Domain distribution and intrinsic disorder in hubs in the human protein–protein interaction network

Ashwini Patil,[1]* Kengo Kinoshita,[2,3] and Haruki Nakamura[4]

[1]Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokane-dai, Minato-ku, Tokyo 108-8639, Japan

[2]Graduate School of Information Sciences, Tohoku University, 6-3-09 Aramaki-aza-aoba, Aoba-ku, Miyagi, 982-0036, Japan

[3]Bioinformatics Research and Development, Japan Science and Technology Corporation, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan

[4]Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan

Abstract: Intrinsic disorder and distributed surface charge have been previously identified as some of the characteristics that differentiate hubs (proteins with a large number of interactions) from non-hubs in protein–protein interaction networks. In this study, we investigated the differences in the quantity, diversity, and functional nature of Pfam domains, and their relationship with intrinsic disorder, in hubs and non-hubs. We found that proteins with a more diverse domain composition were over-represented in hubs when compared with non-hubs, with the number of interactions in hubs increasing with domain diversity. Conversely, the fraction of intrinsic disorder in hubs decreased with increasing number of ordered domains. The difference in the levels of disorder was more prominent in hubs and non-hubs with fewer domains. Functional analysis showed that hubs were enriched in kinase and adaptor domains acting primarily in signal transduction and transcription regulation, whereas non-hubs had more DNA-binding domains and were involved in catalytic activity. Consistent with the differences in the functional nature of their domains, hubs with two or more domains were more likely to connect distinct functional modules in the interaction network when compared with single domain hubs. We conclude that the availability of greater number and diversity of ordered domains, in addition to the tendency to have promiscuous domains, differentiates hubs from non-hubs and provides an additional means of achieving interaction promiscuity. Further, hubs with fewer domains use greater levels of intrinsic disorder to facilitate interaction promiscuity with the prevalence of disorder decreasing with increasing number of ordered domains.

Keywords: protein–protein interactions; interaction networks; hubs; intrinsic disorder; domain enrichment

Introduction

Proteins execute their functions in the cell primarily through interactions with other proteins. A large number of interactions between proteins have been determined through several large scale, or high-throughput, experiments enabling the formation of protein–protein interaction networks and facilitating their systems level study.[1] Proteins in interaction networks have been broadly classified into two categories: those with a large number of interactions, or

hubs, and those with a few interactions, or non-hubs. As a result of their ability to interact with multiple proteins, hubs play an important role in the functioning of the cell.[2] Examples include several tumor suppressors and cell-signaling proteins. Therefore, there has been a special interest in the study of the structural and functional properties of hubs that differentiate them from non-hubs.[3–5]

Previous studies have suggested structural flexibility through large disordered regions,[6–9] and distributed surface charge, especially in small hubs (less than 300 residues) with no disordered regions.[9,10] The presence of multiple domains in hubs has also been suggested,[11,12] the reasoning being that multiple domains correspond to multiple interaction interfaces[13] and greater functional complexity.[14] Indeed, several hubs have multiple domains containing several binding sites.[5] However, some hubs contain a single domain hosting a single interface[5] or multiple overlapping interfaces.[15] This indicates that hubs may acquire interaction promiscuity with the help of either a single promiscuous domain or multiple domains. The preference for a single or multidomain architecture in hubs has been previously studied using small interaction datasets. Ekman *et al.*[7] investigated the prevalence of multiple domains, especially repeating domains, as one of the several properties differentiating hubs and non-hubs. In another study, Taylor *et al.*[16] looked at the propensity and functional nature of multiple domains in different categories of hubs having high or low levels of average gene coexpression with their interaction partners. However, the differences in domain diversity between hubs and non-hubs, or the independent effects of modular or ordered domains and intrinsic disorder on the binding ability in hubs have yet to be investigated. Thus, a comprehensive analysis of the prevalence of distinct and ordered domains in hubs and non-hubs, in relation to disordered regions and functional properties of the domains, is necessary to understand their potential role in interaction promiscuity.

With these goals in mind, we compare the Pfam domain[17] distribution of all domains, distinct domains and ordered domains in hubs and non-hubs using a large high confidence dataset of human protein–protein interactions. We study the impact of the domain diversity and modularity on the binding ability of hubs and attempt to separate the effects of modularity and intrinsic disorder on interaction promiscuity. We further perform an analysis of the functional nature of the domains and functional annotations enriched in hubs and non-hubs. We discuss the biological significance of our findings in the context of the topological properties of hubs and the role they play in the interaction network.

Results and Discussion

Although hubs in protein–protein interaction networks have been defined using several criteria, the definition of hubs as proteins with five or more interactions has been shown to be robust.[2,18] Hence, we used this criterion to identify hubs in our dataset. Further, we defined non-hubs as proteins with one interaction and ignored those with two to four interactions to minimize the effect of potential hubs with unknown interactions. Using these criteria, 4312 hubs and 1929 non-hubs were identified from a high confidence human protein–protein interaction dataset, including direct binary interactions and those derived from protein complexes to obtain better coverage (see Methods for details). Though these criteria result in a greater number of hubs than non-hubs, we confirmed that the underlying protein–protein interaction network had a scale-free topology, and that an alternative definition or exclusion of interactions derived from protein complexes did not affect the nature of the results (Refer Supporting Information for details). The proteins were annotated with Pfam-A domains and Gene Ontology (GO) terms. Disordered regions of 30 residues or more were predicted in these proteins as discussed in the Methods. Pfam domain and GO term enrichment analysis and topological analysis were performed to clarify the biological significance of the results.

Domain counts

For each protein in the interaction network, we defined three types of Pfam domain counts:

1. Total domains: The total number of Pfam domains in the protein, including those overlapping with disordered regions. We used the total domain count to include all possible binding sites in the protein. This count includes duplicate domains.
2. Distinct domains: The number of nonredundant or distinct domains in the proteins. It is a good measure of the number of distinct interactions that a protein can perform.
3. Ordered domains: The total number of Pfam domains that do not contain any predicted disordered regions.

Proteins were categorized as single domain (those with one ordered domain) and multidomain (those with two or more ordered domains) for further comparison of characteristics.

Total domain architecture

The total number of Pfam domains in a protein may be used to define the potential number of its nonoverlapping binding interfaces. A comparison of the domain content in hubs and non-hubs showed that a higher percentage of hubs had two or more domains than non-hubs (Fig. 1, $P = 1.28e\text{-}13$). This is in agreement with earlier results obtained by Ekman *et al.*[7]

We investigated the prevalence of repeating domains in hubs and non-hubs. A total of 25.8%
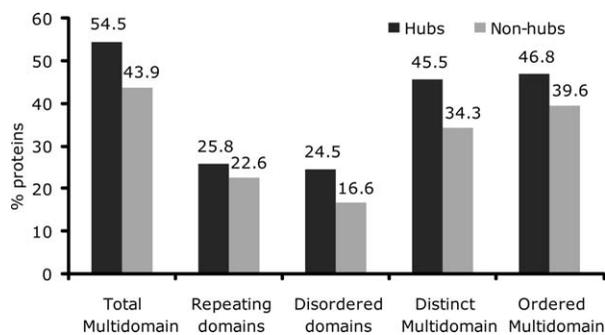
Figure 1. Percent of hubs and non-hubs with the characteristics listed on the *x*-axis: (1) Total Multidomain: % proteins with two or more Pfam domains ($P = 1.28e\text{-}13$), (2) Repeating domains: % proteins with at least one repeating domain ($P = 0.01$), (3) Disordered domains: % proteins with a predicted disordered region overlapping at least one Pfam domain ($P = 4.17e\text{-}13$), (4) Distinct Multidomain: % proteins with two or more unique Pfam domains (excluding domain duplicates, $P < 2.2e\text{-}16$), (5) Ordered Multidomain: % proteins with two or more ordered Pfam domains (excluding those with overlapping disordered regions, $P = 3.03e\text{-}05$). All *P*-values indicate the significance of the difference in the prevalence of the characteristic between hubs and non-hubs.

hubs contained at least one repeating domain when compared with 22.6% in non-hubs (Fig. 1, $P = 0.01$). Thus, more hubs had repeating domains than non-hubs. It has been previously suggested that repeating domains in hubs provide more binding interfaces resulting in greater functional diversity in hubs.[7,19] Further, among hubs with repeating domains, 65% had multiple repeating domains when compared with 57% among non-hubs ($P = 0.007$). This suggests that more than one type of domains are frequently simultaneously repeated in hubs than non-hubs.

Further, an alignment of the Pfam domains with predicted disordered regions showed that 24.5% of hubs had at least one Pfam domain overlapping with a predicted disordered region (Fig. 1). A lower percentage (16.6%) of non-hubs had an overlap between a domain and a disordered region, as expected because of the lower levels of intrinsic disorder in non-hubs ($P = 4.2e\text{-}13$).

The results above illustrate that the presence of repeating domains obscures the role of domain diversity in hubs. Similarly, in several proteins, Pfam domains overlap with disordered regions making it difficult to identify their individual prevalence and effects. This prompted us to investigate the differences in the ordered and distinct domains in hubs and non-hubs.

Domain diversity and interaction promiscuity

To determine if hubs have greater domain diversity than non-hubs, we compared the distinct domain count by excluding the repeating domains. Figure 1 shows that a larger number of hubs had more than

1 distinct domain when compared with non-hubs ($P < 2.2e\text{-}16$). The maximum number of distinct domains found in hubs was 10 as opposed to 7 in non-hubs. From these results, we conclude that hubs are more likely to have multiple distinct domains when compared with non-hubs.

Further investigation showed that the distinct domain count in hubs is positively correlated with the number of interactions (Fig. 2(A), $r = 0.124$, $P < 2.2e\text{-}16$). Thus, multiple distinct domains lead to distinct binding sites potentially providing hubs with the ability to interact with several different proteins. A similar relationship was not observed between the number of interactions and the total number of domains ($r = 0.034$) or the number of ordered domains ($r = 0.026$) in hubs. This lack of correlation probably reflects the fact that the repeat domains included in these domain counts participate in a single interaction or multiple interactions of the same type that are counted as a single interaction in the interaction network.

Domain modularity and intrinsic disorder

We compared the ordered domain content in hubs and non-hubs by eliminating domains containing disordered residues to determine the propensity of modular binding sites outside the disordered regions. A greater fraction of hubs were found to have multiple ordered domains than non-hubs (Fig. 1, $P < 2.2e\text{-}16$). We conclude that hubs are more likely to have multiple ordered domains than non-hubs, suggesting an additional means of interaction promiscuity.

Similar to previous findings, hubs had a higher percentage of disordered residues than non-hubs in this expanded human dataset (Table I, $P < 2.2e\text{-}16$). There was no correlation between the number of interactions and the fraction of intrinsic disorder in hubs. As expected from the higher levels of disorder and a greater tendency for multidomain architectures, hubs were longer than non-hubs (Table I, $P = 7.12e\text{-}10$).

As both disordered regions and multiple domains aid in interaction promiscuity, we hypothesized that hubs with few or no disordered residues would have a greater number of ordered domains to aid their interaction ability. We found that the ordered domain count in hubs was negatively correlated with the percentage of disordered residues, with disorder decreasing with increasing number of ordered domains [Fig. 2(B), $r = -0.183$, $P < 2.2e\text{-}16$]. We conclude that there is a complementary relationship between ordered domains and intrinsic disorder in hubs. This suggests that hubs with fewer domains need greater levels of intrinsic disorder to achieve interaction promiscuity, whereas this task is accomplished by a greater number of domains in multidomain proteins.

The disordered regions in 85% of the hubs with a single ordered domain were located flanking the
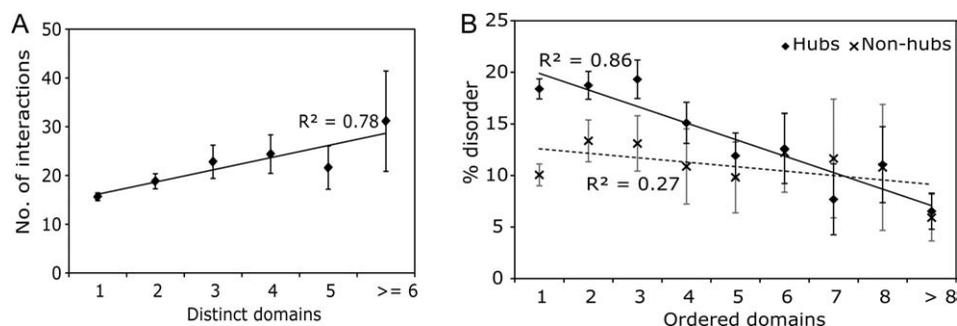
Figure 2. (A) Average number of interactions in hubs with respect to their distinct domain count. (B) Average percentage of disordered residues in hubs and non-hubs with respect to their ordered domain count. Error bars indicate the 95% confidence intervals.

domain at the N- and/or C-terminal regions suggesting that flexible ends in single domain hubs play an important role in forming multiple interactions by providing flexibility and potential binding sites. This is particularly seen in single domain kinases with large disordered regions, where the disordered regions act as the binding domains, whereas the ordered regions act as the functional domain or catalytic domain.[20] Similarly, 84% of the hubs with multiple (two or more) ordered domains contained their disordered regions either between these domains, acting as flexible linkers, or at the N- and/or C- terminal regions. Replication Protein A is a classic example of a hub with a flexible disordered linker between a protein binding domain and two DNA binding domains.[21] In the remaining smaller fraction of hubs (15% of ordered single domain, 16% of ordered multi domain), the disordered regions overlapped with other existing domains in the proteins. Thus in most hubs, the presence of disordered regions outside the ordered domains may result in interaction promiscuity through the increased relative flexibility between the domains or an increase in the number of binding sites located within the disordered regions.

No correlation was observed between the level of intrinsic disorder in hubs and their total domain count ($r = -0.068$) or their distinct domain count ($r = -0.054$). The absence of this correlation may be attributed to the fact that both these domain counts do not exclude domains with overlapping disordered regions. The correlation between intrinsic disorder and the number of ordered domains was also poor in

non-hubs [Fig. 2(B)]. But it was observed that hubs and non-hubs with fewer (1–4) domains showed a greater difference in the levels of intrinsic disorder when compared with those with more domains. This further supports the conclusion that disordered regions predominantly act in hubs with fewer domains. However, it also raises the possibility that the functional nature of domains may be the primary differentiating characteristic between hubs and non-hubs with large number of domains.

Pfam domain enrichment

We determined the differences in the functional nature of the domains enriched in hubs and non-hubs (Fig. 3, Table SI). Kinase domains were most frequent in hubs, with 405 hubs having some kind of kinase activity. Consequently, several other domains, such as SH2, SH3, PDZ, and RRM1, that regulate the catalytic activity of kinases, mediate their interactions and affect their cellular localization,[22] were also enriched in hubs [Fig. 3(A), Table SIA]. Protein kinase domains are known to be one of the most promiscuously interacting domains and have been dubbed as reusable modules (along with SH2 and SH3 domains) because of their ability to bind several targets.[23]

A further breakup of the domain frequencies in single domain and multidomain hubs highlights the differences in their domain content. Signaling domains and adaptor domains are predominant in multidomain hubs. Pkinase and Ras are primarily found in single domain hubs along with nucleic acid binding domains and those functioning in transcription regulation, such as RRM_1, HLH, Hormone_recep, and zf-C4s. Thus single domain and multidomain hubs perform different functions. Further, as single domain hubs have higher levels of intrinsic disorder, it implies that domains enriched in these hubs, that is, kinase or adaptor domains, may preferentially exist with disordered regions.

On the other hand, non-hubs were most enriched in the transcription factor domains, such as classical C2H2 type zinc fingers (zf-C2H2) and

Table I. *Average Percentage of Disordered Residues and Average Length in Hubs and Non-hubs*
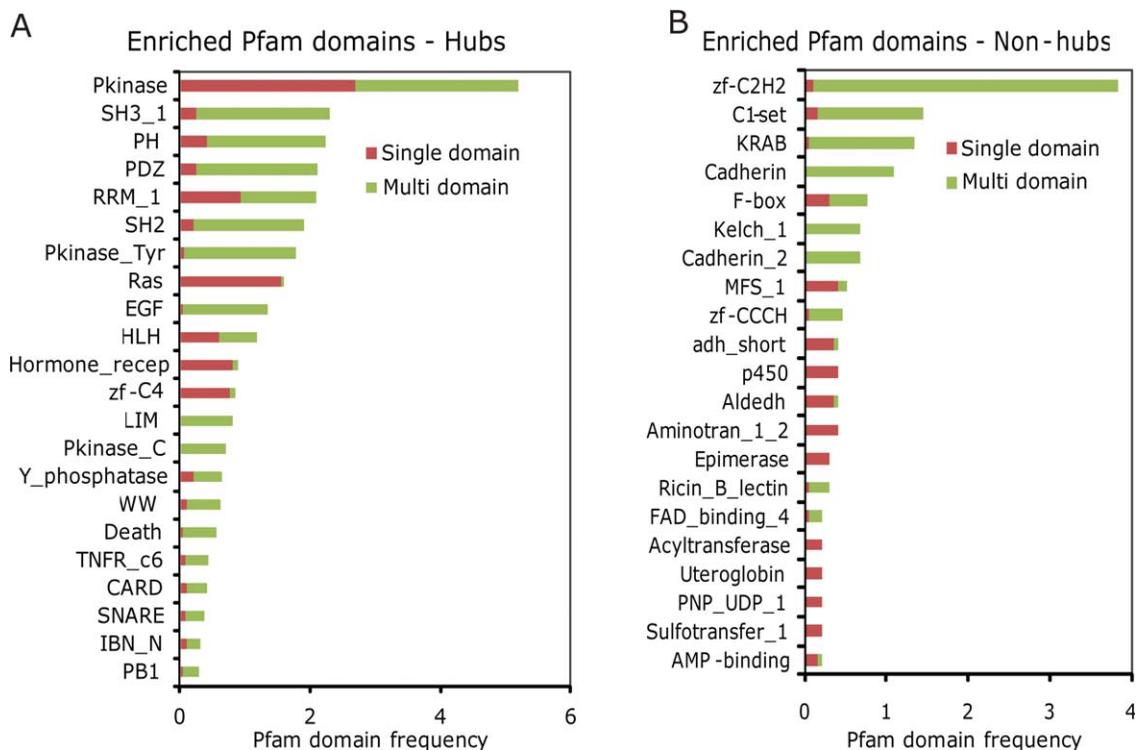
	% Disorder	Length
Non-hubs[a]	13.82 ± 0.87^c	571.79 ± 23.85
Hubs[b]	19.74 ± 0.64	660.75 ± 18.26

All differences are statistically significant ($P \ll 0.0001$).
[a] Non-hubs are proteins with one interaction.
[b] Hubs are proteins with five or more interactions.
[c] Values are 95% confidence intervals.

Figure 3. Pfam domain frequency for domains enriched in (A) hubs and (B) non-hubs ($P < 0.01$). Frequencies in single (1) domain and multi (two or more) domain hubs and non-hubs are shown in red and green, respectively.

KRAB domains [Fig. 3(B), Table SIB]. Single domain non-hubs were mostly enriched in domains, such as the transporter MFS_1 domain, p450, Aldedh, Aminotran_1_2, and Epimerase, all involved in some type of catalytic activity. Cadherin and Cadherin_2 domain, which are important in cell-cell adhesion, were primarily found in multidomain non-hubs.

We conclude that the functional nature of the domains enriched in hubs is distinct from that of non-hubs and appears to play a significant role in their promiscuity. The nature of domains in single versus multidomain hubs and non-hubs is also very distinctive and may reflect the differences in their roles in the interaction network in the case of hubs. Table SII shows the top 10 Pfam domains that are exclusively found in hubs and non-hubs to further highlight their differences.

Gene Ontology term enrichment

The differences in the domain content were reflected in the functions of single and multidomain hubs and non-hubs, as seen from the GO term enrichment analysis. We studied the molecular function, biological process, and cellular component terms of the GO annotations. The enrichment of GO molecular function terms showed that ATP binding was predominant in hubs along with transcription factor activity and RNA binding. Zinc binding was more frequent in multidomain hubs with some propensity for ATP binding [Fig. 4(A), Table SIIIA]. Single domain hubs

were also significantly enriched for the term "unfolded protein binding" indicating that they preferentially bind other disordered proteins with their disordered regions, as has been previously suggested.[20] Confirming results from Pfam domains, the GO term catalytic activity was significantly enriched in single domain non-hubs, whereas nucleic acid binding was predominant in multidomain non-hubs [Fig. 4(B), Table SIIIB]. Table SIV shows the top 10 GO molecular function terms that were exclusively found in hubs and non-hubs. The Biological Process terms also showed a similar pattern with signal transduction and regulation of transcription being the most frequent in hubs, and oxidative reduction and mitochondrial electron transport in non-hubs (Table SV). Similar functional terms have been previously found to be enriched in proteins with high and low levels of intrinsic disorder, respectively.[8,24] The enrichment of the cellular component GO term showed no significant differences in the localizations of hubs and non-hubs. We conclude that hubs are dominant in signal transduction and transcription regulation whereas non-hubs are primarily involved in catalytic activity.

Given these clear demarcations in the annotations and domain content of hubs and non-hubs, it is possible to differentiate between the two, even in the absence of experimental interaction information. This provides a promising approach to predict the interaction ability of proteins using domain and annotation information.

A

Enriched GO Terms - Hubs

B

Enriched GO Terms - Non-hubs
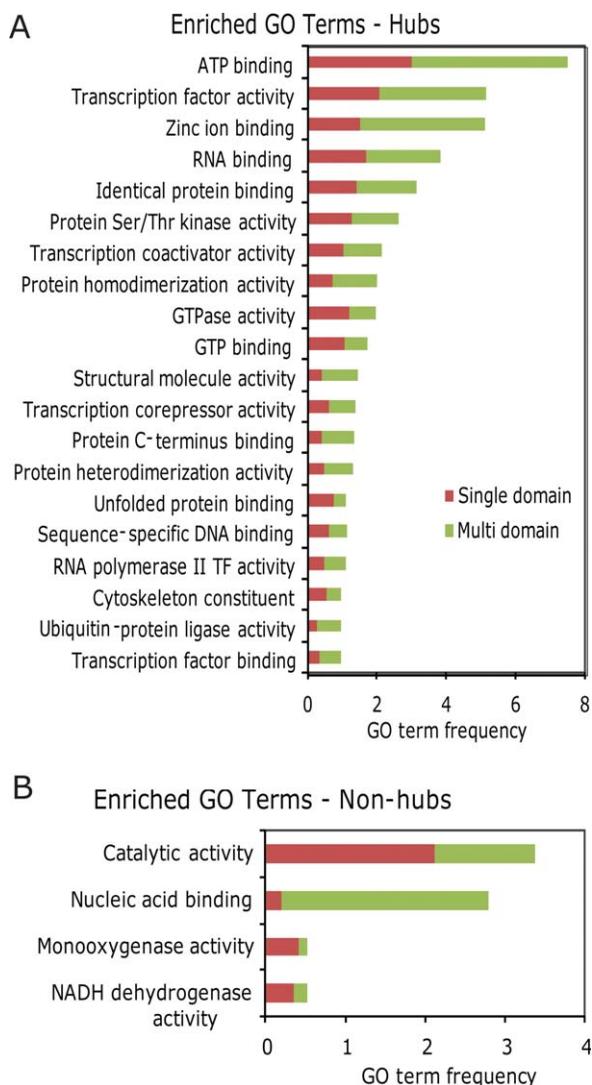
Figure 4. GO term frequency for Molecular Function terms enriched in (A) hubs (first 20 terms) and (B) non-hubs ($P <$ 0.01). Frequencies in single (1) domain and multi (two or more) domain hubs and non-hubs are shown in red and green, respectively.

Location in the interaction network

Given the differences in the characteristics of single domain and multidomain (two or more domains) hubs, we hypothesized that they would occupy different positions in the protein–protein interaction network. To find this difference, we studied two network parameters—clustering coefficient and betweenness centrality. The clustering coefficient of a node represents the interconnectivity of its interaction partners. A high clustering coefficient generally indicates that the neighborhood of the node is highly interconnected and thus the node is part of a clique or module.[25] On the other hand, a low clustering coefficient indicates that the node lies between modules. The betweenness centrality of a node gives the fraction of shortest paths in the network that pass through the node. A node with high betweenness

centrality is considered to be connecting different modules while that with low betweenness centrality is considered to be a part of a module.[26]

Single domain hubs had, on average, a lower betweenness centrality when compared with multi-domain hubs (Table II, $P = 0.0001$). This implies that multidomain hubs are more likely to connect different modules than single domain hubs. These findings are in agreement with previous results by Taylor *et al.*, who found that intermodular hubs (coepxressed with their interaction partners in specific tissues) had a higher average domain count than intramodular hubs (coepxressed with their interaction partners in all tissues).[16] Though single domain hubs had a greater clustering coefficient than multidomain hubs, the difference was not statistically significant (Table II, $P = 0.042$). The functional differences in the domains enriched in single domain and multidomain hubs, as seen in the previous section, also reflects the differences in their role in the interaction network.

Conclusion

The results of this study confirm that proteins with greater domain modularity and domain diversity are over-represented in hubs as compared to non-hubs. The functional nature of the domains enriched in hubs and non-hubs was different, with hubs being enriched in reusable and promiscuous domains. Hubs with fewer ordered domains were found to have more intrinsic disorder with the level of disorder decreasing as the number of domains increased suggesting a complementary relationship. Functional analysis revealed that hubs were primarily involved in signal transduction and transcription regulation, whereas non-hubs were predominant in catalytic activity. Further, single and multi-domain hubs were enriched in different kinds of domains and performed different roles in the interaction network with multi-domain hubs more likely to be inter-modular. Thus, along with higher levels of intrinsic disorder and surface charge, the presence of greater numbers and types of modular binding sites also differentiate hubs from non-hubs. The

Table II. *Average Betweenness Centrality and Clustering Coefficient in Single and Multidomain Hubs*

Hubs	Betweenness centrality[c]	Clustering coefficient[d]
Single domain[a]	5.72e-03 ± 7.90e-05[e]	0.155 ± 0.007
Multidomain[b]	7.66e-03 ± 1.14e-05	0.138 ± 0.006

P-values indicate the statistical significance of the difference between single and multidomain hubs.
[a] Hubs with one ordered domain.
[b] Hubs with two or more ordered domains.
[c] $P = 0.0001$.
[d] $P = 0.042$.
[e] Values are 95% confidence intervals.

Domain Distribution and Disorder in Hub Proteins

structures of hubs have evolved to facilitate their function as central participants in the interaction network and several characteristics have been acquired to fulfill this role. Disordered regions, high surface charge and presence of multiple ordered domains along with many others enable hubs to perform the required functions.

Materials and Methods

Dataset

We used the human protein–protein interaction network to study the domain distributions of hubs and non-hubs. Human protein–protein interactions were obtained from IntAct,[27] BIOGRID,[28] and Human Protein Reference Database.[29] To obtain maximum information coverage, we used 28718 interactions obtained from yeast two hybrid experiments and 18260 interactions derived from protein complexes identified using pull down, COIP, and so forth, using the spoke model (in which the bait protein is considered to interact with each of the prey proteins).[30] Derived binary interactions are known to have greater accuracy and coverage than those obtained from yeast two hybrid experiments and have minimal overlap with them.[31,32] Further, only those binary interactions, which had a high reliability score, based on a combination of genomic features, were used in the analysis to reduce the number of spurious interactions included.[32] These binary interactions were collected and annotated with Pfam-A domains and GO terms using the database Hintdb (as of March 2009).[33] This resulted in 46,978 interactions among 8954 human proteins. We defined hubs as proteins with five or more interactions and non-hubs as proteins with only one interaction, giving 4312 hubs and 1929 non-hubs. Proteins with two to four interactions were not considered to reduce the number of false non-hubs in the dataset. As we wanted to identify the characteristics of hubs and non-hubs within a single genome (in this case, human), we considered the known interactions between all proteins in the genome. Therefore, proteins with similar sequences were also considered in this analysis.

Prediction of disordered regions

Disordered regions having a length of more than 30 consecutive residues were predicted using a meta-predictor Meta-PrDOS, which uses a combination of several disorder predictors.[34] The prediction was done at a false positive rate of 5%.

Statistical significance

Statistical significance values were calculated using the Wilcoxon rank sum test for nonparametric distributions to test for a location shift greater than 0 with a significance threshold of 0.025 after correction for multiple testing.

Correlation coefficient

Correlation between two data sets was calculated using Pearson's correlation coefficient (r). Further support for correlation was obtained using the Spearman's rank correlation coefficient (ρ) (Refer Supporting Information for details). Statistical significance was calculated for all correlation coefficients.

Pfam domain and GO term enrichment

Pfam domain frequency was calculated as the fraction of the proteins having the Pfam domain. This frequency was calculated for each Pfam domain for all proteins in the dataset, and for hubs and non-hubs separately. Enrichment was calculated as the ratio of the observed number of proteins and the expected number of proteins with the Pfam domain. The hypergeometric distribution was used to calculate the statistical significance of the enrichment of Pfam domains in hubs and non-hubs as compared to the entire dataset. The frequencies of single domain and multidomain hubs and non-hubs were calculated based on the ordered Pfam domain counts in each. GO term enrichment was determined in a similar manner. The terms "protein binding," "molecular function," and "binding" were removed before performing the analysis of GO molecular function term enrichment.

Network parameters

The Betweenness Centrality and Clustering Coefficient of each protein were calculated in reference to the entire protein–protein interaction network using the Network Analyzer plugin[35] in Cytoscape.[36]

References

1. Yamada T, Bork P (2009) Evolution of biomolecular networks—lessons from metabolic and protein interactions. Nat Rev Mol Cell Biol 10:791–803.
2. Jeong H, Mason SP, Barabasi AL, Oltvai ZN (2001) Lethality and centrality in protein networks. Nature 411:41–42.
3. Han J-DJ, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJM, Cusick ME, Roth FP, Vidal M (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature 430:88–93.

4. Batada NN, Hurst LD, Tyers M (2006) Evolutionary and physiological importance of hub proteins. PLoS Comput Biol 2:e88. doi: 10.1371/journal.pcbi.0020088.

5. Kim PM, Lu LJ, Xia Y, Gerstein MB (2006) Relating three-dimensional structures to protein networks provides evolutionary insights. Science 314:1938–1941.

6. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN (2005) Flexible nets. FEBS J 272:5129–5148.

7. Ekman D, Light S, Bjorklund A, Elofsson A (2006) What properties characterize the hub proteins of the protein-protein interaction network of *Saccharomyces cerevisiae*? Genome Biol 7:R45. doi: 10.1186/gb-2006-7-6-r45.

8. Haynes C, Oldfield CJ, Ji F, Klitgord N, Cusick ME, Radivojac P, Uversky VN, Vidal M, Iakoucheva LM (2006) Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. PLoS Comput Biol 2:e100. doi: 10.1371/journal.pcbi.0020100.

9. Patil A, Nakamura H (2006) Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks. FEBS Lett 580:2041–2045.

10. Patil A, Nakamura H (2007) The role of charged surface residues in the binding ability of small hubs in protein-protein interaction networks. Biophysics 3:27–35.

11. Nobeli I, Favia AD, Thornton JM (2009) Protein promiscuity and its implications for biotechnology. Nat Biotechnol 27:157–167.

12. Tsai CJ, Ma B, Nussinov R (2009) Protein-protein interaction networks: how can a hub protein bind so many different partners? Trends Biochem Sci 34:594–600.

13. Vogel C, Bashton M, Kerrison ND, Chothia C, Teichmann SA (2004) Structure, function and evolution of multidomain proteins. Curr Opin Struct Biol 14:208–216.

14. Levitt M (2009) Nature of the protein universe. Proc Natl Acad Sci USA 106:11079–11084.

15. Humphris EL, Kortemme T (2007) Design of multispecificity in protein interfaces. PLoS Comput Biol 3:e164. doi: 10.1371/journal.pcbi.0030164.

16. Taylor IW, Linding R, Warde-Farley D, Liu Y, Pesquita C, Faria D, Bull S, Pawson T, Morris Q, Wrana JL (2009) Dynamic modularity in protein interaction networks predicts breast cancer outcome. Nat Biotechnol 27:199–204.

17. Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz H-R, Ceric G, Forslund K, Eddy SR, Sonnhammer ELL, Bateman A (2008) The Pfam protein families database. Nucl Acids Res 36:D281–D288.

18. Vallabhajosyula RR, Chakravarti D, Lutfeali S, Ray A, Raval A (2009) Identifying hubs in protein interaction networks. PLoS ONE 4:e5344. doi: 10.1371/journal.pone.0005344.

19. Keskin O, Gursoy A, Ma B, Nussinov R (2008) Principles of protein-protein interactions: what are the preferred ways for proteins to interact? Chem Rev 108:1225–1244.

20. Kim PM, Sboner A, Xia Y, Gerstein M (2008) The role of disorder in interaction networks: a structural analysis. Mol Syst Biol 4:179. doi: 10.1038/msb.2008.16.

21. Jacobs DM, Lipton AS, Isern NG, Daughdrill GW, Lowry DF, Gomes X, Wold MS (1999) Human replication protein A: global fold of the N-terminal RPA-70 domain reveals a basic cleft and flexible C-terminal linker. J Biomol NMR 14:321–331.

22. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S (2002) The protein kinase complement of the human genome. Science 298:1912–1934.

23. Schuster-Bockler B, Bateman A (2007) Reuse of structural domain-domain interactions in protein networks. BMC Bioinformatics 8:259. doi: 10.1186/1471-2105-8-259.

24. Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Uversky VN, Obradovic Z (2007) Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. J Proteome Res 6:1882–1898.

25. Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nat Rev Genet 5:101–113.

26. Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M (2007) The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. PLoS Comput Biol 3:e59. doi: 10.1371/journal.pcbi.0030059.

27. Kerrien S, Alam-Faruque Y, Aranda B, Bancarz I, Bridge A, Derow C, Dimmer E, Feuermann M, Friedrichsen A, Huntley R, Kohler C, Khadake J, Leroy C, Liban A, Lieftink C, Montecchi-Palazzi L, Orchard S, Risse J, Robbe K, Roechert B, Thorneycroft D, Zhang Y, Apweiler R, Hermjakob H (2007) IntAct—open source resource for molecular interaction data. Nucl Acids Res 35:D561–D565.

28. Breitkreutz B-J, Stark C, Reguly T, Boucher L, Breitkreutz A, Livstone M, Oughtred R, Lackner DH, Bahler J, Wood V, Dolinski K, Tyers M (2008) The BioGRID Interaction Database: 2008 update. Nucl Acids Res 36:D637–D640.

29. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A, Balakrishnan L, Marimuthu A, Banerjee S, Somanathan DS, Sebastian A, Rani S, Ray S, Harrys Kishore CJ, Kanth S, Ahmed M, Kashyap MK, Mohmood R, Ramachandra YL, Krishna V, Rahiman BA, Mohan S, Ranganathan P, Ramabadran S, Chaerkady R, Pandey A (2009) Human Protein Reference Database—2009 update. Nucl Acids Res 37:D767–D772.

30. Bader GD, Hogue CW (2002) Analyzing yeast protein-protein interaction data obtained from different sources. Nat Biotechnol 20:991–997.

31. von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, Bork P (2002) Comparative assessment of large-scale data sets of protein-protein interactions. Nature 417:399–403.

32. Patil A, Nakamura H (2005) Filtering high-throughput protein-protein interaction data using a combination of genomic features. BMC Bioinformatics 6:100. doi: 10.1186/1471-2105-6-100.

33. Patil A, Nakamura H (2005) HINT: a database of annotated protein-protein interactions and their homologs. Biophysics 1:21–24.

34. Ishida T, Kinoshita K (2008) Prediction of disordered regions in proteins based on the meta approach. Bioinformatics 24:1344–1348.

35. Assenov Y, Ramirez F, Schelhorn S-E, Lengauer T, Albrecht M (2008) Computing topological parameters of biological networks. Bioinformatics 24:282–284.

36. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504.