Disorder and Sequence Repeats in Hub Proteins and Their Implications for Network Evolution

ARTICLE *in* JOURNAL OF PROTEOME RESEARCH · DECEMBER 2006

Impact Factor: 4.25 · DOI: 10.1021/pr060171o · Source: PubMed

5 AUTHORS, INCLUDING:

Zsuzsanna Dosztányi
Eötvös Loránd University
41 PUBLICATIONS **3,287** CITATIONS
SEE PROFILE

Jake Chen
Indiana University-Purdue University Indiana…
138 PUBLICATIONS **1,671** CITATIONS
SEE PROFILE

István Simon
Hungarian Academy of Sciences
136 PUBLICATIONS **8,616** CITATIONS
SEE PROFILE

Peter Tompa
Vlaams Instituut voor Biotechnologie
130 PUBLICATIONS **8,502** CITATIONS
SEE PROFILE

research articles Journal of proteome research

Disorder and Sequence Repeats in Hub Proteins and Their Implications for Network Evolution

Zsuzsanna Dosztányi,[†] Jake Chen,[‡,§] A. Keith Dunker,[||] István Simon,[†] and Peter Tompa*[,†]

Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, 1518 Budapest, Hungary, School of Informatics, Indiana University, Indianapolis, Indiana 46202, Department of Computer and Information Science, Purdue School of Science, Indianapolis, Indiana 46202, and Center for Computational Biology and Bioinformatics, School of Medicine, Indiana University, Indianapolis, Indiana 46202

Received April 14, 2006

Protein interaction networks display approximate scale-free topology, in which hub proteins that interact with a large number of other proteins determine the overall organization of the network. In this study, we aim to determine whether hubs are distinguishable from other networked proteins by specific sequence features. Proteins of different connectednesses were compared in the interaction networks of *Saccharomyces cerevisiae*, *Drosophila melanogaster*, *Caenorhabditis elegans*, and *Homo sapiens* with respect to the distribution of predicted structural disorder, sequence repeats, low complexity regions, and chain length. Highly connected proteins ("hub proteins") contained significantly more of, and greater proportion of, these sequence features and tended to be longer overall as compared to less connected proteins. These sequence features provide two different functional means for realizing multiple interactions: (1) extended interaction surface and (2) flexibility and adaptability, providing a mechanism for the same region to bind distinct partners. Our view contradicts the prevailing view that scaling in protein interactomes arose from gene duplication and preferential attachment of equivalent proteins. We propose an alternative evolutionary network specialization process, in which certain components of the protein interactome improved their fitness for binding by becoming longer or accruing regions of disorder and/or internal repeats and have therefore become specialized in network organization.

Keywords: disordered protein • unstructured protein • protein−protein interaction • interaction network • hub protein

Introduction

Protein function at the cellular level has a significant contextual component determined by the multitude of interactions among proteins in the living cell. Therefore, a primary focus of post-genomic molecular biology has been to catalog and interpret all the interactions of the proteome. As a result of large-scale proteomic efforts, significant progress has been made. Interaction network data have been generated for several organisms, including *Saccharomyces cerevisiae* (*S. cerevisiae*),[1] *Drosophila melanogaster* (*D. melanogaster*),[2] *Caenorhabditis elegans* (*C. elegans*),[3] and *Homo sapiens* (*H. sapiens*).[4] Although these distinct interactomes differ in many details, all seem to display a scale-free topology, or at least an approximation thereof,[5] characterized by a power-law distribution of the degree of connectivity.[6−8] In such networks, most proteins (the network nodes) are connected to relatively fewer, highly connected proteins (the hubs). These hub proteins play essential roles in organizing the network. The presence of hubs explains the salient features of biological networks such as robustness, because random removal of nodes is much better tolerated in a scale-free network than in a random network. This robustness resulting from the scale-free topology is of prime importance in cell survival.[8]

Because protein interactomes share the scale-free topology with many other networks in nature, it has been suggested that their emergence has been governed by the same underlying principles, i.e., steady and random growth and preferential attachment to already highly connected nodes.[6] While this random-growth (gene-duplication) model agrees with a variety of considerations and observations,[9,10] it oversimplifies the biology of protein interactions. In particular, this model does not consider differences among proteins nor the potential of proteins to adapt their interaction capacity to their specialized functions through molecular evolution. In fact, recently it has been formally demonstrated that scale-free topology in protein interactomes can arise from varying fitness values of nodes[11] and can be explained by simple genetic events, without assuming a selective pressure on network topology itself.[12] Furthermore, it has been suggested that the large interaction capacity of hubs might be directly manifested in discernible

* Corresponding author. Telephone: (361) 279-3143. Fax: (361) 466-5465. E-mail: tompa@enzim.hu.
[†] Hungarian Academy of Sciences.
[‡] School of Informatics, Indiana University.
[§] Purdue University.
[||] School of Medicine, Indiana University.

10.1021/pr060171o CCC: $33.50

Published on Web 10/06/2006

Table 1. Selected Properties of the Databases[a]

	YEAST	YEAST_CORE	WORM	FLY	FLY_CONF	HUMAN
Interactions[a]	10741	6600	3992	20433	4733	25207
proteins[b]	4358	2640	2616	7003	4646	7560
proteins in IC_1[c]	1578	793	1527	2282	2569	1997
sequences in genome[d]	6357	6357	19957	18484	18484	32035
no. of HUBS (cutoff)[e]	766 (5)	141 (5)	282 (5)	1910 (5)	213 (5)	2329 (5)
no. of HUBS (cutoff)[f]	398 (9)	228 (12)	229 (6)	651 (15)	360 (4)	721 (14)
maximum interaction[g]	288	111	187	175	40	188

[a] The table shows the total number of interactions (a), the total number interacting proteins (b), the number of proteins with one interaction (c), the number of sequences in the genomes (d), the number of HUB proteins (the cutoff value used to define HUB proteins) with fixed cutoff (e), the number of HUB proteins (the cutoff value used to define HUB proteins) with floating cutoff (f), and the maximum interaction of a single protein (g) for the various datasets used.

physicochemical feature(s).[9] Our goal in this study is to explore whether hub proteins are enriched in features such as intrinsic structural disorder and sequence repeats.

Recently, it has become clear that a large fraction of eukaryotic proteins lack a well-defined 3D structure, but manifest their functions in an intrinsically unstructured or disordered state.[13−19] Computational studies have shown that this feature increases with the increasing complexity of the organisms and prevails in regulatory and signal-transduction proteins.[20,21] This structural state confers important functional features, such as increased interaction capacity,[22−24] enhanced association rates,[25−27] and adaptability to different partners.[28,29] Apparently, these features are of potential benefit in functions realized in protein interaction networks, as suggested previously.[30,31] In fact, disorder has been noted to contribute to hub characteristics in several ways, where hubs may be mostly disordered, partially disordered, or ordered (i.e. highly structured), but interacting with disordered partners.[31] Disordered proteins or segments are often generated by the expansion of internal repeat regions[32] and often concomitantly exhibit low-complexity amino acid compositions.[33,34] Internal repeat regions can also encode for recurring structural elements in ordered proteins, whose presence could lead to generation of novel proteins or functional variants.[35] Whether disordered or ordered, sequence repeats might afford proteins enhanced evolutionary prospects due to an enlarged available surface area, which predisposes them for functioning via protein−protein interactions.[35] Since many disordered regions do not contain the salient sequence features of low complexity or repeats,[36] and low complexity does not represent an absolute discriminator for order and disorder,[36] disorder prediction is needed to indicate the lack of specific 3D structure in the absence of ligands and partners.[16−19] In fact, because of the significant difference in the various attributes of sequences encoding for disordered and ordered proteins, prediction of intrinsic disorder in proteins can play a crucial role in the comparison and analysis of different proteomes.[21,37]

Motivated by these implications, we have used bioinformatics methods to predict protein disorder, sequence repeats, and segments of low complexity in the interaction networks of *S. cerevisiae* (YEAST), *D. melanogaster* (FLY), *C. elegans* (WORM), and *H. sapiens* (HUMAN). In this study, we found that hub proteins tend to be larger and contain significantly longer and more frequent regions of these sequence features, which indicate such features as the structural basis for the large interaction capacity of hub proteins. Our extensive global analysis and findings provide strong evidence supporting the generalization of recent observations regarding the importance of disorder for a few hub proteins[31] and also for a highly restricted interactome dataset.[38] In more general terms, our observations challenge the simplistic view that the evolution of scale-free behavior by random growth leads to preferential attachment.[6] We offer a more dynamic and realistic evolutionary perspective, in which network specialization is primarily accomplished via evolution of hub proteins through the accrual of "hub-friendly" sequence features.

Materials and Methods

Sequence Features. Proteins were analyzed for four sequence features: length, regions of predicted protein disorder, low complexity, and sequence repeats. Sequences were downloaded from GenBank and were studied using the programs available via the Internet as follows: protein disorder, IUPred[39,40] available at http://iupred.enzim.hu/; low complexity, SEG[33] downloaded from the ftp site ftp://ftp.ncbi.nlm.nih.gov/pub/seg/; and sequence repeats, internal repeat finder[41] available at http://nihserver.mbi.ucla.edu/Repeats/. Disorder and low complexity could be assigned to residues, from which the percentage of the sequence covered could be calculated. In the case of repeats, however, the program returns only an estimation of the repeat number and repeat length, from which percentage coverage, instead of the actual location of the repeat region, could be calculated. Various combinations of these properties, calculated for each protein as the maximum percentage of the properties, were also analyzed.

Datasets. The list of all protein−protein interactions of yeast (YEAST) was downloaded from the BIND database (2004 November release). A "core dataset" of the interactome is also defined as the subset of interactions observed by several different large- or small-scale experiments, or confirmed by studies on paralogues. This YEAST_CORE dataset was downloaded from the DIP database.[42] The *C. elegans* interactions (WORM) were also taken from BIND.[3] The Drosophila interactome (FLY) was downloaded from the CuraGen database at http://www.curagen.com/.[2] The interaction file also specifies a reliability score for each interaction. The subset of the "confident interactions" (FLY_CONF) was also separately analyzed. The data for human interacting proteins (HUMAN) were downloaded from human protein reference database http://www.hprd.org/.[4] The major features of the datasets are shown in Table 1.

Hub proteins (HUBs) were defined in two different ways. On one hand, we applied a fixed cutoff, when proteins with five or more interactions were defined as candidate hubs, similarily to a previous work.[38] On the other hand, because hub function is a system property and it cannot be appropriately defined at the level of individual proteins, we also applied a floating cutoff definition, in which a unique cutoff was set for each interactome depending on the dataset. In this case, hub proteins were

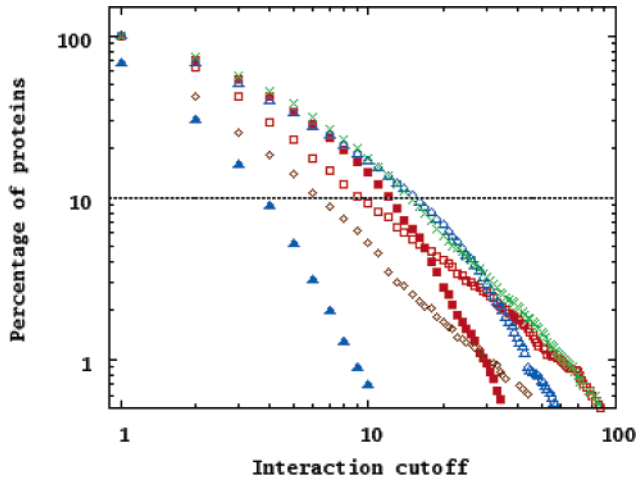

Figure 1. Degree of connectivity in the interaction databases. The percentage of proteins with interacting partners above the given cutoff value is shown for YEAST (red squares), FLY (blue triangles), WORM (brown diamonds), and HUMAN (green x) databases on a log−log scale. The filled symbols correspond to confident/core interactions. A power-law distribution assumes a linear relationship.

defined as the top 10% of proteins with the highest number of interacting partners (cf. Table 1, Figure 1). The properties of hub proteins were compared to two reference datasets. IC_1 contains proteins with exactly one interaction. The other reference dataset was RAN, a random sample of genome sequences. To generate RAN, the appropriate genome sequences were downloaded from the COGENT database at http://cgg.ebi.ac.uk/services/cogent/info.html. Due to computational considerations, a maximum of 10 000 proteins, representative of the whole genomes, were randomly selected and their sequence features were individually determined. We chose to compare hub proteins to these reference datasets primarily because of the above-mentioned uncertainty in hub definition. Due to this, the dataset complementary to hubs at any chosen cutoff value will also contain proteins that behave as hubs themselves, which will compromise the statistical difference between hubs and nonhubs. Comparing two distinct reference datasets, RAN, which statistically represents the whole interactome and thus contains hubs, and IC_1, which by definition does not, are expected to provide the information necessary to resolve this dilemma.

Statistical Approaches. For each sequence, the total number or proportion of residues with a given feature was determined. Beside the average of these values, their distributions were also analyzed, to obtain a more detailed picture. The range of the values was divided into five bins so that roughly equal portions of RAN sequences fell into each bin. Since occasionally more than one-fifth of proteins lack disorder/low complexity/repeats, the first bin can cover more than 20% of proteins, whereby the remaining proteins could then be divided into four equal bins. Nevertheless, these unequal bins with similar number of data points give statistically more reliable results than bins of equal width, where the first bin would contain most data points and the last bin only a few.

The distribution of properties was compared between HUB, IC_1, and RAN proteins (for definition and explanation, see Datasets). The statistical significance of the differences was also assessed. The basic assumption was that deviations arise because of limited sample size. This was estimated by scooping into the reference dataset, selecting as many random proteins as HUB contained, and repeating this procedure 5000 times. These random samples were then used to calculate the standard deviation of the percentage of proteins falling into the given bin. The deviation was also approximated with the square root of the actual data points in the given bin. This estimation gave very similar results (not shown).

Given the distribution of the percentage of hub proteins and the reference proteins (IC_1 or RAN) in each bin and their deviations, the likelihood that they derive from the same distribution can also be estimated using χ^2 statistics: $\chi^2 = \Sigma(a_i - b_i)^2/(2 \times \mathrm{dev}_i^2)$. For each i bin, a_i and b_i is the percentage of proteins falling in the given bin for HUB proteins and for proteins in the reference set, respectively, and dev_i is the standard deviation of the percentage of proteins in the bin, calculated from the 5000 random samples taken from the reference set, and used for both sets. By virtue of this value, we can test the hypothesis that the two distributions are not the same. If this value exceeds 13.3 (using 4 as the degrees of freedom), the two distributions are different at a confidence level of 99%.

Results

Connectivity of the Interactomes Studied. The scale-free topology of interaction networks is primarily manifested in a power-law distribution of the degree of connectivity. Different interactomes, however, have been determined by different experimental techniques and vary in coverage, which may affect their global topological features.[5] For a comparison of the actual databases used herein (Table 1), Figure 1 shows the percentage of proteins versus their numbers of interactions on a log−log scale. The YEAST and WORM data follow most closely the linear relationship expected, but the other datasets deviate significantly from a straight line, as already noted.[2,5] Despite the deviation from strict linearity, however, the data are amenable to our proposed analyses, with a relatively small fraction of proteins having large numbers of interactions.

Sequence Features in Hubs and Nonhubs, and the Effect of Cutoff Choice. Four sequence features (sequence size and the number of residues with either of the three sequence attributes: structural disorder, sequence repeats, and low complexity) have been compared between hubs and nonhubs by calculating the differences between the averages for hubs proteins (HUB) and proteins with exactly one interaction (IC_1) and a random sample of genome sequences (RAN), for all four species (Table 2). The mean and standard deviation values show that hubs are signicantly longer, and contain more of, and a greater proportion of, structural disorder, sequence repeats, and low complexity than nonhubs. Hubs in this experiment have been defined by a fixed cutoff of five or more interaction partners. These comparisons show that the length of the protein and disorder are the strongest discriminators of hubs from reference proteins, with repeats and low complexity displaying smaller, but still significant, differences. Worth noting is the unusual behavior of the WORM database, where IC_1 proteins also seem to be biased. A further point is that these data allow an insight into the evolution of hub function. Comparing the four species, hubs appear to have gained mostly in length and disorder, with a smaller incerase in the other two features. Thus, these four features not only characterize hubs today, they also have contributed to the evolution of hub function and the increase in complexity of protein interactomes.

Table 2. Number of Residues and Proportion with Sequence Features of Hubs in Four Interactomes with Fixed Cutoff[a]

			IC_1		RAN	
datasets	property	HUB average	average	SD	average	SD
			Average Number of Residues			
YEAST	disorder	143.69	87.38	5.09	92.11	5.45
YEAST	lc	45.20	32.86	1.79	34.35	1.93
YEAST	repeat	23.44	14.52	2.43	17.84	2.94
YEAST	length	532.61	430.23	12.01	469.72	13.61
WORM	disorder	140.34	144.09	12.82	77.94	9.54
WORM	lc	49.28	47.66	3.90	33.62	3.43
WORM	repeat	49.20	61.20	14.14	36.65	10.39
WORM	length	494.90	506.24	26.00	436.73	22.87
FLY	disorder	174.55	144.07	5.37	163.40	6.30
FLY	lc	68.90	57.95	2.20	65.65	2.55
FLY	repeat	43.22	38.43	3.35	42.75	3.58
FLY	length	482.66	507.85	10.06	541.08	11.45
HUMAN	disorder	203.34	168.29	5.91	119.14	4.22
HUMAN	lc	70.22	59.76	1.70	43.76	1.42
HUMAN	repeat	94.51	72.45	5.60	56.48	3.73
HUMAN	length	698.33	621.82	11.40	467.36	8.99
			Proportion of Residues			
YEAST	disorder	0.2460	0.1532	0.0760	0.1730	0.0079
YEAST	lc	0.0818	0.0713	0.0034	0.0767	0.0036
YEAST	repeat	0.0331	0.0287	0.0041	0.0302	0.0040
WORM	disorder	0.2493	0.2124	0.0149	0.1636	0.0136
WORM	lc	0.1006	0.0906	0.0077	0.0808	0.0069
WORM	repeat	0.0873	0.0660	0.0099	0.0588	0.0095
FLY	disorder	0.3148	0.2280	0.0059	0.2594	0.0063
FLY	lc	0.1381	0.1010	0.0030	0.1140	0.0031
FLY	repeat	0.0875	0.0568	0.0034	0.0633	0.0034
HUMAN	disorder	0.2722	0.2292	0.0050	0.2389	0.0054
HUMAN	lc	0.1027	0.0999	0.0023	0.0955	0.0024
HUMAN	repeat	0.0967	0.0785	0.0039	0.0864	0.0042

[a] For each species (*S. cerevisiae*, *D. melanogaster*, *C. elegans*, and *H. sapiens*) and each sequence feature (disorder, low complexity (lc), repeats, and total length in the case of absolute numbers) the distribution of the number of residues or the proportion of the given feature in the sequences of hub proteins is compared to that of both the genome sequences (RAN) and the IC_1 (proteins with exactly one interaction). HUB proteins were defined as proteins with at least five interactions. The averages were calculated for HUB proteins, for the RAN and IC_1 datasets. The SD refers to the standard deviation calculated from the average properties over random samples of proteins, matching the number of HUB proteins but selected from RAN and IC_1 datasets, respectively (see Materials and Methods).

The literature offers no clue regarding the specification of the boundary between hubs and nonhubs; indeed, the concept of being a hub is qualitative rather than quantitative. Since the choice of the cutoff could affect the above results, an experiment was carried out to examine this uncertainty. Figure 2 shows the results of how the mean of the difference between hubs and random sample of proteins changes with the cutoff value. The mean for hub proteins is higher than the average for RAN and tends to increase with increasing cutoff values up to very large values with the exception of HUMAN, which shows a diminution of the difference with high values. The reason for this deviation is not apparent but may be related to the distinction that most of the data in the human interactome[4] have come from curated individual observations, whereas the other three datasets have been generated from high-throughput studies. Of course, as the boundary value increases, the number of hub proteins decreases, which increases the standard deviation and impairs the estimation of the significance of the difference. Nevertheless, these data support the overall conclusion that hubs have discernible structural characteristics that do not depend on hub definition, i.e., on the choice of the cutoff value.

Characterizing Hubs by Applying a Floating Cutoff Definition. Here we further address the point that hub function is a system property and cannot be exactly defined at the level of the individual proteins. Furthermore, various interactomes have been determined by different techniques, which are of different sensitivity and may not provide the same information even on the very same protein. We therefore also adopted an alternative definition using a floating cutoff definition in which hub proteins were considered as the top 10% of proteins sorted according to the number of interaction partners. This defines hubs in terms of their relation to the entire interactome and thereby increases the statistical power of our analysis.

For the four species, the occurrence of the four attributes in hubs has been compared to that in IC_1 and RAN, by dividing the range of values obtained into five bins so that roughly an equal portion of RAN sequences fell into each bin (for details, see Statistical Approaches). The results are shown in Figure 3. In practically all the cases, hub proteins are underrepresented in the first or first two bins and are overrepresented in the last or last two bins. In other words, hubs tend to have more disorder, more sequence repeats, or more low complexity regions than nonhubs. The trend is similar for the sizes of proteins: the polypeptide chains of hubs are significantly longer than those of nonhubs, with the possible exception of Drosophila, in which case there is an excess of hubs with medium lengths (centered at around 400 and 500 amino acids).

A concise description of all these comparisons is given in Table 3. Here, the difference between the distributions of sequence features and the total length of hubs and both random genome samples and proteins with exactly one interaction is determined. The difference is characterized by a χ^2 value and the corresponding probability that the two sets of data are of different distributions. Hubs and the reference datasets significantly differ in practically all features for all the

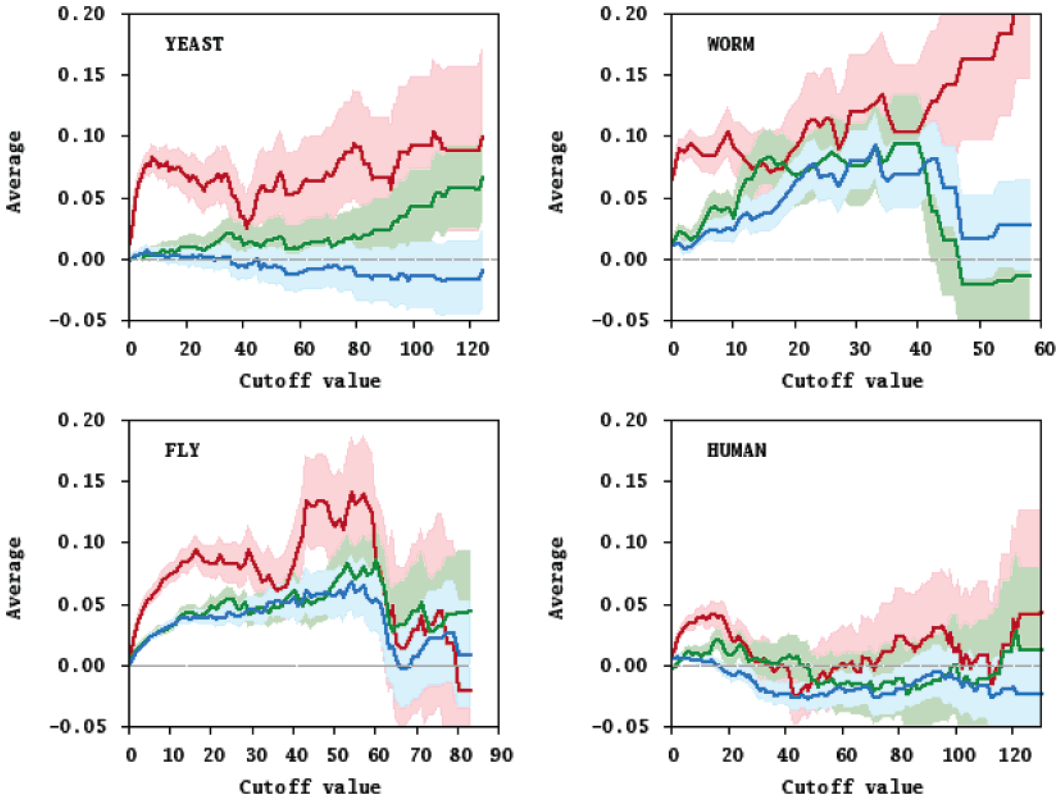

Figure 2. Effect of hub definition on the difference between hubs and random proteins. Three sequence features, disorder (red), repeat coverage (green), and low complexity (blue) are compared for YEAST, WORM, FLY, and HUMAN hubs (HUB) and a random selection of proteins (RAN). The difference between the average values for HUB and RAN is shown as a function of the number of interactions, above which a protein is considered a hub (cutoff). The light-colored stripe around the mean corresponds to the standard deviation.

species, with the possible exception of the low complexity in WORM and length in YEAST_CORE. It is also to be noted that differences are more significant in some cases with RAN than IC_1, which suggests IC_1 as a class is not representative of the whole genome but is biased due to the methods used for high-throughput screening of protein−protein interactomes. In general, there are some minor differences between the datasets. Disorder is a strong distinguishing feature in the various interactomes, with length and repeats being also convincing in most cases. In several instances low complexity appears to be the least obvious but is still significant. Apart from these minor differences, however, it is safe to conclude that all three features are important, and thus conserved, in defining hub behavior.

Proportion of Various Sequence Features in Hubs. In addition to the overall length of regions with a particular feature, it is also worth investigating how the proportion of regions with a given feature differs between hubs and nonhubs. Regions with a given feature were identified and normalized by the size of the protein for the four species, as shown in Figure 4. Again, the difference in favor of hubs is significant in most cases, with the exception of low complexity in YEAST_CORE, as quantitatively rendered in Table 4. Compared to the total length of features, there are some differences in the order of the importance of features, which are probably of secondary importance. Overall, these data strongly suggest that not only lengthy regions of disorder/repeats/low complexity but also a high proportion of these features is likely to be important for conferring advantages in terms of hub behavior.

Interdependence of Sequence Features. Because the three features studied are not independent of each other, it is important to determine the extent of their correlation. Although intrinsically unstructured or disordered proteins are often composed of repeats[32] and low sequence complexity correlates with the lack of a well-defined structure,[36] the three features are not perfectly correlated characteristics. Indeed their combinations are stronger indicators of hubs than any single one. To characterize their interdependence, the difference of the averages of individual features and their combinations between HUB and RAN proteins has been determined (Figure 5). Among the three properties, disorder exhibited the largest increase in all species, whereas their combination increased the difference even further. In contrast, low complexity in any combination led to only a minor increase and showed a relatively high correlation with both disorder and repeats (data not shown).

All Interactions versus Confident Interactions. A comparison of interactomes obtained in different studies has shown that individual studies may have provided a low coverage of the total interactome and contain a significant fraction of false positive interactions.[4,43] This suggests that the actual interaction databases may not be representative, which may cause artificial results in our studies. To minimize this possibility, we characterized specific subsets covering reliable interactions only. These are available for the Drosophila (FLY_CONF) and yeast (YEAST_CORE) interactome (for definitions, see Datasets).

Restricting our analysis to the subset of confident interactions did not alter the differences in any significant way (Tables 3 and 4), thus corroborating the prior major conclusions. The differences of the averages are significant in most of the cases by the measures χ^2 and probability of difference in the distributions of hubs and reference datasets.

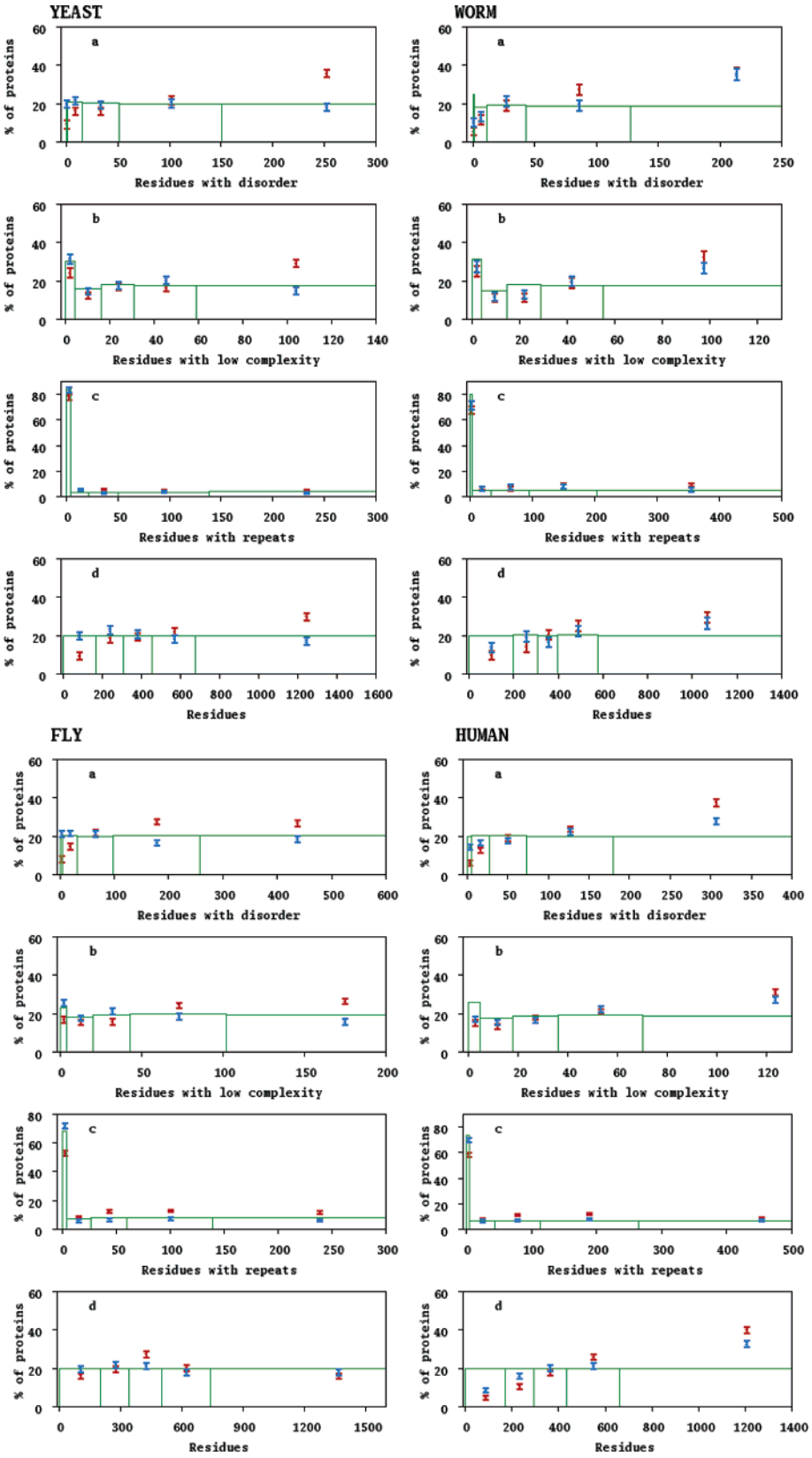

Figure 3. Total length of sequence features of hubs in interactomes. Comparison of the total number of residues with disorder (a), low complexity (b), repeats (c), and the length of protein (d) for the yeast (YEAST), worm (WORM), Drosophila (FLY), and human (HUMAN) interactomes. Green columns represent the borders of bins that contain about the same number of proteins calculated for the random sample of genome sequences (RAN). Hub proteins are shown in red, and the random sample of proteins with exactly one interaction (IC_1) are shown in blue. The height of the columns and the position of symbols represent the percent of proteins in the given database that fall into the given bin, i.e., range of feature. The horizontal position of symbols is arbitrary, because it represents all data within the given bin. The error bars correspond to the standard deviation calculated from IC_1.

Table 3. Number of Residues with Sequence Features of Hubs in Four Interactomes with Floating Cutoff[a]

		compared to RAN		compared to IC_1	
datasets	property	χ^2	probability	χ^2	probability
YEAST	disorder	49.14	>0.999 99	58.81	>0.999 99
YEAST	lc	23.94	0.999 91	38.87	>0.999 99
YEAST	repeat	16.05	0.997 04	13.27	0.989 99
YEAST	length	27.53	0.999 98	42.27	>0.999 99
YEAST_CORE	disorder	24.89	0.999 94	8.79	0.933 46
YEAST_CORE	lc	13.43	0.990 64	4.74	0.685 45
YEAST_CORE	repeat	58.93	>0.999 99	36.29	>0.999 99
YEAST_CORE	length	21.65	0.999 76	3.36	0.500 88
WORM	disorder	53.13	>0.999 99	8.24	0.916 94
WORM	lc	25.53	0.999 96	2.71	0.392 13
WORM	repeat	20.43	0.999 58	4.31	0.633 74
WORM	length	17.55	0.998 48	5.45	0.755 59
FLY	disorder	53.67	>0.999 99	84.06	>0.999 99
FLY	lc	27	0.999 98	54.14	>0.999 99
FLY	repeat	63.67	>0.999 99	106.29	>0.999 99
FLY	length	17.84	0.998 67	12.28	0.984 60
FLY_CONF	disorder	29.62	0.999 99	28.75	0.999 99
FLY_CONF	lc	3.49	0.520 86	4.57	0.665 33
FLY_CONF	repeat	15.72	0.996 58	23.30	0.999 89
FLY_CONF	length	32.9	0.999 99	24.83	0.999 95
HUMAN	disorder	125.77	>0.999 99	40.93	>0.999 99
HUMAN	lc	64.37	>0.999 99	5.66	0.774 17
HUMAN	repeat	79.14	>0.999 99	41.68	>0.999 99
HUMAN	length	170.41	>0.999 99	28.93	0.999 99

[a] For each species (*S. cerevisiae*, *D. melanogaster*, *C. elegans*, and *H. sapiens*) and each sequence feature (disorder, low complexity (lc), repeats, and total length) the distribution of the number of residues with the given feature in the sequences of hub proteins is compared to that in both the genome sequences (RAN) and the IC_1 (proteins with exactly one interaction). The difference is characterized by χ^2 values (see Materials and Methods) and the corresponding probability that the two sets of data are of different distributions.

Discussion

The interaction networks of different species show remarkable similarity in terms of the global feature of near scale-free topology. In the interactomes examined, we found significant deviations from a strict scale-free behavior (cf. Figure 1), with an increasing deficiency of hubs toward higher connectedness. In principle, this may be attributed to the suppression of hubs,[2] to an approximation of scale-free behavior due to limited sampling,[5] or to some other factors, such as the limited size of the network. Notwithstanding these reservations, the fundamental features of these networks derive primarily from the presence of highly connected hubs. Thus, the molecular basis of the function of hub proteins is key to understanding how interaction networks provide the bases for cell function. In this paper, we present evidence that hub proteins are significantly larger, have more predicted disorder, and contain more sequence repeats and/or low-complexity regions than nonhubs. Our studies into the interdependence of the features have shown that disordered segments and repeat regions are relatively independent, and their presence in hubs provides synergistic structural rationale for hub behavior. Low complexity, on the other hand, is much worse in distinguishing hubs from nonhubs and may actually compromise discrimination between these functional classes (cf. Figure 5). Interpretation of this sequence feature in terms of hub function, thus, may lead to misleading conclusions.

As noted in the Introduction, structural disorder confers many functional advantages, several of which provide the rationale for its prevalence in hubs,[16,18,19] as also suggested in previous works.[24,31] The observation that either the total length of disordered segments or the proportion of disorder are equally good attributes of hub function may actually imply two alternative structural strategies. For example, the open structure of intrinsically unstructured or disordered proteins provides a large interaction surface, which enhances the capacity of the protein for interactions.[23] This is of clear benefit to proteins, which interact simultaneously with many partners such as the so-called party hubs.[31,44] The presence of long repeat regions may be rationalized on a similar ground.[35] For several actual examples of hubs, such as for caldesmon, BRCA1, and estrogen receptor α, for example,[31] the advantage of a large amount of disorder has been demonstrated. Hub function may also benefit because disorder significantly increases the association rate of protein interactions,[17,25] as formulated in the fly casting[26] and protein fishing[27] models. In addition, multicomponent complexes do not assemble very well from rigid components due to steric clashes. In contrast, the use of coupled folding and binding by flexible subunits facilitates the formation of such multicomponent complexes by avoiding the steric problems encountered by rigid subunits.[45]

A somewhat different structural logic may apply to proteins that do not necessarily have long disordered regions but that have a high proportion of disorder. These hubs may rely on the malleability of their structures, which enable them to adapt to distinct partners. Such adaptability has been described for cyclin-dependent protein kinase inhibitors,[28] for glycogen synthase kinase 3β,[46] for α-synuclein,[47] and for the hypoxia inducible factor 1α,[18] and has been generalized as moonlighting[29] or polymorphism in the bound state.[24] A variation on this theme might be represented by ordered hubs, such as calmodulin,[31] or 14-3-3 proteins,[48] for which the partners utilize disorder in an adaptative process. While such a use of disorder by hub partners definitely occurs in some cases, we found no statistical difference between hub partners and other proteins (data not shown). To further investigate whether disorder may be important for some hub partners, it will be necessary to separately evaluate the partners of a large collection of highly ordered hubs. Overall, the excess of the features studied in hubs can be rationalized in terms of the functional specialization of these proteins. It would be interesting to test additionally whether the features are also correlated with the connectedness of hubs. Due to the relatively low number of hubs and the extremely wide range of the number of interaction partners (Table 1), the significance of this possible correlation could not be established given the current limited dataset (data not shown).

In addition, the excess of disorder and repeat regions in hubs also has general evolutionary implications. Hubs play a central role in defining the scale-free topology of interaction networks, but the prevailing model for the emergence of such a contextual arrangement fails to capture the basic capacity of proteins to undergo evolutionary changes. The most influential model of network evolution assumes that random growth and preferential attachment to already highly connected nodes explains the emergence of scale-free behavior.[6,7] The underlying evolutionary mechanism has been assumed to be gene duplication, which, due to mere chance, prefers nodes already connected to nodes with multiple links.[9,10] Although scale-free topology in principle may confer several selective advantages, such as error tolerance,[8] avoidance of jamming,[49] and hierarchical modularity,[44] upon which selective pressure may act, its suggested development oversimplifies the situation in which deletion of gene products, rewiring of physical contacts, and

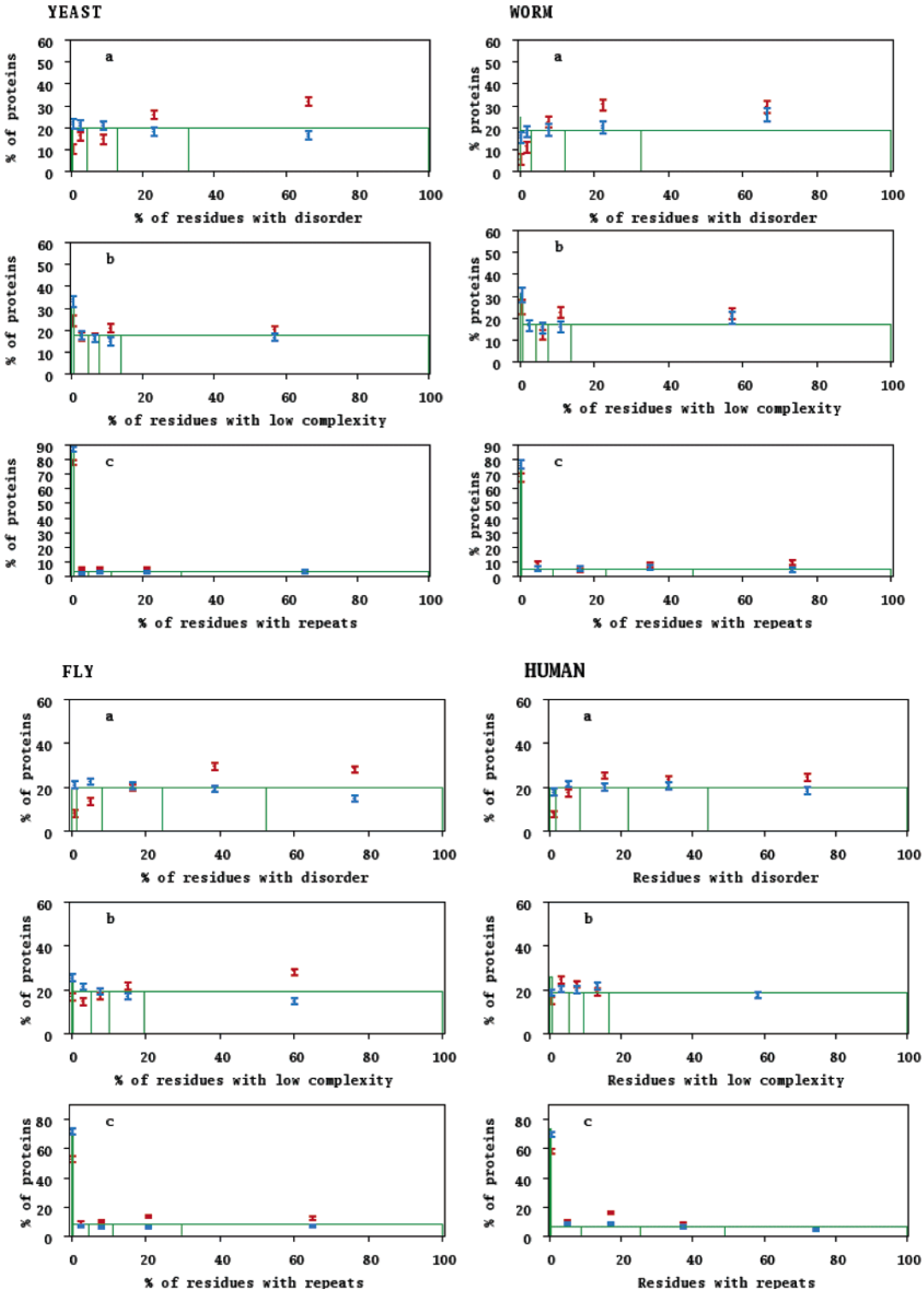

Figure 4. Proportion of disorder, sequence repeats, and low complexity in hubs in interactomes. Comparison of the relative abundance of disorder (a), repeats (b), and low complexity (c) for the yeast (YEAST), worm (WORM), Drosophila (FLY), and human (HUMAN) interactomes. Hub proteins (red), a random sample of proteins with exactly one interaction (blue), and a random sample of genome sequences (green) are shown as in Figure 3.

critical differences between individual proteins all need to be taken into account.[12] In fact, it has been formally shown that this topology also arises simply if hub proteins attract novel partners due to their physicochemical nature that predisposes them for interactions.[11] It is of relevance here that IUPred, the algorithm used for assessing disorder,[39,40] relies on estimating the energy content that a given protein segment can realize. This incorporates energy terms for both intramolecular and protein−solvent interactions. The importance of the latter in both network evolution and intrinsic disorder has been discussed recently.[50] A recent analysis of genetic regulatory networks has in fact shown that, for the evolution of a network with the observed global and local features, elements of both node copying (gene duplication) and link mutation (change in

Table 4. Proportion of Disorder, Sequence Repeats, and Low Complexity in Four Interactomes[a]

		compared to RAN		compared to IC_1	
datasets	property	χ^2	probability	χ^2	probability
YEAST	disorder	38.92	>0.999 99	67.88	>0.999 99
YEAST	lc	6.23	0.817 45	14.50	0.994 14
YEAST	repeat	17.05	0.998 11	32.66	>0.999 99
YEAST_CORE	disorder	31.28	>0.999 99	19.17	0.999 27
YEAST_CORE	lc	4.68	0.678 18	3.20	0.475 50
YEAST_CORE	repeat	59.17	>0.999 99	68.92	>0.999 99
WORM	disorder	48.10	>0.999 99	20.68	0.999 63
WORM	lc	8.29	0.918 54	6.09	0.807 22
WORM	repeat	23.25	0.999 89	13.61	0.991 35
FLY	disorder	69.66	>0.999 99	115.33	>0.999 99
FLY	lc	31.99	>0.999 99	72.41	>0.999 99
FLY	repeat	65.08	>0.999 99	112.49	>0.999 99
FLY_CONF	disorder	28.84	0.999 99	30.15	>0.999 99
FLY_CONF	lc	8.47	0.924 29	12.03	0.982 86
FLY_CONF	repeat	17.70	0.998 59	22.77	0.999 86
HUMAN	disorder	47.55	>0.999 99	43.79	>0.999 99
HUMAN	lc	34.03	>0.999 99	10.01	0.959 82
HUMAN	repeat	111.64	>0.999 99	52.17	>0.999 99

[a] For each species (*S. cerevisiae*, *D. melanogaster*, *C. elegans*, and *H. sapiens*) and each sequence feature (disorder, low complexity (lc), repeats) the distribution of the number of the proportion of the given feature for hub proteins is compared to that in both the genome sequences (RAN) and the IC_1 (proteins with exactly one interaction). The difference is characterized by χ^2 values (see Materials and Methods) and the corresponding probability that the two sets of data are of different distributions.

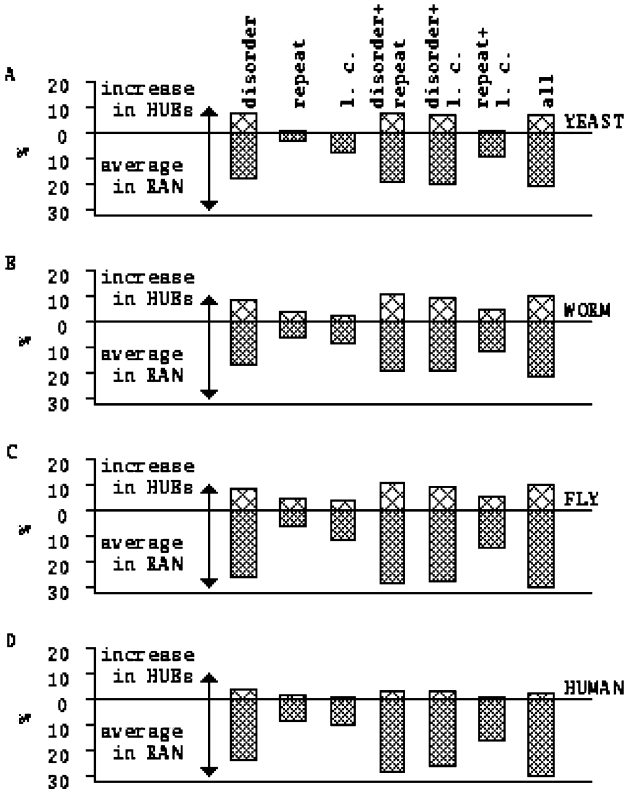

Figure 5. Averages and correlations of the sequence features. The average percentage of the sequence properties and their various combinations in HUB proteins for the four species calculated. The region below the zero line corresponds to the average in the random genome subsets (RAN), whereas the region above the zero line shows the increase in HUB proteins for the three primary sequence properties (disorder, repeats, and low complexity, lc) and their various combinations. These latter ones were defined as the maximum of the two or three properties for each protein, averaged over the dataset.

interaction) events have to be invoked.[51] Although protein−protein interaction networks studied in our work and genetic regulatory networks differ in some basic aspects, the evolutionary complexity for one regulatory network supports our point that a more elaborate evolutionary model of biological networks is likely needed in general for biological networks.

By examining hubs of the four species, it is clear their length and disorder, and to a lesser degree their repeats and low complexity regions, tend to increase as the complexity of the organism and the underlying interactomes increases in complexity on the evolutionary tree. These observations support the idea that the evolution of protein interaction networks has involved an element of selection of certain proteins toward functioning in network organization, i.e., by becoming hubs. In this process, the generation and extension of internal repeat regions and the increase in disorder is proposed to have played an active role. This scenario fully conforms to the logic of specialization that enables biological entities to occupy more niches. Additionally, direct functional advantages could also derive from this specialization process. Disordered proteins are frequently involved in regulated interaction processes, due to their disposition for posttranslational modification.[16,52] This is of significant functional benefit, as interaction networks are very dynamic objects, prone to undergo profound reorganization events mostly conducted by "transient"[53] or "date"[44] hubs. An interesting possibility is that date hubs may also draw a functional advantage not from the disorder and ensuing adaptability of their own but that of their partners. In principle, this might alleviate the demand of structural adaptability of the hub and provide a simpler solution for the inclusion of the hub in distinct and functionally/structurally unrelated complexes. This possibility has been discussed previously.[31]

Another significant feature might derive from the observation that, since intrinsically disordered proteins are typically unfolded, they undergo little change upon treatment with heat or chemical denaturants. This resistance may provide protection against elimination of hub function, to which scale-free networks are very sensitive.[8] A further pertinent point is that intrinsically unstructured or disordered regions often bind with their partner(s) by virtue of an extended surface with interaction sites dispersed over the surface of the ordered protein partner.[19] A change in any of these sites might not entirely eliminate the interaction and may thus provide resistance against point mutations. This may be a good explanation why evolutionary variability shows very weak correlation with the number of interaction partners,[54] whereas removal of a hub is three times more probable to be lethal than other proteins.[55]

In summary, hub proteins are found to be enlarged and also to be enriched in predicted disorder, in sequence repeats, and in low complexity regions and in combinations involving two or more of these features. All of these characteristics and their combinations facilitate binding to multiple partners. The enrichment of these features over evolutionary time is probably necessary to explain these observations, suggesting a more complicated evolutionary history than the commonly accepted mechanisms based on simple, random gene duplication. Experimental studies to further test these proposed roles of intrinsic disorder in protein−protein interaction networks would be useful.

Acknowledgment. This work was supported by grants from the Economic Competitiveness Operative Program (GVOP),

Grant 3.1.1.-2004-05-0195/3.0; the Hungarian Scientific Research Fund (OTKA), Grants K60694, F043609, and T049073; the National R&D Program (NKFP), Grant MediChemBats21/A/005/2004; the NIH, Grant R01 LM007688; and the Indiana Genomics Initiative, funded in part by the Lilly Endowment. We also acknowledge the Bolyai János fellowships for Z.D. and P.T. and the Wellcome Trust International Senior Research Fellowship ISRF 067595 for P.T. Finally, A.K.D. would like to thank Zoran Obradovic, Vladimir Uversky, and Pedro Romero for their continuing collaborations that include discussions of importance to the current work.

References

(1) Gavin, A. C.; Bosche, M.; Krause, R.; Grandi, P.; Marzioch, M.; Bauer, A.; Schultz, J.; Rick, J. M.; Michon, A. M.; Cruciat, C. M.; Remor, M.; Hofert, C.; Schelder, M.; Brajenovic, M.; Ruffner, H.; Merino, A.; Klein, K.; Hudak, M.; Dickson, D.; Rudi, T.; Gnau, V.; Bauch, A.; Bastuck, S.; Huhse, B.; Leutwein, C.; Heurtier, M. A.; Copley, R. R.; Edelmann, A.; Querfurth, E.; Rybin, V.; Drewes, G.; Raida, M.; Bouwmeester, T.; Bork, P.; Seraphin, B.; Kuster, B.; Neubauer, G.; Superti-Furga, G. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **2002**, *415*, 141−147.

(2) Giot, L.; Bader, J. S.; Brouwer, C.; Chaudhuri, A.; Kuang, B.; Li, Y.; Hao, Y. L.; Ooi, C. E.; Godwin, B.; Vitols, E.; Vijayadamodar, G.; Pochart, P.; Machineni, H.; Welsh, M.; Kong, Y.; Zerhusen, B.; Malcolm, R.; Varrone, Z.; Collis, A.; Minto, M.; Burgess, S.; McDaniel, L.; Stimpson, E.; Spriggs, F.; Williams, J.; Neurath, K.; Ioime, N.; Agee, M.; Voss, E.; Furtak, K.; Renzulli, R.; Aanensen, N.; Carrolla, S.; Bickelhaupt, E.; Lazovatsky, Y.; DaSilva, A.; Zhong, J.; Stanyon, C. A.; Finley, R. L., Jr.; White, K. P.; Braverman, M.; Jarvie, T.; Gold, S.; Leach, M.; Knight, J.; Shimkets, R. A.; McKenna, M. P.; Chant, J.; Rothberg, J. M. A protein interaction map of *Drosophila melanogaster*. *Science* **2003**, *302*, 1727−1736.

(3) Li, S.; Armstrong, C. M.; Bertin, N.; Ge, H.; Milstein, S.; Boxem, M.; Vidalain, P. O.; Han, J. D.; Chesneau, A.; Hao, T.; Goldberg, D. S.; Li, N.; Martinez, M.; Rual, J. F.; Lamesch, P.; Xu, L.; Tewari, M.; Wong, S. L.; Zhang, L. V.; Berriz, G. F.; Jacotot, L.; Vaglio, P.; Reboul, J.; Hirozane-Kishikawa, T.; Li, Q.; Gabel, H. W.; Elewa, A.; Baumgartner, B.; Rose, D. J.; Yu, H.; Bosak, S.; Sequerra, R.; Fraser, A.; Mango, S. E.; Saxton, W. M.; Strome, S.; Van Den Heuvel, S.; Piano, F.; Vandenhaute, J.; Sardet, C.; Gerstein, M.; Doucette-Stamm, L.; Gunsalus, K. C.; Harper, J. W.; Cusick, M. E.; Roth, F. P.; Hill, D. E.; Vidal, M. A map of the interactome network of the metazoan *C. elegans*. *Science* **2004**, *303*, 540−543.

(4) Gandhi, T. K.; Zhong, J.; Mathivanan, S.; Karthick, L.; Chandrika, K. N.; Mohan, S. S.; Sharma, S.; Pinkert, S.; Nagaraju, S.; Periaswamy, B.; Mishra, G.; Nandakumar, K.; Shen, B.; Deshpande, N.; Nayak, R.; Sarker, M.; Boeke, J. D.; Parmigiani, G.; Schultz, J.; Bader, J. S.; Pandey, A. Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. *Nat. Genet.* **2006**, *38*, 285−293.

(5) Han, J. D.; Dupuy, D.; Bertin, N.; Cusick, M. E.; Vidal, M. Effect of sampling on topology predictions of protein−protein interaction networks. *Nat. Biotechnol.* **2005**, *23*, 839−844.

(6) Barabasi, A. L.; Oltvai, Z. N. Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.* **2004**, *5*, 101−113.

(7) Babu, M. M.; Luscombe, N. M.; Aravind, L.; Gerstein, M.; Teichmann, S. A. Structure and evolution of transcriptional regulatory networks. *Curr. Opin. Struct. Biol.* **2004**, *14*, 283−291.

(8) Albert, R.; Jeong, H.; Barabasi, A. L. Error and attack tolerance of complex networks. *Nature* **2000**, *406*, 378−382.

(9) Eisenberg, E.; Levanon, E. Y. Preferential attachment in the protein network evolution. *Phys. Rev. Lett.* **2003**, *91*, 138701.1−138701.4.

(10) Pastor-Satorras, R.; Smith, E.; Sole, R. V. Evolving protein interaction networks through gene duplication. *J. Theor. Biol.* **2003**, *222*, 199−210.

(11) Caldarelli, G.; Capocci, A.; De Los Rios, P.; Munoz, M. A. Scale-free networks from varying vertex intrinsic fitness. *Phys. Rev. Lett.* **2002**, *89*, 258702.1−258702.4.

(12) Aloy, P.; Russell, R. B. Taking the mystery out of biological networks. *EMBO Rep.* **2004**, *5*, 349−350.

(13) Wright, P. E.; Dyson, H. J. Intrinsically unstructured proteins: Reassessing the protein structure−function paradigm. *J. Mol. Biol.* **1999**, *293*, 321−331.

(14) Uversky, V. N.; Gillespie, J. R.; Fink, A. L. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* **2000**, *41*, 415−427.

(15) Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z. Intrinsically disordered protein. *J. Mol. Graphics Modell.* **2001**, *19*, 26−59.

(16) Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradovic, Z. Intrinsic disorder and protein function. *Biochemistry* **2002**, *41*, 6573−6582.

(17) Tompa, P. Intrinsically unstructured proteins. *Trends Biochem. Sci.* **2002**, *27*, 527−533.

(18) Dyson, H. J.; Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197−208.

(19) Tompa, P. The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett.* **2005**, *579*, 3346−3354.

(20) Iakoucheva, L.; Brown, C.; Lawson, J.; Obradovic, Z.; Dunker, A. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J. Mol. Biol.* **2002**, *323*, 573−584.

(21) Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* **2004**, *337*, 635−645.

(22) Fuxreiter, M.; Simon, I.; Friedrich, P.; Tompa, P. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J. Mol. Biol.* **2004**, *338*, 1015−1026.

(23) Gunasekaran, K.; Tsai, C. J.; Kumar, S.; Zanuy, D.; Nussinov, R. Extended disordered proteins: Targeting function with less scaffold. *Trends Biochem. Sci.* **2003**, *28*, 81−85.

(24) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. Showing your ID: Intrinsic disorder as an ID for recognition, regulation and cell signaling. *J. Mol. Recognit.* **2005**, *18*, 343−384.

(25) Pontius, B. W. Close encounters: Why unstructured, polymeric domains can increase rates of specific macromolecular association. *Trends Biochem. Sci.* **1993**, *18*, 181−186.

(26) Shoemaker, B. A.; Portman, J. J.; Wolynes, P. G. Speeding molecular recognition by using the folding funnel: The fly casting mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8868−8873.

(27) Evans, P. R.; Owen, D. J. Endocytosis and vesicle trafficking. *Curr. Opin. Struct. Biol.* **2002**, *12*, 814−821.

(28) Kriwacki, R. W.; Hengst, L.; Tennant, L.; Reed, S. I.; Wright, P. E. Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: Conformational disorder mediates binding diversity. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 11504−11509.

(29) Tompa, P.; Szasz, C.; Buday, L. Structural disorder throws new light on moonlighting. *Trends Biochem. Sci.* **2005**, *30*, 484−489.

(30) Dunker, A. K.; Obradovic, Z. The protein trinity-linking function and disorder. *Nat. Biotechnol.* **2001**, *19*, 805−806.

(31) 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.* **2005**, *272*, 5129−5148.

(32) Tompa, P. Intrinsically unstructured proteins evolve by repeat expansion. *Bioessays* **2003**, *25*, 847−855.

(33) Wootton, J. C.; Federhen, S. Statistics of local complexity in amino acid sequences and sequence databases. *Comput. Chem. (Oxford)* **1993**, *17*, 149−163.

(34) Wootton, J. C. Sequences with "unusual" amino acid compositions. *Curr. Opin. Struct. Biol.* **1994**, *4*, 413−421.

(35) Andrade, M. A.; Perez-Iratxeta, C.; Ponting, C. P. Protein repeats: Structures, functions, and evolution. *J. Struct. Biol.* **2001**, *134*, 117−131.

(36) Romero, P.; Obradovic, Z.; Li, X.; Garner, E. C.; Brown, C. J.; Dunker, A. K. Sequence complexity of disordered protein. *Proteins* **2001**, *42*, 38−48.

(37) Oldfield, C. J.; Cheng, Y.; Cortese, M. S.; Brown, C. J.; Uversky, V. N.; Dunker, A. K. Comparing and combining predictors of mostly disordered proteins. *Biochemistry* **2005**, *44*, 1989−2000.

(38) 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.* **2006**, *580*, 2041−2045.

(39) Dosztányi, Z.; Csizmok, V.; Tompa, P.; Simon, I. The pairwise energy content estimated from amino acid composition discriminates between folded and instrinsically unstructured proteins. *J. Mol. Biol.* **2005**, *347*, 827−839.

(40) Dosztányi, Z.; Csizmok, V.; Tompa, P.; Simon, I. IUPred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **2005**, *21*, 3433−3434.

(41) Pellegrini, M.; Marcotte, E. M.; Yeates, T. O. A fast algorithm for genome-wide analysis of proteins with repeated sequences. *Proteins* **1999**, *35*, 440−446.
(42) Deane, C. M.; Salwinski, L.; Xenarios, I.; Eisenberg, D. Protein interactions: Two methods for assessment of the reliability of high throughput observations. *Mol. Cell Proteomics* **2002**, *1*, 349−356.
(43) Bader, G. D.; Hogue, C. W. Analyzing yeast protein−protein interaction data obtained from different sources. *Nat. Biotechnol.* **2002**, *20*, 991−997.
(44) Han, J. D.; Bertin, N.; Hao, T.; Goldberg, D. S.; Berriz, G. F.; Zhang, L. V.; Dupuy, D.; Walhout, A. J.; Cusick, M. E.; Roth, F. P.; Vidal, M. Evidence for dynamically organized modularity in the yeast protein−protein interaction network. *Nature* **2004**, *430*, 88−93.
(45) Namba, K. Roles of partly unfolded conformations in macromolecular self-assembly. *Genes Cells* **2001**, *6*, 1−12.
(46) Dajani, R.; Fraser, E.; Roe, S. M.; Yeo, M.; Good, V. M.; Thompson, V.; Dale, T. C.; Pearl, L. H. Structural basis for recruitment of glycogen synthase kinase 3beta to the axin-APC scaffold complex. *Embo. J.* **2003**, *22*, 494−501.
(47) Uversky, V. N. A protein-chameleon: conformational plasticity of alpha-synuclein, a disordered protein involved in neurodegenerative disorders. *J. Biomol. Struct. Dyn.* **2003**, *21*, 211−234.
(48) Bustos, D. M.; Iglesias, A. A. Intrinsic disorder is a key characteristic in partners that bind 14-3-3 proteins. *Proteins: Struct., Funct., Bioinf.* **2006**, *63*, 35−42.
(49) Toroczkai, Z.; Bassler, K. E. Network dynamics: Jamming is limited in scale-free systems. *Nature* **2004**, *428*, 716.
(50) Fernandez, A.; Scott, R.; Berry, R. S. The nonconserved wrapping of conserved protein folds reveals a trend toward increasing connectivity in proteomic networks. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 2823−2827.
(51) Louzoun, Y.; Muchnik, L.; Solomon, S. Copying nodes versus editing links: The source of the difference between genetic regulatory networks and the WWW. *Bioinformatics* **2006**, *22*, 581−588.
(52) Iakoucheva, L. M.; Radivojac, P.; Brown, C. J.; O'Connor, T. R.; Sikes, J. G.; Obradovic, Z.; Dunker, A. K. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* **2004**, *32*, 1037−1049.
(53) Luscombe, N. M.; Babu, M. M.; Yu, H.; Snyder, M.; Teichmann, S. A.; Gerstein, M. Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* **2004**, *431*, 308−312.
(54) Fraser, H. B.; Hirsh, A. E.; Steinmetz, L. M.; Scharfe, C.; Feldman, M. W. Evolutionary rate in the protein interaction network. *Science* **2002**, *296*, 750−752.
(55) Jeong, H.; Mason, S. P.; Barabasi, A. L.; Oltvai, Z. N. Lethality and centrality in protein networks. *Nature* **2001**, *411*, 41−42.

PR060171O