

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of August 19, 2010):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/296/5569/910>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/cgi/content/full/296/5569/910#related-content>

This article **cites 13 articles**, 4 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/296/5569/910#otherarticles>

This article has been **cited by** 575 article(s) on the ISI Web of Science.

This article has been **cited by** 91 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/cgi/content/full/296/5569/910#otherarticles>

This article appears in the following **subject collections**:

Cell Biology

http://www.sciencemag.org/cgi/collection/cell_biol

unspliced RNA in the cytoplasm (23, 24) is not supported by the observation that the majority of these do not accumulate significantly in a *upf3Δ* strain [suppl. fig. 3 (7)]. The expectation based on intronic small nucleolar RNA-processing phenotypes that accumulation of introns in the *dbp1Δ* mutant should be inversely related to intron size (25) seems not to hold either, most likely because of Dbr1p-independent mechanisms of intron turnover (suppl. fig. 4). We do not observe correlation between a nonconsensus 5' splice site or a U-rich region near the 5' splice site and strong dependence on Nam8p (26) for splicing in vivo (Suppl. figs. 5 and 6). We also see no correlation between the presence of a U residue upstream of the branchpoint sequence (27) or the presence of a polypyrimidine tract before or after the branchpoint and strong dependence on Mud2p (suppl. figs. 7 and 8). These data indicate that using any one intron as a reporter may cause the importance of a factor to be overemphasized or missed. Genomewide analysis allows perturbations of splicing to be evaluated on every intron at once, in effect using the entire genome as a reporter.

These studies present the first genome-wide view of splicing for any organism. The ability to distinguish differently spliced forms of RNA by using oligonucleotide microarrays opens the way for expression profiling that accounts for alternative splicing and splicing regulation in higher cells. Estimates suggest that 40 to 60% of human genes produce alternatively spliced transcripts (28, 29). In a growing number of key cases, alternatively spliced mRNAs produce proteins of distinct or even antagonistic function [e.g. (30)]. Improved expression profiling technologies must resolve changes in alternative splicing not simply by estimating exon representation [e.g. (31)], but by providing direct evidence for exon joining. The results we describe here demonstrate that oligonucleotide arrays designed to detect specific splicing products will be key to accurate parallel analysis of alternative splicing in higher organisms.

References and Notes

1. J. P. Staley, C. Guthrie, *Cell* **92**, 315 (1998).
2. D. L. Black, *Cell* **103**, 367 (2000).
3. C. A. Davis, L. Grate, M. Spingola, M. Ares Jr., *Nucleic Acids Res.* **28**, 1700 (2000).
4. M. Spingola, L. Grate, D. Haussler, M. Ares Jr., *RNA* **5**, 221 (1999).
5. J. L. DeRisi, V. R. Iyer, P. O. Brown, *Science* **278**, 680 (1997).
6. D. J. Lockhart, E. A. Winzler, *Nature* **405**, 827 (2000).
7. Supplementary on Science Online at www.sciencemag.org/cgi/content/full/296/5569/907/DC1. Microarray data are available at the Gene Expression Omnibus (GEO) at www.ncbi.nlm.nih.gov/geo/, under accession numbers GSE34 and GSE35.
8. J. Banroques, J. N. Abelson, *Mol. Cell. Biol.* **9**, 3710 (1989).
9. S. P. Bjorn, A. Soltyk, J. D. Beggs, J. D. Friesen, *Mol. Cell. Biol.* **9**, 3698 (1989).
10. Fluorescently labeled target sequence sample preparation and hybridization were performed as described (5) by using 20 μg of total RNA primed with a mixture of oligo(dT) and random hexamers (7). Ar-

rays were scanned and analyzed by using a GenePix 4000A scanner and GenePix Pro 3.0 software from Axon Instruments (Union City, CA).

11. D. Yan et al., *Mol. Cell. Biol.* **18**, 5000 (1998).
12. R. Perriman, M. Ares Jr., *Genes Dev.* **14**, 97 (2000).
13. K. B. Chapman, J. D. Boeke, *Cell* **65**, 483 (1991).
14. The splice junction (SJ) index is the ratio of the mutant/wild-type ratios derived from the normalized signals from the splice junction probe to the exon 2 probe: SJ index = SJ_{mut}/SJ_{wt} divided by $E2_{mut}/E2_{wt}$, obtained by subtracting the log₂ ratio of the exon 2 probe from the log₂ ratio of the splice junction probe. The intron accumulation (IA) index is obtained by subtracting the log₂ ratio of the exon 2 probe from the log₂ ratio of the intron probe. Because probe performance may not be directly related to absolute transcript amount, these indexes depend idiosyncratically on the sequences of the probes. We also calculated the precursor/mature (PM) index, which is obtained by subtracting the log₂ ratio of the splice junction probe from the log₂ ratio of the intron probe. This index mimics the unspliced/spliced ratio used in classical splicing studies (15).
15. C. W. Pikielny, M. Rosbash, *Cell* **41**, 119 (1985).
16. H. V. Colot, F. Stutz, M. Rosbash, *Genes Dev.* **10**, 1699 (1996).
17. P. Fortes et al., *Mol. Cell. Biol.* **19**, 6543 (1999).
18. E. C. Shen, T. Stage-Zimmermann, P. Chui, P. A. Silver, *J. Biol. Chem.* **275**, 23718 (2000).
19. M. H. Jones, D. N. Frank, C. Guthrie, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 9687 (1995).
20. X. Zhang, B. Schwer, *Nucleic Acids Res.* **25**, 2146 (1997).
21. Zhang and Schwer define the brp-to-3' ss distance

starting from 2 bases downstream of the brp adenosine to the Y of the 3' ss YAG sequence. Therefore, their distance of 12 nt corresponds to 17 nt from the actual branched A to the 3' splice site G.

22. T. A. Clark, C. W. Sugnet, M. Ares Jr., unpublished observations.
23. F. He, S. W. Peltz, J. L. Donahue, M. Rosbash, A. Jacobson, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 7034 (1993).
24. M. J. Lelivelt, M. R. Culbertson, *Mol. Cell. Biol.* **19**, 6710 (1999).
25. S. L. Ooi, D. A. Samarsky, M. J. Fournier, J. D. Boeke, *RNA* **4**, 1096 (1998).
26. O. Puig, A. Gottschalk, P. Fabrizio, B. Seraphin, *Genes Dev.* **13**, 569 (1999).
27. J. C. Rain, P. Legrain, *EMBO J.* **16**, 1759 (1997).
28. E. S. Lander et al., *Nature* **409**, 860 (2001).
29. B. Modrek, C. Lee, *Nature Genet.* **30**, 13 (2002).
30. L. H. Boise et al., *Cell* **74**, 597 (1993).
31. D. D. Shoemaker et al., *Nature* **409**, 922 (2001).
32. <http://genome-www.stanford.edu/Saccharomyces/>
33. Thanks go to H. Igel for technical assistance; and to J. DeRisi, T. Ferea, T. Furey, G. Hartzog, D. Haussler, H. Heynecker, R. Perriman, T. Powers, L. Shiue, and R. Samaha for advice. Funding was provided by a W. M. Keck Foundation Grant to the Center for Molecular Biology of RNA, by National Cancer Institute grant CA77813 to M.A. and D. Haussler, and by NIH grant GM40478 to M.A. T.C. was supported by a University of California Biotechnology training grant, and an NIH Training Grant. C.S. is a Predoctoral Fellow of the Howard Hughes Medical Institute.

27 December 2001; accepted 29 March 2002

Specificity and Stability in Topology of Protein Networks

Sergei Maslov¹ and Kim Sneppen²

Molecular networks guide the biochemistry of a living cell on multiple levels: Its metabolic and signaling pathways are shaped by the network of interacting proteins, whose production, in turn, is controlled by the genetic regulatory network. To address topological properties of these two networks, we quantified correlations between connectivities of interacting nodes and compared them to a null model of a network, in which all links were randomly rewired. We found that for both interaction and regulatory networks, links between highly connected proteins are systematically suppressed, whereas those between a highly connected and low-connected pairs of proteins are favored. This effect decreases the likelihood of cross talk between different functional modules of the cell and increases the overall robustness of a network by localizing effects of deleterious perturbations.

With the growth of experimental information about basic biochemical mechanisms of life, molecular networks operating in living cells are becoming better defined. Direct physical interactions between pairs of proteins form one such network, which serves as a backbone for functional and structural relationships among its nodes and defines pathways for the propagation of various signals such as

phosphorylation and allosteric regulation of proteins. The information about specific binding of proteins to each other has recently grown by an unprecedented amount as a result of high-throughput two-hybrid experiments (1, 2). The production and degradation of proteins participating in the interaction network is controlled by the genetic regulatory network of the cell, formed by all pairs of proteins in which the first protein directly regulates the abundance of the second. The majority of known cases of such regulation happens at the level of transcription, in which a transcription factor positively or negatively regulates the RNA transcription of the controlled protein. The large-scale structure of

¹Department of Physics, Brookhaven National Laboratory, Upton, NY 11973, USA. ²Department of Physics, Norwegian University of Science and Technology, N-7491 Trondheim, Norway.

*To whom correspondence should be addressed. E-mail: maslov@bnl.gov

both these networks is characterized by a high degree of interconnectedness, where most pairs of nodes are linked to each other by at least one path. One may wonder how such a heavily intertwined and mutually dependent dynamical system can perform multiple functional tasks and still remain stable against deleterious perturbations.

We analyzed the topological properties of interaction and transcription regulatory networks in yeast *Saccharomyces cerevisiae*, which at present is perhaps the best characterized model organism. The interaction network used in this study consists of 4549 physical interactions between 3278 yeast proteins as measured in the most comprehensive two-hybrid screen of yeast proteins (2), whereas the genetic regulatory network is formed by 1289 directed positive or negative direct transcriptional regulations within a set of 682 proteins as listed in the Yeast Protein Database (Incyte Genomics, Palo Alto, CA) (3). The protein interaction network is a representative of the broad class of scale-free networks (4–6) in which the number of nodes with a given number of neighbors (connectivity) K scales as a power law $\propto 1/K^\gamma$. In our case the histogram of connectivities can be fitted by a power law with $\gamma = 2.5 \pm 0.3$ for K ranging from 2 to about 100 (7, 8). A small part of the protein interaction network, formed by proteins known to be localized in the nucleus and to interact with at least one other nuclear protein, was visualized (Fig. 1). One remarkable feature of this graph is the abundance of highly connected proteins that are mostly connected to those with low connectivity, and thus well separated from each other.

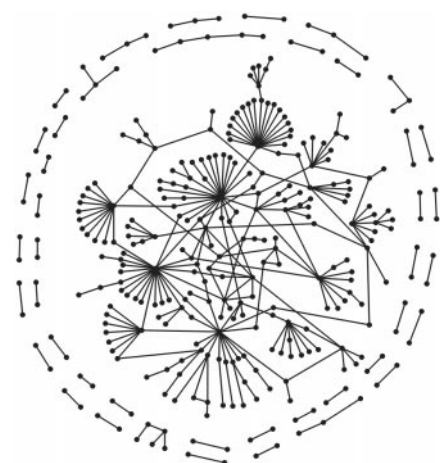


Fig. 1. Network of physical interactions between nuclear proteins. Here, we show the part of the network reported in (2), consisting of all proteins that are known to be localized in the yeast nucleus (3), and which interact with at least one other protein in the nucleus. This subset consists of 318 interactions between 329 proteins. Note that most neighbors of highly connected nodes have rather low connectivity.

To test for correlations in connectivities of nodes for each of the above two networks, we calculated the likelihood $P(K_0, K_1)$ that two proteins with connectivities K_0 and K_1 are connected to each other by a link and compared it to the same quantity $P_r(K_0, K_1)$ measured in a randomized version of the same network. In this “null model” network, all proteins have exactly the same connectivity as in the original one, whereas the choice of their interaction partners is totally random. The transcription regulatory network is naturally directed, whereas the network of physical interactions among proteins in principle lacks directionality. However, for poorly understood reasons, the two-hybrid experimental data have a significant asymmetry between baits and preys, with bait hybrids being more likely to be highly connected than their prey counterparts. This can be seen, e.g., in the fact that average connectivity of baits with at least one interaction partner is close to 3, whereas the same quantity measured for preys is only 1.8. Because each reported interaction involves one bait and one prey protein, this asymmetry needs to be taken into account when constructing an uncorrelated “null” model for the interaction network. For

this purpose, in our randomization procedure we would treat the two-hybrid data as a directed network with an arrow on each edge pointing out from bait to prey hybrid. Randomized versions of these two networks were constructed by randomly reshuffling links, while keeping the in- and out-degree of each node constant. A convenient numerical algorithm performing such randomization consists of first randomly selecting a pair of directed edges $A \rightarrow B$ and $C \rightarrow D$. The two edges are then rewired in such a way that A becomes connected to D , while C connects to B . However, in case one or both of these new links already exist in the network, this step is aborted and a new pair of edges is selected. This last restriction prevents the appearance of multiple edges connecting the same pair of nodes. A repeated application of the above rewiring step leads to a randomized version of the original network. Multiple sampling of randomized networks allowed us to calculate both the average expectation and the standard deviation for any particular property of the random network.

Correlations in connectivities manifest themselves as systematic deviations of the ratio $P(K_0, K_1)/P_r(K_0, K_1)$ from 1. We calcu-

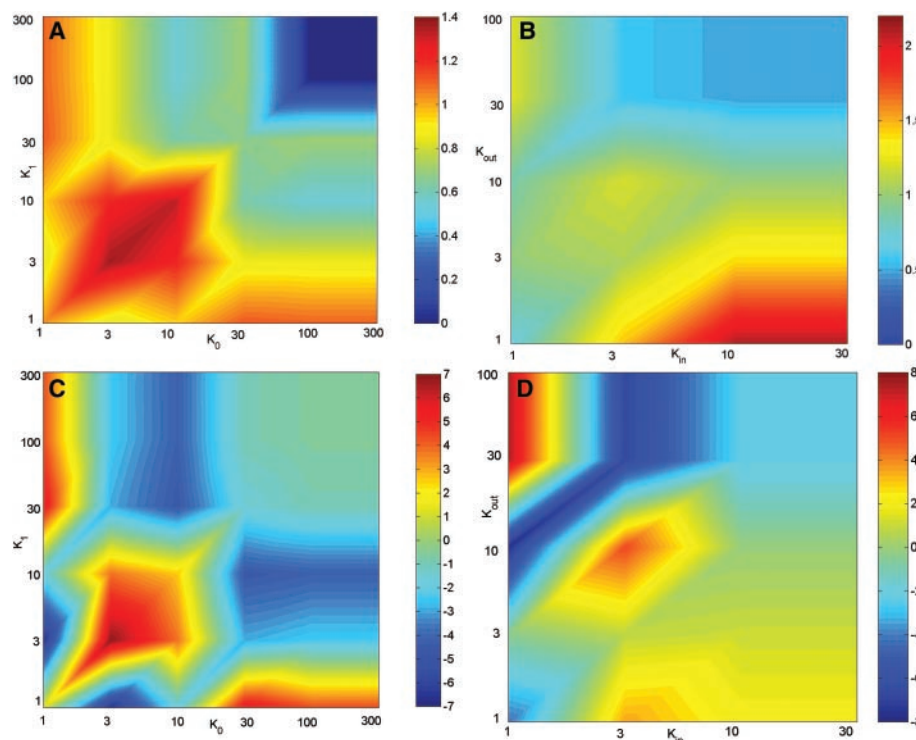


Fig. 2. Correlation profiles of protein interaction and regulatory networks in yeast. (A) The ratio $P(K_0, K_1)/P_r(K_0, K_1)$, where $P(K_0, K_1)$ is the probability that a pair of proteins with total numbers of interaction partners given by K_0, K_1 correspondingly, directly interact with each other in the full set of (2), while $P_r(K_0, K_1)$ is the same probability in a randomized version of the same network. (B) The same as in (A) but for a protein with the in-degree K_{in} to be regulated by that with the out-degree K_{out} in the transcription regulatory network (3). (C) Z-scores for connectivity correlations from (A): $Z(K_0, K_1) = [P(K_0, K_1) - P_r(K_0, K_1)]/\sigma_r(K_0, K_1)$, where $\sigma_r(K_0, K_1)$ is the standard deviation of $P_r(K_0, K_1)$ in 1000 realizations of a randomized network. (D) As in (C), but for incoming and outgoing links in the transcription regulatory network. To improve statistics, the connectivities in all four panels of this figure were logarithmically binned into 2 bins per decade.

lated this ratio for interaction (Fig. 2A) and regulatory (Fig. 2B) networks, with K_0 and K_1 being the total number of interaction partners of two interacting proteins (for the interaction network), and out- and in-degrees of two nodes connected by a directed edge $0 \rightarrow 1$ (for the regulatory network). Thus, by the very construction, $P(K_0, K_1)$ is symmetric for the physical interaction network but not for the regulatory network. We also estimated the statistical significance $Z(K_0, K_1)$ of the above deviations in the interaction (Fig. 2C) and regulatory (Fig. 2D) networks, by dividing each observed deviation from the null model by the standard deviation in multiple realizations of a randomized network. The combination of these two plots reveals the regions on the K_0 - K_1 plane, where connections between proteins in the real network are significantly enhanced or suppressed, compared to the null model. In particular, red regions in the upper left and the lower right corners reflect the tendency of highly connected nodes (hubs) to associate with nodes of low connectivity, while the blue/green region in the upper right corner reflects the reduced likelihood that two hub centers are directly linked to each other. One should also note a prominent feature on the diagonal of Fig. 2, A and C, corresponding to an enhanced affinity of proteins with between four and nine interaction partners to physically interact with each other. This feature can be tentatively attributed to the tendency of members of multiprotein complexes to interact with other proteins from the same complex. The above range of connectivities thus corresponds to a typical number of direct interaction partners of a protein in a complex. When we checked for interactions between proteins in this range of connectivities, we found 39 pairs of interacting proteins to belong to the same complex in a recent high-

throughput study (9), which is four times more than one would expect to find by pure chance alone.

To further quantify and compare correlation patterns in interaction and regulatory networks, we calculated the average connectivity $\langle K_1 \rangle$ of nearest neighbors of a node, as a function of its own connectivity K_0 (Fig. 3A). In order to simplify the comparison between two networks here, we characterize each node in the regulatory network by its total number of neighbors $K = K_{in} + K_{out}$. For both interaction and regulatory networks, the average connectivity $\langle K_1 \rangle$ shows a gradual decline with K_0 , which can be fitted with a power law $\langle K_1 \rangle \propto 1/K_0^{0.6 \pm 0.1}$ over approximately two decades. This observation gives an additional credit to the affinity between correlation patterns in these two protein networks visible in Fig. 2. It was recently found (10) that the internet, defined as the set of interconnected routers, in addition to a scale-free distribution of node connectivities similar to the protein interaction network, is characterized by the same correlation pattern between connectivities of neighboring nodes: $\langle K_1 \rangle \propto 1/K_0^{0.5}$. This extends by one step an intriguing similarity in the topology of these networks of completely different nature.

For the scale-free physical interaction network, we also plotted the probability distribution of the nearest-neighbor connectivity K_1 , measured separately for nodes with small connectivity $K_0 \leq 3$, and for those with large connectivity $K_0 \geq 100$ (Fig. 3B). In the absence of correlations, this conditional probability does not depend on K_0 and is proportional to $K_1/K_1^\gamma \sim 1/K_1^{1.5}$. This uncorrelated form holds approximately true for neighbors of a protein with low connectivity. It is only violated at the far tail of the distribution due

to an excess likelihood of it being connected to a protein with very high connectivity, as was mentioned above. On the other hand, the distribution of connectivities K_1 of neighbors of highly connected proteins scales as $\propto 1/K_1^{2.5}$ and thus differs from that of lowly connected ones by a factor of $1/K_1$.

When analyzing molecular networks, one should consider possible sources of errors in the underlying data. Two-hybrid experiments give rise to false positives of two kinds. In one case, the interaction between proteins is real but it never happens in the course of the normal life cycle of the cell, due to spatial or temporal separation of participating proteins. In another case, an indirect physical interaction is mediated by one or more unknown proteins localized in the yeast nucleus. Conversely, in a high-throughput two-hybrid screen, one should expect a sizeable number of false negatives. Primarily, a binding may not be observed if the conformation of the bait or prey heterodimer blocks relevant interaction sites or if the corresponding heterodimer altogether fails to fold properly. Secondly, 391 proteins out of the potential 5671 baits in (2) were not tested as possible bait hybrids, because they were found to activate transcription of the reporter gene in the absence of any prey proteins.

Unlike for the interaction network, our data for the genetic regulatory network do not come from a single large-scale project. Instead, they derive from a collection of numerous experiments performed with different experimental techniques in different labs. Therefore, it is not feasible even to list possible sources of errors present in such a diverse data set. In particular, one should worry about a hidden anthropomorphic factor present in such a network: Some proteins just constitute more attractive subjects of research and are, therefore, relatively better studied than others. One should also note that the transcription regulation network is only a subset of a larger genetic regulatory network, which in addition to transcriptional regulation includes translational regulation, RNA editing, and so forth. An encouraging sign was that when we separately analyzed the set representing the current knowledge (3) about this later more complete network, consisting of 1750 genetic regulations among 848 proteins, we reproduced all of our empirical results for the transcriptional network.

The observed suppression of connections between nearest neighbors of highly connected proteins is consistent with compartmentalization and modularity characteristic of control of many cellular processes (11). In fact, it suggests the picture of functional modules of the cell organized around individual hubs. To further test the extent of modularity of hubs and their immediate neighborhood in each network, we selected 15 highest connected nodes. To pro-

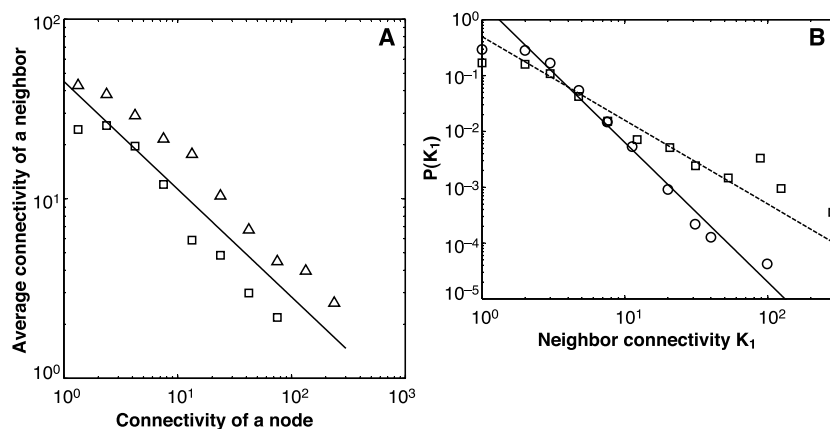


Fig. 3. Correlations in connectivities of neighbors. (A) The average connectivity $\langle K_1 \rangle$ of nearest neighbors of proteins with the connectivity K_0 in the physical interaction network (triangles) and the regulatory network (squares). The solid line is a power law fit, $\propto 1/K_0^{0.6}$. (B) The probability distribution of connectivities K_1 in the physical interaction network, calculated separately for neighbors of proteins with small connectivity $K_0 \leq 3$ (squares), and with large connectivity $K_0 \geq 100$ (circles). Lines are power laws $\propto 1/K_1^{1.5}$ (dashed) and $\propto 1/K_1^{2.5}$ (solid).

vide an unbiased sample of hubs from the point of view of in- and out-connectivity, half of those nodes were selected as the highest out-degree hubs (8 baits with $K_{\text{bait}} \geq 90$ for the interaction network and 7 nodes with $K_{\text{out}} \geq 34$ for the regulatory network), while half were the highest in-degree hubs (7 preys with $K_{\text{prey}} \geq 20$ for the interaction network and 8 nodes with $K_{\text{out}} \geq 8$ for the regulatory network). In agreement with the correlation properties described above, direct connections between hubs were significantly suppressed. In the interaction network, we observed 20 links between different hubs in this group, which is significantly below 56 ± 7.5 links in the randomized network. In the transcription regulatory network, there were 16 links between hubs in real network, as opposed to 35 ± 6.5 in its randomized version. Not only are direct links between hubs suppressed in both studied networks, but hubs also tend to share fewer of their neighbors with other hubs, thereby extending their isolation to the level of next-nearest neighbor connections. The total number of paths of length 2 between the set of 15 hubs in the interaction network is equal to 418, whereas in the null model we measured this number to be 653 ± 56 . Similarly, for the transcriptional network the number of paths of length 2 is equal to 186 in the real network, whereas from the null model one expects it to be 262 ± 30 . Since the number of paths of length 2 between a pair of proteins is equal to the number of their common interaction partners, one concludes that both the hub node itself and its immediate surroundings tend to separate from other hubs, reinforcing the picture of functional modules clustered around individual hubs.

A further implication of the observed correlation is in the suppression of the propagation of deleterious perturbations over the network. It is reasonable to assume that certain perturbations such as, e.g., significant changes in the concentration of a given protein (including its vanishing altogether in a null-mutant cell) with a ceratin probability can affect its first, second, and sometimes even more distant neighbors in the corresponding network. While the number of immediate neighbors of a node is by definition equal to its own connectivity K_0 , the average number of its second neighbors, given by $K_0 \langle (K_1 - 1) \rangle_{K_0}$, is sensitive to correlation patterns of the network. Because highly connected nodes serve as powerful amplifiers for the propagation of deleterious perturbations, it is especially important to suppress this propagation beyond their immediate neighbors. It was argued that scale-free networks in general are very vulnerable to attacks aimed at highly connected nodes (12, 13). The anticorrelation presented above implies a reduced branching ratio around these nodes and thus provides a certain degree of protection against such attacks. This may be the reason why the correlation between the connectivity of a given protein and the

lethality of the mutant cell lacking this protein is not particularly strong (8).

It is feasible that molecular networks in a living cell have organized themselves in an interaction pattern that is both robust and specific. Topologically, the specificity of different functional modules can be enhanced by limiting interactions between hubs and suppressing the average connectivity of their neighbors. We have seen that such a correlation pattern appears in a similar way in two different layers of molecular networks in yeast, and thus presumably is a universal feature of all molecular networks operating in living cells.

References and Notes

1. P. Uetz et al., *Nature* **403**, 623 (2000).
2. T. Ito et al., *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4569 (2001).
3. M. C. Costanzo et al., *Nucleic Acids Res.* **29**, 75 (2001).
4. A. L. Barabási, R. Albert, *Science* **286**, 509 (1999).
5. A. Broder et al., *Comput. Networks* **33**, 309 (2000).
6. H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, A.-L.

Barabási, *Nature* **407**, 651 (2000); D. A. Fell, A. Wagner, *Nature Biotechnol.* **18**, 1121 (2000).

7. A. Wagner, *Mol. Biol. Evol.* **18**, 1283 (2001).
8. H. Jeong, S. Mason, A.-L. Barabási, Z. N. Oltvai, *Nature* **411**, 41 (2001).
9. A.-C. Gavin et al., *Nature* **415**, 141 (2002).
10. R. Pastor-Satorras, A. Vazquez, A. Vespignani, *Phys. Rev. Lett.* **87**, 258701 (2001).
11. L. H. Hartwell, J. J. Hopfield, S. Leibler, A. W. Murray, *Nature* **402** (6761 suppl.), C47 (1999).
12. R. Albert, H. Jeong, A.-L. Barabási, *Nature* **406**, 378 (2000).
13. B. Vogelstein, D. Lane, A. J. Levine, *Nature* **408**, 307 (2000).
14. Supported in part by the NSF under grant PHY99-07949. Work at Brookhaven National Laboratory was carried out under Contract No. DE-AC02-98CH10886, Division of Material Science, U.S. Department of Energy. We thank the Institute for Theoretical Physics at the University of California at Santa Barbara, Nordita, and the University of Tokyo (for S.M.) for hospitality, K. Eriksen for valuable comments on the manuscript, T. Hwa for suggesting the subject of two-hybrid experiments, and T. Ito for providing the information on limitations of his experimental techniques.

6 August 2001; accepted 25 March 2002

Partitioning of Lipid-Modified Monomeric GFPs into Membrane Microdomains of Live Cells

David A. Zacharias,^{1,3*†} Jonathan D. Violin,^{1,2*} Alexandra C. Newton,¹ Roger Y. Tsien^{1,3‡}

Many proteins associated with the plasma membrane are known to partition into submicroscopic sphingolipid- and cholesterol-rich domains called lipid rafts, but the determinants dictating this segregation of proteins in the membrane are poorly understood. We suppressed the tendency of *Aequorea* fluorescent proteins to dimerize and targeted these variants to the plasma membrane using several different types of lipid anchors. Fluorescence resonance energy transfer measurements in living cells revealed that acyl but not prenyl modifications promote clustering in lipid rafts. Thus the nature of the lipid anchor on a protein is sufficient to determine submicroscopic localization within the plasma membrane.

Subcellular compartmentalization of signaling increases the specificity and efficiency of signal transduction. Caveolae and lipid rafts are related microdomains of the plasma membrane that are enriched in cholesterol, sphingolipids, and many signaling proteins (1, 2). Whereas protein-protein interactions maintain many signaling complexes (3, 4), specific lipid modifications are believed sufficient

to sequester proteins in lipid rafts and caveolae. In particular, acylated proteins may preferentially partition into these compartments (5, 6). Unambiguous observation of these small (<100-nm) microdomains in living cells is beyond the resolution of visible light microscopy; thus destructive assays such as cellular fractionation (7) or immunolocalization by electron microscopy (8) have been relied upon to study lipid rafts and caveolae. We now use fluorescence resonance energy transfer (FRET) between nondimerizing cyan (CFP) and yellow (YFP) variants of *Aequorea* green fluorescent protein (9, 10) to show which lipid modifications are sufficient to cause such test proteins to aggregate within lipid rafts inside living cells. FRET from CFP to YFP is advantageous, because it nondestructively detects proximities at nanometer

¹Department of Pharmacology, ²Biomedical Sciences Graduate Program, and ³Howard Hughes Medical Institute, University of California, San Diego, La Jolla, CA 92093-0647, USA.

*These authors contributed equally to this work.

†Present address: Merck Research Laboratories, 3535 General Atomics Court, MRLSDB1, San Diego, CA 92121, USA.

‡To whom correspondence should be addressed. E-mail: rtsien@ucsd.edu