, in our dataset some hubs do have more than one interface, but generally do not have a very large number of interfaces spread across their surfaces. Figure 4 shows two such hubs, Ubiquitin and Ferredoxin, that have two binding sites each, using which they bind their known partners. Until recently, Ubiquitin was known to have only one binding site that it uses to bind its targets. This site is primarily hydrophobic and centered around Ile44. Several proteins were known to bind this interface somewhat differently with varying weak binding affinities[17]. A new site has now been found using which Ubiquitin binds Rabex5. This interface is slightly charged and centered around Asp58[18]. In the case of Ferredoxin, two highly charged binding sites are used to interact with Ferredoxin-$NADP^+$ reductase and sulfite reductase, respectively[19].

The charged residues on the exposed surface of a hub may increase its binding ability by sufficiently reducing the electrostatic binding free energy to make interactions with several proteins favourable. The electrostatic binding free energy of two proteins is the sum of (i) the desolvation contributions of the two proteins arising from the loss of interactions of the charged residues with the solvent at the interface, and (ii) contributions of solvent screened electrostatic interactions in the bound state between the charged residues at the interfaces of the two proteins in the form of salt bridges and hydrogen bonds. Of these, the first component is unfavourable for binding while the second one is favourable and accounts for the effect of charged residues at the interface[20]. The net effect of electrostatics at the interface is generally to destabilize protein binding due to the large desolvation penalties. However, in several cases, this desolvation penalty is reduced by favourable intra-molecular interactions amongst the surrounding charged residues of one or both the proteins[21]. Intra-molecular interactions are electrostatic interactions, exposed during binding, that take place among functional groups within the same binding partner, as opposed to inter-molecular interactions that take place across the interfaces of two binding partners[22]. Thus, along with the charged residues at the interface, those that are present around the interface also play an important role in reducing the electrostatic binding free energy thus promoting binding[21].

Charged residues on the exposed surface are also known to make significant contributions to binding through long-range electrostatic interactions[22]. These “action-at-a-distance” interactions, or electrostatic steering, act through regions of solvent and thus improve the overall complementarity of the interfaces of the interacting proteins and modulate the binding free energy to achieve an appropriate level of binding affinity[23]. Thus the charged residues on the exposed surface of a hub can increase its binding ability considerably, through enhanced intra-molecular interactions and long-range inter-molecular electrostatic interactions, without participating in direct contact with the interfaces of the partner proteins.

Hydropathy results show that the interfaces of hubs are

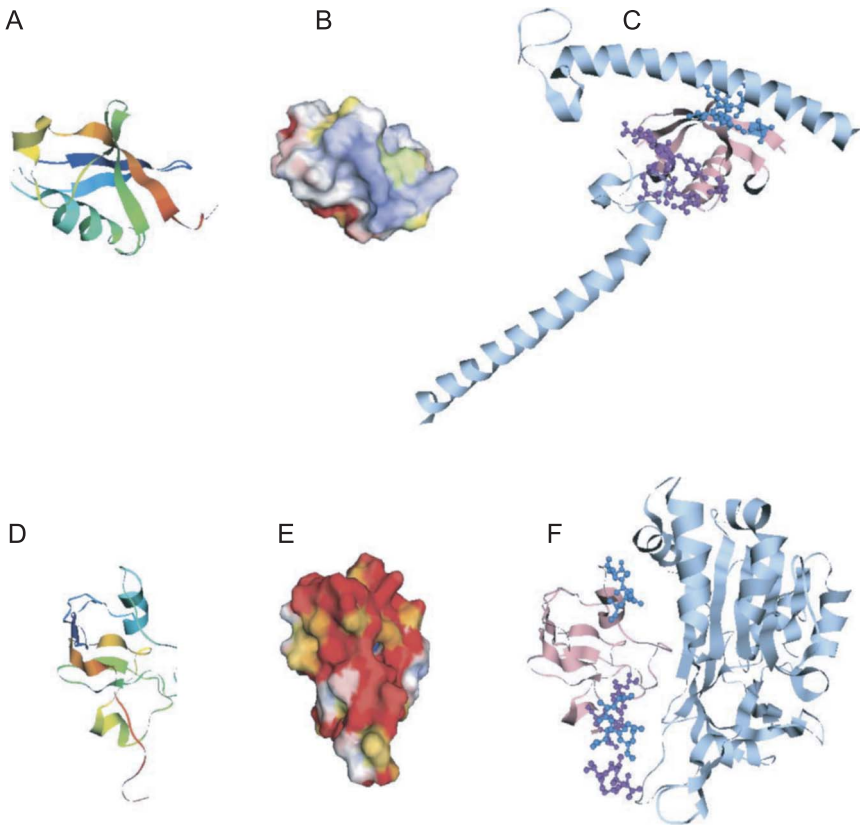

Figure 4 Binding interfaces of hubs and non-hubs. (A) Ubiquitin (PDB ID: 1WR1-A) (B) Surface potential of Ubiquitin (1wr1_1-A from eF-site[12]). (C) Ubiquitin (pink) bound to two molecules of Rabex5 (light blue) using distinct binding interfaces shown in blue (Lys6, Thr7, Gly10, Arg42, Ile44, Ala46, Gly47, His68, Val70) and purple (Ser20, Glu51, Arg54, Thr55, Ser57, Asp58, Tyr59, Asn60, Gln62), respectively (PDB ID: 2C7N). (D) Ferredoxin (PDB ID: 1GAQ-B) (E) Surface potential of Ferredoxin (1gaq-B from eF-site[12]). (F) Ferredoxin (pink) bound to Ferredoxin $NADP^+$ reductase (FNR) (light blue) using the interface residues shown in blue (Gln61, Leu64, Asp65, Asp66, Gln68, Leu95, Thr96, Gly97, Ala98). Purple residues (Glu29, Glu30, Asp34, Glu92, Glu93, Glu94) show the binding site of Ferredoxin to Sulphite reductase (SiR) (PDB ID: 1GAQ). Red indicates negative potential, blue indicates positive potential and yellow indicates hydropathy in B and E. Figure created using jV 3.2[12].

more hydrophilic than those of non-hubs. Along with the results of enrichment in Figure 3, this indicates that even though the charged residues on hub surfaces predominantly act from the exposed surfaces, they are more enriched at the hub interfaces as compared to non-hubs. Thus, the interfaces of hubs have some charged residues that participate in direct binding. But the bulk of the charged residues are present on the exposed surface. An example of this is the hub Ferredoxin, which has two highly charged interfaces along with a very charged exposed surface (Figure 4D, E, F).

Interface flexibility and charge

The high prevalence of Arg at the hub interface provides a lot of binding flexibility to the hub. Arg forms multiple types of favourable interactions[6]. The Arg side chain is considerably flexible, forming a good anchor residue as a result of its ability to anchor its side chain into a binding groove of an interaction partner[24]. It also allows binding across interfaces in various orientations. The dihedral angle, CG-CD-NE-CZ, of Arg varies through a spectrum of values, from –120, –60, 60, 120, to 180 degrees, in the structures in our data set (Supplementary Table S4). For instance, a superposition of all structures of Ubiquitin with its various interaction partners binding the interface centered at Ile44, shows the various side-chain orientations for Arg42 which also participates in binding (Figures 5A, B). Along with salt bridges[25] and hydrogen bonds[16], Arg is also involved in cation-π interactions with aromatic residues[16,26] across the interface. A cation-π interaction of Arg with Tyr in the hub, Son of Sevenless-1, is shown in Figure 5C. Arg also provides increased specificity of interaction[16].

Met is also found predominantly at hub interfaces. It is enriched in interaction hotspots, and contributes favourably to the binding free energy through reduced entropic penalty due to the lack of rotatable bonds and through hydrophobic effects[6,7,25]. Met, along with Arg, is a good anchor residue with a flexible side chain[24]. Tyr and His, which are enriched at hub interfaces, are known interface hotspots[6,27]. Tyr has the ability to form multiple interactions[6], while His has the ability to form hydrogen bonds[27]. Thus, clearly, hub inter-

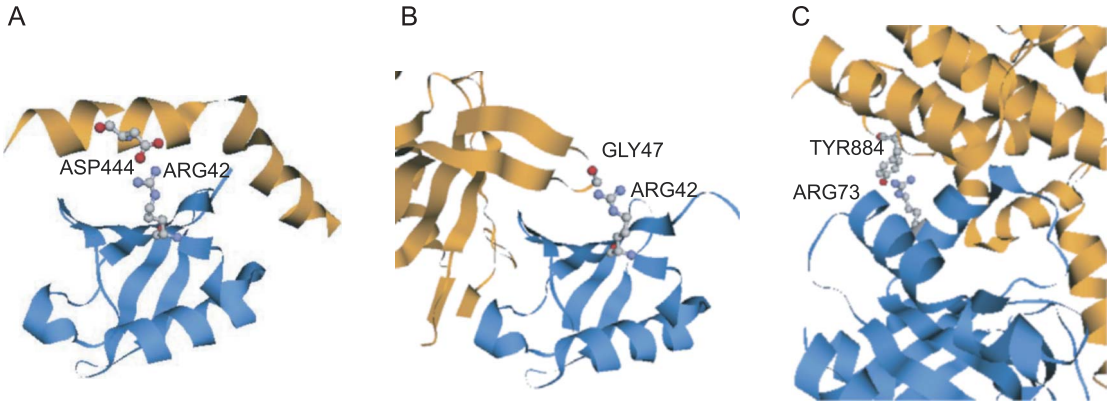

Figure 5 Selected views of Arginine participating in various types of interactions in different binding orientations. (A) Arg42 of Ubiquitin (blue) participating in a salt bridge with Asp444 of the CUE domain of Vacuolar protein sorting-associated protein VPS9 (orange). The dihedral angle, CG-CD-NE-CZ, of Arg42 is 131.9 degrees. (PDB ID: 1P3Q). (B) Arg42 of Ubiquitin (blue) forming a hydrogen bond with Gly47 backbone (distance 2.8 Å) of Tumor susceptibility gene 101 protein TSG101 (orange). The dihedral angle, CG-CD-NE-CZ, of Arg42 is –143.4 degrees. (PDB ID: 1S1Q). (C) Arg73 of Son Of Sevenless-1 (blue) participating in a cation-π interaction with Tyr884 of H-RAS (orange). The dihedral angle, CG-CD-NE-CZ, of Arg73 is 124.2 degrees. (PDB ID: 1BKD) Figure created using jV 3.2[12].

faces prefer residues that allow formation of multiple flexible interactions to further increase their binding ability. Gln is also enriched at hub interfaces. Though it is not an interface hotspot residue, it has been previously found to be enriched at the interfaces[7,27].

Charged and polar residues at interfaces have been associated with transient or non-obligatory interactions[8,28]. They are also enriched in small sized interfaces[25]. This goes well with the functional requirements of small hubs since they need to bind multiple partners during their lifetime and hence participate predominantly in transient interactions. Due the highly charged nature of their exposed surfaces, the binding affinity of the hubs participating in these transient interactions may also be greatly affected by changes in the physiological conditions in the cellular compartment (e.g. pH, temperature, concentrations of ions, chemicals and proteins)[29].

Though we have used multiple complex structures for hubs, where available, the determination of interface and exposed surface residues is limited by the absence of several complex structures for hubs with multiple interfaces. However, there is still no consensus on whether most hubs use a single interface or multiple interfaces [30,31], though hubs with a single interface are predominant in our data set.

The correlation coefficients of the conserved interface residues of hubs and non-hubs with those of previously determined interface hotspot residues indicate that these reflect the trend at hub interfaces only. Thus, the enrichment of residues at non-hub interfaces are not adequately represented by interface hotspot residues determined by Bogan and Thorn[6], Hu et al.[27] and Ma et al.[7].

Solubility

Solubility of proteins is known to be important for their functions[32]. The charged residues on the surfaces of hubs may also increase their binding ability by providing increased surface hydrophilicity and hence increased solubility. This increased solubility may significantly enhance association rates of hubs that are normally limited by rates of diffusion in the solvent[5].

Conclusion

We show that hubs have more charged interfaces and exposed surfaces as compared to non-hubs. Using multipole expansion of electrostatic potential, we show that the combination of quadrupole and octupole moments is dominant in small hubs. This indicates that the charges on the surfaces of hubs are spread out over their surfaces. This distributed charge is predominantly found on the exposed surface and most likely increases the promiscuity of small hubs through short-range intra-molecular interactions and long-range action-at-a-distance interactions through the solvent. The interfaces of hubs are enriched in Arg, most probably due to its ability to form multiple interactions. The interfaces are also enriched in other polar residues, Try, His, Gln and the hydrophobic Met, which provide a hydrophobic effect or the ability to form multiple interactions. The absence or relative reduction of the characteristics discussed above, in non-hubs, further stresses their role in the promiscuity of hubs. Thus the charges on the exposed surfaces and at the interfaces of hubs are optimized for participation in the interactions with several other proteins. Undoubtedly, these factors must also affect the binding ability of large hubs even if they have the flexibility afforded by disordered regions, though we have limited the current study to small hubs.

Acknowledgements

We would like to thank Dr. Kengo Kinoshita of Tokyo

University for his invaluable help with eF-site and jV 3.2, and Dr. Daron Standley from PDBj for his help with the analysis of PDB files. This study was supported by the grant-in-aid from the Institute for Bioinformatics Research and Development of the Japan Science and Technology Agency (BIRD-JST), and by the grant-in-aid for Scientific Research on priority areas No. 18054013 from the Ministry of Education, Science, Sports and Culture of Japan. AP would like to thank the Department of Biotechnology, India, for its support through the Centre Of Excellence grant.

References

1. Patil, A. & Nakamura, H. Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks. *FEBS Lett.* **580**, 2041–2045 (2006).
2. Haynes, C., Oldfield, C. J., Ji, F., Klitgord, N., Cusick, M. E., Radivojac, P., Uversky, V. N., Vidal, M. & Iakoucheva, L. M. Intrinsic Disorder Is a Common Feature of Hub Proteins from Four Eukaryotic Interactomes. *PLoS Comput. Biol.* **2**, e100 (2006).
3. Dunker, A. K., Cortese, M. S., Romero, P., Iakoucheva, L. M. & Uversky, V. N. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.* **272**, 5129–5148 (2005).
4. Sinha, N. & Smith-Gill, S. J. Electrostatics in protein binding and function. *Curr. Protein. Pept. Sci.* **3**, 601–614 (2002).
5. Sheinerman, F. B., Norel, R. & Honig, B. Electrostatic aspects of protein-protein interactions. *Curr. Opin. Struct. Biol.* **10**, 153–159 (2000).
6. Bogan, A. A. & Thorn, K. S. Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **280**, 1–9 (1998).
7. Ma, B., Elkayam, T., Wolfson, H. & Nussinov, R. Protein-protein interactions: Structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5772–5777 (2003).
8. Jones, S. & Thornton, J. M. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13–20 (1996).
9. 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.* **28**, 235–242 (2000).
10. Hermjakob, H., Montecchi-Palazzi, L., Lewington, C., Mudali, S., Kerrien, S., Orchard, S., Vingron, M., Roechert, B., Roepstorff, P., Valencia, A., Margalit, H., Armstrong, J., Bairoch, A., Cesareni, G., Sherman, D. & Apweiler, R. IntAct: an open source molecular interaction database. *Nucleic Acids Res.* **32**, D452–455 (2004).
11. Patil, A. & Nakamura, H. Filtering high-throughput protein-protein interaction data using a combination of genomic features. *BMC Bioinformatics* **6**, 100 (2005).
12. Kinoshita, K. & Nakamura, H. eF-site and PDBjViewer: database and viewer for protein functional sites. *Bioinformatics* **20**, 1329–1330 (2004).
13. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).
14. Glaser, F., Rosenberg, Y., Kessel, A., Pupko, T. & Ben-Tal, N. The ConSurf-HSSP database: The mapping of evolutionary conservation among homologs onto PDB structures. *Proteins* **58**, 610–617 (2005).
15. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).
16. Crowley, P. B. & Golovin, A. Cation-π interactions in protein-protein interfaces. *Proteins* **59**, 231–239 (2005).
17. Hicke, L., Schubert, H. L. & Hill, C. P. Ubiquitin-Binding Domains. *Nat. Rev. Mol. Cell. Biol.* **6**, 610–621 (2005).
18. Harper, J. W. & Schulman, B. A. Structural complexity in ubiquitin recognition. *Cell* **124**, 1133–1136 (2006).
19. Saitoh, T., Ikegami, T., Nakayama, M., Teshima, K., Akutsu, H. & Hase, T. NMR Study of the Electron Transfer Complex of Plant Ferredoxin and Sulfite Reductase: Mapping the interaction sites of Ferredoxin. *J. Biol. Chem.* **281**, 10482–10488 (2006).
20. Lee, L.-P. & Tidor, B. Optimization of electrostatic binding free energy. *J. Chem. Phys.* **106**, 8681–8690 (1997).
21. Chong, L. T., Dempster, S. E., Hendsch, Z. S., Lee, L. P. & Tidor, B. Computation of electrostatic complements to proteins: a case of charge stabilized binding. *Protein Sci.* **7**, 206–210 (1998).
22. Lee, L. P. & Tidor, B. Optimization of binding electrostatics: charge complementarity in the barnase-barstar protein complex. *Protein Sci.* **10**, 362–377 (2001).
23. Joughin, B. A., Green, D. F. & Tidor, B. Action-at-a-distance interactions enhance protein binding affinity. *Protein Sci.* **14**, 1363–1369 (2005).
24. Rajamani, D., Thiel, S., Vajda, S. & Camacho, C. J. Anchor residues in protein-protein interactions. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 11287–11292 (2004).
25. Tsai, C. J., Lin, S. L., Wolfson, H. J. & Nussinov, R. Studies of protein-protein interfaces: A statistical analysis of the hydrophobic effect. *Protein Sci.* **6**, 53–64 (1997).
26. Gallivan, J. P. & Dougherty, D. A. Cation-π interactions in structural biology. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9459–9464 (1999).
27. Hu, Z., Ma, B., Wolfson, H. & Nussinov, R. Conservation of polar residues as hot spots at protein interfaces. *Proteins* **39**, 331–342 (2000).
28. De, S., Krishnadev, O., Srinivasan, N. & Rekha, N. Interaction preferences across protein-protein interfaces of obligatory and non-obligatory components are different. *BMC Struct. Biol.* **5**, 15 (2005).
29. Nooren, I. M. A. & Thornton, J. M. Diversity of protein-protein interactions. *EMBO J.* **22**, 3486–3492 (2003).
30. Kim, W. K., Henschel, A., Winter, C. & Schroeder, M. The Many Faces of Protein-Protein Interactions: A Compendium of Interface Geometry. *PLoS Comput. Biol.* **2**, e124 (2006).
31. Keskin, O. & Nussinov, R. Similar Binding Sites and Different Partners: Implications to Shared Proteins in Cellular Pathways. *Structure* **15**, 341–354 (2007).
32. Shanahan, H. P. & Thornton, J. M. An examination of the conservation of surface patch polarity for proteins. *Bioinformatics* **20**, 2197–2204 (2004).
33. Kraulis, P. J. MOLSCRIPT: A Program to Produce Both Detailed and Schematic Plots of Protein Structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
34. Merritt, E. A. & Bacon, D. J. Raster3D: Photorealistic Molecular Graphics. *Meth. Enzymol.* **277**, 505–524 (1997).