From The Cover: A simple physical model for scaling in protein-protein interaction networks

Eric J. Deeds, Orr Ashenberg, and Eugene I. Shakhnovich

PNAS 2006;103;311-316; originally published online Dec 29, 2005;
doi:10.1073/pnas.0509715102

A simple physical model for scaling in protein–protein interaction networks

Eric J. Deeds*, Orr Ashenberg†, and Eugene I. Shakhnovich‡§

*Department of Molecular and Cellular Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA 02138; †Harvard College, 12 Oxford Street, Cambridge, MA 02138; and ‡Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138

Communicated by David Chandler, University of California, Berkeley, CA, November 10, 2005 (received for review September 23, 2005)

It has recently been demonstrated that many biological networks exhibit a "scale-free" topology, for which the probability of observing a node with a certain number of edges (k) follows a power law: i.e., $p(k) \sim k^{-\gamma}$. This observation has been reproduced by evolutionary models. Here we consider the network of protein–protein interactions (PPIs) and demonstrate that two published independent measurements of these interactions produce graphs that are only weakly correlated with one another despite their strikingly similar topology. We then propose a physical model based on the fundamental principle that (de)solvation is a major physical factor in PPIs. This model reproduces not only the scale-free nature of such graphs but also a number of higher-order correlations in these networks. A key support of the model is provided by the discovery of a significant correlation between the number of interactions made by a protein and the fraction of hydrophobic residues on its surface. The model presented in this paper represents a physical model for experimentally determined PPIs that comprehensively reproduces the topological features of interaction networks. These results have profound implications for understanding not only PPIs but also other types of scale-free networks.

biological networks | hydrophobic effect | scale-free networks

Many studies in recent years have revealed that a large variety of systems, from the World Wide Web to the network of chemical reactions catalyzed in a cell, exhibit a particularly interesting "scale-free" topology when represented as graphs (1–6). In these systems the probability of finding an object (or node) that connects k other nodes in the graph follows a power-law; i.e., the degree distribution [or $p(k)$] has the form $p(k) \sim k^{-\gamma}$ (1). This observation has (in general) been explained in terms of dynamical models based on the principles of network growth and an effective "preferential attachment" whereby objects that have many links at some point in time are more likely to acquire nodes as the graph grows than objects with fewer connections (1, 7). The fact that scale-free networks are so often observed in biological systems has lead to the proposal that many evolutionary processes exhibit mechanisms similar to preferential attachment that are based on the duplication and divergence of genes (4, 8–11).

One of the biological networks that has undergone considerable study is the set of interactions between proteins in the cell. The advent of high-throughput methods for measuring the binding of one protein to another using the yeast two-hybrid (Y2H) system has allowed for the characterization of large numbers of interactions between proteins in organisms such as *Saccharomyces cerevisiae*, *Helicobacter pylori*, *Caenorhabditis elegans*, and *Drosophila melanogaster* (12–16). Two major independent Y2H experiments have been performed to determine the "interactome" of *S. cerevisiae* (12, 13), and graphs of these interactions reveal that these systems constitute scale-free networks with power-law exponents ranging from ≈2.0 to ≈2.7 (1, 3, 12, 13, 17, 18).

It has long been noted, however, that Y2H screens are rather inaccurate and can lead to relatively "noisy" sets of interactions (19–22). Indeed, when the two major *S. cerevisiae* protein–protein interaction (PPI) experiments are compared with one another, one finds that only ≈150 of the thousands of interactions identified in each experiment are recovered in the other experiment (22). A similar lack of agreement has recently been found for independent Y2H experiments in *D. melanogaster* (23). Although computational methods have been proposed that may allow for some reduction of noise, it is clear that the rate of false positives and false negatives in these experiments may be quite high (19–22). Moreover, it is known that when a protein is used as bait (i.e., fused to the DNA-binding component of the Y2H), it will tend to exhibit more interactions than when used as prey (19). It is thus very clear that these experiments may contain a large number of artifacts.

In the present work we have explored these potential artifacts by considering the hypothesis that the interactions reported by the Y2H method are dominated by nonspecific interactions between proteins. This hypothesis is primarily motivated by our observation that, in general, the connectivity of a given protein is not well correlated between the Uetz *et al*. (12) and Ito *et al*. (13) experiments (see Fig. 1). We propose an entirely physical model to explain how two networks with essentially uncorrelated connectivities could nonetheless display profoundly similar (scale-free) topologies. We demonstrate that this model, when combined with an elemental source of experimental noise, reproduces the degree distributions of the experimentally determined PPI networks. The exposure of random surfaces between experiments (and thus a varying number of hydrophobic residues that thermodynamically drive interactions) is sufficient to explain the lack of correlation between two experiments that exhibit scaling in their degree distributions. We further show that "higher-order" features of these networks, such as the scaling of the clustering coefficient of a node with its connectivity (i.e., C as a function of k), are also recovered in this model. These results indicate that the observation of such topological features is not contingent on any specific evolutionary dynamics or evolutionary pressure for such networks to be "robust," "hierarchical," or "modular," as has been previously proposed. Finally, we observe a strong correlation between the hydrophobicity of a protein and its number of interacting partners, a finding that is in complete agreement with our physical model. Together these results demonstrate that the PPIs as assayed by the Y2H techniques need not report only evolved and specific interactions and that the interesting (nonrandom) topological features of these graphs need not have an evolutionary origin. Although our results do not indicate that these networks contain no evolutionarily or biologically important information, they do imply that a large number of observations in these and (perhaps) other biological networks might contain considerable influences from nonspecific interactions.

Conflict of interest statement: No conflicts declared.

§To whom correspondence should be addressed. E-mail: eugene@belok.harvard.edu.

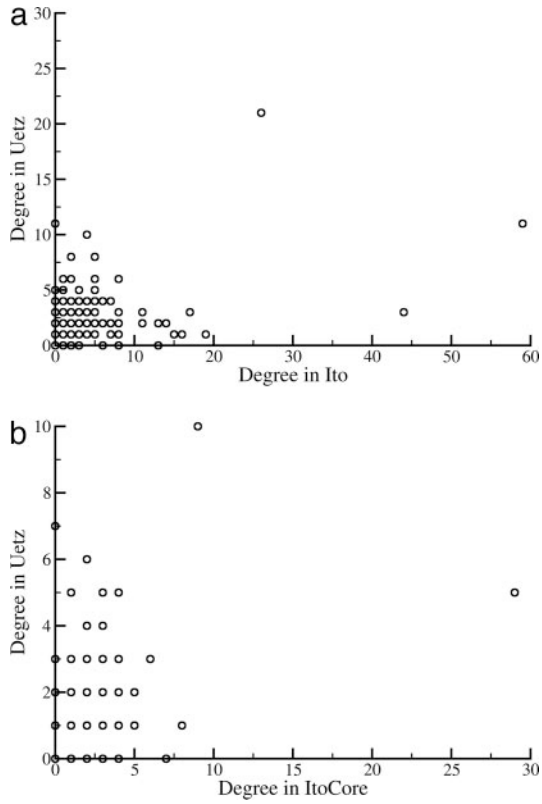

Fig. 1. Correlation between PPI networks. (*a*) The correlation between the network degree of a given protein in the Ito (13) and Uetz (12) data sets. Each point corresponds to a particular protein that exhibited interactions in both experiments. (*b*) A plot similar to *a* but comparing the ItoCore data set with Uetz.

Materials and Methods

Interaction Data. All of the interaction data used in this paper was obtained from the web site maintained by the authors of the relevant references, i.e., Uetz (12) and Ito/ItoCore (13). Interactions were obtained from these experiments and were not modified or filtered in any way.

Hydrophobicity. To determine the exposed hydrophobicity of proteins in these experiments (for use in determining the surface fraction of hydrophobic residues, p, and creating the correlations as discussed below), we employ an approximate "homology modeling" procedure whereby solvent accessibilities obtained from crystal structures are transferred to residues in the yeast proteins that align to the structurally determined homolog. This procedure is discussed in greater detail in the supporting information, which is published on the PNAS web site.

Results and Discussion

Correlations in the Number of Interacting Partners. To further explore the scale-free graphs obtained from these potentially noisy experiments we considered the graph of interactions between proteins for the 676 proteins that exhibited interactions in both the Uetz *et al.* (12) and the Ito *et al.* (13) experiments. We then compared the number of interactions measured for a given protein in one of the assays to the number of interactions for that same protein observed in the other assay. As evidenced by Fig. 1*a*, the correlation between the degree of a given protein in the two experiments is quite weak, with an R^2 of 0.18 for nodes of all degrees and an R^2 of 0.068 if the three outliers are ignored (i.e., considering only nodes of degree <20). The situation is much the same when the comparison is made with the more reliable ItoCore data set (13) (Fig. 1*b*). These very low R^2 values are striking, considering that they represent the same proteins from the same organism assayed in very similar experiments, and it is clear that these two graphs, although topologically similar, are statistically unrelated. If one set of interactions is assumed to represent the "true" set of evolved PPIs in yeast, it follows that the other graph must consist largely of experimental noise, a finding that casts doubt on the reliability of either data set. Indeed, this observation may indicate why the number of interactions made by a protein in PPI networks is only very weakly correlated with evolutionary rates (24).

The fact that these networks are scale-free, however, rules out the possibility that apparent PPIs in either case are entirely random: If they were, the graph would represent a random graph and one would observe a Poisson or Gaussian degree distribution in the resulting networks (1). To reconcile these two observations, we posit a simple physical model of PPIs. First, we assume that much of the free energy of binding that characterizes a particular PPI is due to the burial and desolvation of hydrophobic groups at the binding interface (25–28). In this case, we hypothesize that the low correlation in connectivity between the two data sets is largely due to the exposure of different surfaces for each protein in each of the Y2H experiments.

The MpK Model. Suppose there are N surface residues for a particular protein, and a given fraction p of them are hydrophobic. Say that M of those residues are actually exposed and involved in binding the other proteins in the experiment, and that K out of those M residues are hydrophobic. If we assume that M is sampled from N randomly and independently, it is clear that the probability of finding K hydrophobic residues within M follows a binomial distribution:

$$p(K) = \binom{M}{K} p^K (1-p)^{M-K}. \quad [1]$$

In this case, each PPI will result in the burial of a certain total number of hydrophobic groups; i.e., $K_{ij} = K_i + K_j$ (see Fig. 2). The desolvation of K_{ij} hydrophobic residues is related to the free energy of protein binding and represents a standard way to treat the strength of hydrophobic interactions (25–27). In this case we simply take the free energy of binding F_{ij} to be equal to $-K_{ij}$. The Y2H experiments are based on binding affinity, not binding free energy, and it follows from statistical mechanics and thermodynamics that the affinity A_{ij} between two proteins i and j will follow $A_{ij} \sim \exp(-F_{ij})$ if we set the temperature scale of our experiment such that $kT = 1$. To build a PPI network we define an experimental limit of sensitivity A_C corresponding to the weakest interaction (the interaction that buries the fewest hydrophobic groups) that is nonetheless sufficiently strong to be detected by the experiment. A_C is directly related to the number of hydrophobic residues that must be buried to observe an interaction (i.e., K_C).

To simulate this model we must first understand the distribution of p values for proteins in the experiment; therefore, we employ a simple homology modeling procedure (described in the supporting information) to transfer solvent accessibilities from proteins of known structures to their corresponding homologs from the Ito Y2H data set. We find that this distribution is well fit by a Gaussian function (see Fig. 2*b*). In our model of the Y2H experiment, we sample 3,200 values of p from a Gaussian distribution with the same mean and standard deviation (Fig. 2*b*). We use the same value of M for each protein in the experiment given that the stereotypical size of the binding surface is not determined by the surface area of the protein itself but rather the average size of the interface across all of the other proteins in the experiment. The choice of M is essentially

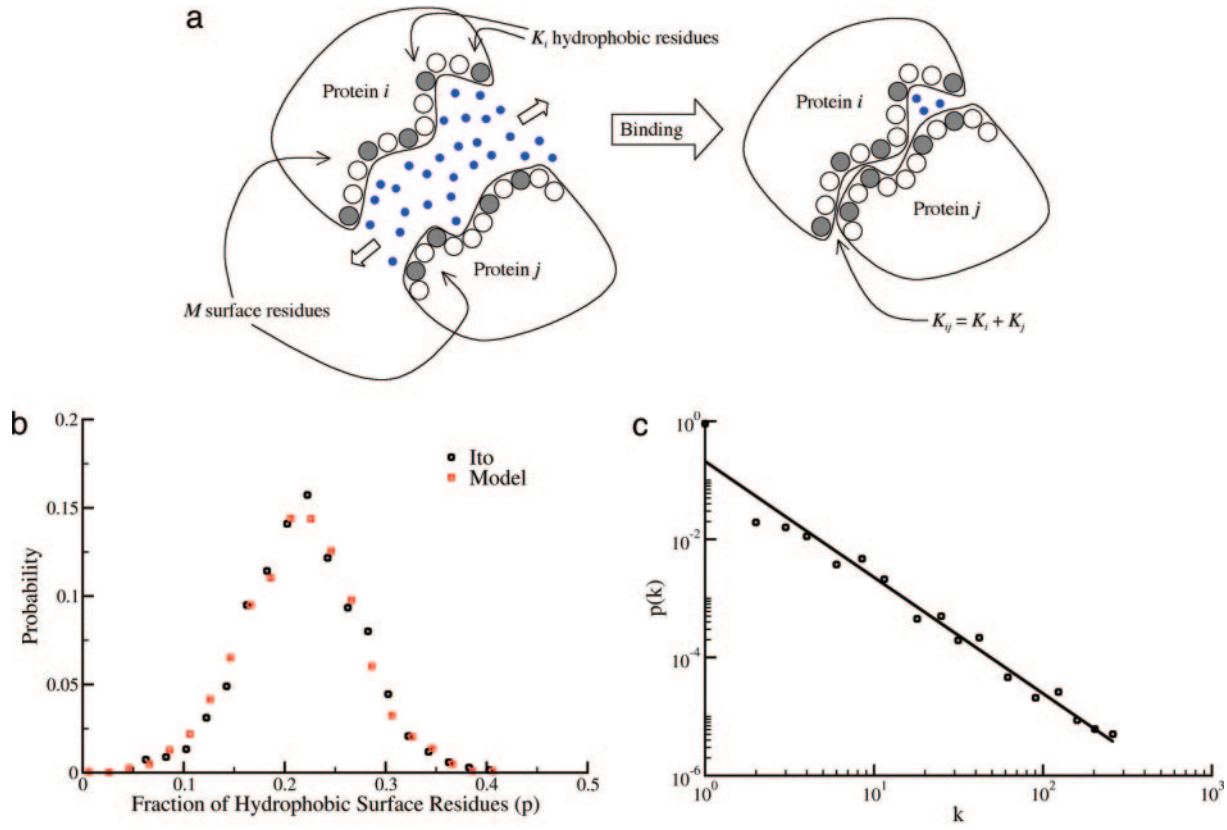

Fig. 2. A physical model for PPI measurements. (*a*) A schematic of the model described in the text. Association free energies are largely the result of desolvation of the two protein surfaces. The overall burial of hydrophobic groups is represented by the sum of the contributions from each protein. (*b*) The distribution of surface hydrophobicities in yeast proteins. The fraction of surface residues that are hydrophobic (defined as residues AVILMFYW) is calculated according to the description in the supporting information. This distribution is taken from proteins in the Ito experiment (13). The red squares represent the model hydrophobicities sampled from a Gaussian distribution with the same mean and standard deviation as the Ito proteins themselves. (*c*) A degree distribution for the realization of the model used in *b*. The cutoff was chosen such that the power-law fit gives an exponent of approximately −2.0, close to that of Ito graph. The degrees in this plot are shifted by +1 to allow for orphans (nodes of degree 0) to be displayed on a log–log plot. Note that the fraction of orphans in the graph is very high.

arbitrary (see the discussion of A_C below and in the supporting information), and in the case of our results it is set to be 100.

We find that, within certain ranges of A_C (and its logarithm K_C), the networks created by this MpK model discussed above exhibit degree distributions that are well fit by power-law functions; a representative example is shown in Fig. 2*c* (for a discussion of the variance in the degree distributions for different realizations of this model, see the supporting information). This model indicates that, at stringent cutoffs, many of the nodes in the graph are orphans, a finding that fits well with experimental observations from both the Uetz and ItoCore data sets (12, 13) (note that, in contrast to the graphs from Fig. 3, orphans are displayed on the log–log plot in Fig. 2*c* by adding 1 to the degree of each node). This finding indicates that the apparent scaling in these systems could very easily arise from a set of completely nonspecific interactions that contain no evolutionary information. A_C determines the apparent power-law exponent γ and is the only truly fittable parameter in the model; for any value of M that is sufficiently large to capture the differences in p that exist in the population, one may obtain a degree distribution of a given γ simply by changing the value of A_C. The dependence of γ on the cutoff parameter, as well as the distribution of γ values obtained from different realizations of the MpK model at a given cutoff, are explored in greater detail in the supporting information. We have also solved the MpK model analytically in the limit of high connectivity and find that the power-law fit we observe is well justified given the limited number of proteins we are simulating (for a discussion of this analytical work, see the supporting information). This solution explicitly demonstrates that the power-law exponent should be related to the cutoff parameter (A_C or the cutoff in buried hydrophobicity, K_C). Although this model is mathematically related to other static models of scale-free networks (29), it is important to note that our model represents a model of PPI networks that attempts to consider the physics of protein binding and is based on a Gaussian distribution of some underlying property. It should also be noted that, although the MpK model represents a very useful model for comparison with the experimental results (see below), it is actually simply one member of a large class of models that produced scale-free networks based on Gaussian distributions of quantities from which graphs are built (see the analytical solution in the supporting information for an example of one such related model).

Random Noise. The above model, although suggestive, is not necessarily a complete model of all of the PPI experiments; for instance, in the case of the original Ito data set, the number of orphans is much smaller (the experiment reports many more connected nodes than our model predicts), and the degree distribution deviates from power-law behavior at small values of k (6). To better model both of these experimental observations, we add an elemental source of noise to our model by linking a

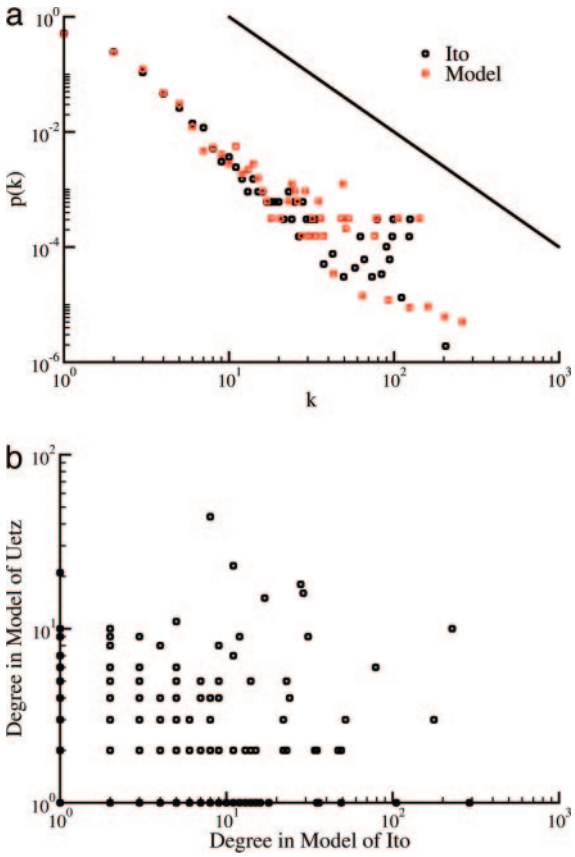

Fig. 3. Degree distributions and correlations for model PPI networks. (*a*) Comparison of degree distributions for the Ito data set (13) and a realization of the random linking model. In this case, all orphans from the model graph of 3,200 nodes are connected to one node that does exhibit connections randomly. The line represents a power-law with an exponent of −2. The degrees in this plot are not shifted as they are in Fig. 2*c*. (*b*) The correlation between degrees for in the model of Ito (13) compared with the model of Uetz (12). In this case, the different experiments are represented as independent sampling of values of *K* from a population of proteins with the distribution of *p* values shown in Fig. 2*b*. The Ito model is equivalent to the random linking results in *a*, and the Uetz degrees are taken from its random linking model (the degree distribution for that graph may be found in the supporting information). The linear correlation is 0.04 in this case.

number of orphans to randomly chosen nodes in the graph. To model a particular data set, we first fit the value of A_C to the power-law exponent that is observed in the experimental data. We randomly connect a number of orphans to nodes in the graph to obtain a number of connected nodes in the graph exactly equal to the number of connected nodes in the data set. The degree distributions of these random linking graphs exhibit surprisingly good agreement with the experimental results in all cases (see Fig. 3*a* for the Ito model and the supporting information for the ItoCore and Uetz models). The relationship between Ito and ItoCore is quite natural in this model: ItoCore is a data set that is obtained at higher stringency [representing the greater number of colonies needed to count an interaction in the ItoCore case (13)] with less random noise, and this is represented in the model by a higher value of A_C and fewer random links. The number of edges in the resulting graph is generally very close to that in the data set that is being modeled; for instance, in the case shown in Fig. 3*a* the number of edges is 2% larger than the number in the Ito data set (the results are similar for ItoCore and Uetz; see the supporting information). Although the above algorithm represents only one method of adding random noise to the system, other random linking strategies (such as adding a random link to every node in the graph regardless of connectivity) yield similar results (see the supporting information). In this model, the exposure of two different surfaces for individual proteins in the Uetz and Ito experiments represents creating graphs from two independent realizations of the MpK model, holding *p* fixed for each model protein but resampling *K* independently. Two model graphs sampled in this way exhibit a very low correlation between connectivities, as expected (Fig. 3*b*). In this case, the value of R^2 is 0.012, an order of magnitude smaller than that observed for the Ito vs. Uetz data sets, indicating that samplings of the surfaces in the two experiments are most likely not completely unrelated to one another. Nonetheless, the above results demonstrate that a purely physical model can produce networks that are unrelated topologically but nonetheless scale-free from a single population of proteins.

Recent studies have indicated that topological properties aside from the degree distribution also exhibit interesting scaling behaviors in the PPI and other biological networks (6, 17, 30–32). Perhaps most interesting is the fact that the clustering coefficient of a node (a measure of the tendency of a node's neighbors to contact one another, denoted *C*) scales with the degree of the node; i.e., $C(k) \sim k^{-2}$ (17, 30, 32). This finding has been explained in terms of a tendency for such networks to display "hierarchical modularity," but we find that our purely physical model displays similar scaling behavior in the absence of any considerations of modularity (supporting information). We also find that other, higher-order features of the graph, such as the relationship between the connectivity of a node and the average connectivity of its neighbors (31), is also observed in our physical model (supporting information). It is therefore clear that interpretation of global topological features in light of evolutionary or functional pressures is difficult to evaluate in the absence of purely physical, nonevolutionary controls, and these results highlight the potential utility of our model as a "null model" for understanding such observations in the future. Although it has also been shown that more local properties of a graph may potentially contain interesting evolutionary traces (33, 34), we leave exploration of those properties to future work.

Correlation between Connectivity and Hydrophobicity. Although the above graph theoretic results are suggestive, our model makes another key testable prediction that explicitly relates the MpK model to the physical reality of protein–protein binding. Specifically, our model suggests that a relationship should exist between the connectivity of a protein in the PPI network and the surface hydrophobicity of that protein. In the case of the experimental data, we do not know which specific surface is exposed in the experiment; we only know the (approximate) value of *p* for a subset of proteins. In the MpK model, it is clear that the average value of *K* will follow

$$\langle K \rangle = Mp, \qquad [2]$$

with a standard deviation (σ_K) of

$$\sigma_K = \sqrt{M(p)(1-p)}. \qquad [3]$$

These features of the binomial distribution of *K* indicate that averaging over populations of proteins with similar values of *p* should provide a method for overcoming the inherent uncertainty in the relationship between *p* and *K* (especially at values of *p* of ≈0.5, where σ_K is maximal). We thus expect a strong correlation between $\langle \log(k) \rangle$ and $\langle p \rangle$ at some bin size in *p* and a weak correlation between $\log(k)$ and *p* for individual proteins [given that affinity, not free energy, determines connectivity, the $\log(k)$ gives a stronger correlation than *k*]. The model also predicts that σ_K will increase with increasing *p* up to $p = 0.5$. The

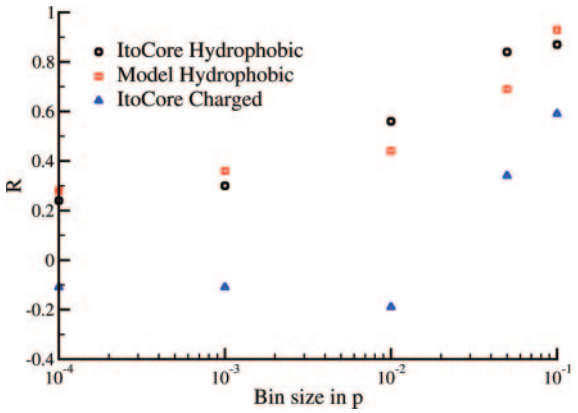

Fig. 4. Correlations between hydrophobicity and connectivity. The dependence of the correlation between $\langle \log(k) \rangle$ and $\langle p \rangle$ as a function of the bin size in p used to define the populations over which p and $\log(k)$ are averaged. The dependence for hydrophobic residues is very similar between the ItoCore data and the physical model for ItoCore. In the case of the charged data set, p is taken to be the percentage of charged residues on the surface, and the correlation dependence is calculate exactly as for the hydrophobic residues. None of the correlations for the charged data set are statistically significant.

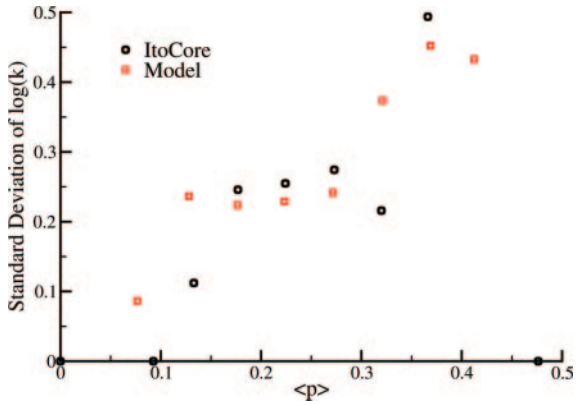

Fig. 5. Standard deviation in connectivity. As predicted by the MpK model, in the model and the ItoCore network, the standard deviation of $\log(k)$ in a given bin in p increases with increasing $\langle p \rangle$ for that bin. The bin size is set at 0.05 for the model and experimental networks. The lack of a maximum at $p = 0.5$ (as predicted by Eq. **3**) is due to the fact that very few proteins exist in those bins with large $\langle p \rangle$, decreasing the standard deviation for the most hydrophobic bin in each case.

MpK model therefore provides a mechanism whereby the empirically observed hydrophobicity distribution presented in Fig. 2*b* can be compared with the connectivities obtained from the Y2H experiments.

The above analysis introduces a new parameter into the system; namely, the bin size in p over which the averaging occurs. In general, larger bin sizes result in larger correlations but weaker statistical significance given the fact that fewer points are used to calculate the correlation. If we take the largest bin size in p that nonetheless results in a statistically significant correlation ($P < 0.05$), we find that the correlations between $\langle \log(k) \rangle$ and $\langle p \rangle$ are 0.84 for ItoCore, 0.79 for Uetz, and 0.17 for Ito (P values of 0.012, 0.025 and 0.014, respectively). The bin size for ItoCore, Uetz, and Ito in this case is 0.05, 0.05, and 0.001 units in p, respectively (although it is important to note that a correlation of 0.74 exists for Ito at a bin size of 0.05, but the P value in this case is 0.052, just above the P-value cutoff for significance). The maximum correlation is displayed for ItoCore and the ItoCore model graph (the graph with the degree distribution shown in Fig. 3*a*) in the supporting information (the maximum for the ItoCore model is 0.89 and occurs at a bin size of 0.05). The dependence of R on the bin size is similar between the model and the data (Fig. 4 for ItoCore and the supporting information for Uetz and Ito), although the correlation at intermediate bin sizes is somewhat larger in the model in all cases. The lower correlations in the experimental data are likely due to the fact that p is only approximately known for the proteins in the data sets but is exactly known for the model and the fact that every hydrophobic residue contributes equally to binding, whereas, for the experimental PPI networks, more bulky hydrophobic residues may contribute more to stickiness and thus to connectivity. In the case of the Ito data, our random-linking model indicates that there is a significant amount of noise at low values of k (especially for those nodes with $k = 1$). Consistent with this finding, the maximal correlation between $\langle \log(k) \rangle$ and $\langle p \rangle$ in Ito increases to 0.89 ($P = 0.019$) at a bin size of 0.05 when all $k = 1$ nodes are removed from the data set (see the supporting information).

It is important to note that the binning procedure does not produce statistically significant correlations between connectivity and other types of amino acids. For instance, we observe no statistically significant correlation between the percentage of charged amino acids (DEKR) on the surface and connectivity, despite the fact that such residues have been implicated in the determination of specificity in PPIs. The raw correlation between the percentage of charged residues on the surface and the log of connectivity is -0.09 for the ItoCore data set, and we observe no statistically significant correlation at any bin size (see Fig. 4; P values in this case are between 0.15 and 0.3 for all bin sizes, indicating a lack of statistical significance). As a further control, we calculated the correlation between connectivity and the percentage of eight randomly chosen amino acids and again find no statistically significant correlations at any bin size (data not shown). From these results it is clear that binning alone does not guarantee strong and statistically significant correlations.

A clear and nontrivial prediction of our model is that the standard deviation in the log of the connectivity will increase as the hydrophobicity of the surface increases up to a maximum at 0.5. The increase in standard deviation arises from the fact that higher values of p simply represent the possibility that a protein will expose a large number of hydrophobic residues (and thus exhibit a large connectivity) but does not ensure that the subset of residues that actually are involved in binding are actually hydrophobic. Consistent with this prediction, we find an increase in the dispersion in connectivity with increasing $\langle p \rangle$ (see Fig. 5 for the ItoCore and model results). We observe the same behavior is true for all of the experimental data sets (data not shown). Although the correlation results themselves represent strong support of our model, one could posit that proteins that make a large number of specific connections (as reported by the Y2H system) have simply evolved to be more hydrophobic. This argument would predict, however, that hydrophobic proteins would have universally large numbers of specific, evolved interactions, a finding that is directly contrary to our observation of increasing dispersion with increasing $\langle p \rangle$. It is therefore clear that the MpK model not only reproduces the lack of correlation between independent realizations of the same PPI experiment but also the dependence of both the average and standard deviation in connectivity as a function of hydrophobicity.

The model discussed above represents an extremely attractive alternative to evolutionary models of these graphs given that this model is inherently simpler and is based on very basic physical properties and not elaborate evolutionary mechanisms. This model also explains a number of the observations that have been made regarding PPI networks: the existence of scale-free net-

works in very noisy experiments, the lack of correlation between degrees in Uetz and Ito, the promiscuity of baits when compared with preys, and the scaling of C with k. The model reproduces such features of the experimental PPI networks based on only one fittable parameter (A_C). To our knowledge, no evolutionary model has exhibited all of the above features. Indeed, evolutionary models that produce pseudobipartite structures (35) are inherently unable to reproduce the observed $p(C)$ distribution and $C(k)$ behavior because each node in these networks has $C = 0$. Finally, the correlation of k with the fraction of hydrophobic surface residues (and the strong similarity in the behavior of this correlation between the model and the data as a function of the bin size in p) is a straightforward demonstration of the feasibility of our physical model.

Although the results of our model are very suggestive, our findings do not imply that the PPI experiments or especially curated online PPI databases do not contain any relevant biological or evolutionary information at all. Indeed it is possible to find weak correlations between biological observables and PPI network quantities (24), just as it is possible to find cases in which network features replicate findings from more careful biological studies. These weak correlations are completely consistent with a picture in which the majority of links in the network are the result of nonspecific interactions or experimental noise. Our results strongly indicate, however, that the interpretation of graph theoretic features of high-throughput experiments in the light of evolutionary processes must be tempered by the exploration of alternative physical hypotheses; indeed, this physical picture represents a null model against which future results regarding PPI networks should be measured. The model discussed above, although very important in terms of the PPI network, might also be used in various forms to describe other types of scale-free networks. Physical models based on additive or multiplicative processes could be used to provide explanations for many scale-free graphs, especially those that involve networks of macromolecules that bind one another.

We thank Drs. Nikolay Dokholyan, Boris Shakhnovich, and Leonid Mirny for comments on the manuscript. This work was supported by the National Institutes of Health. E.J.D. was supported by a Howard Hughes Medical Institute Predoctoral Fellowship

1. Albert, R. & Barabasi, A.-L. (2002) *Rev. Mod. Phys.* **74,** 47–97.
2. Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabasi, A. L. (2000) *Nature* **407,** 651–654.
3. Jeong, H., Mason, S. P., Barabasi, A. L. & Oltvai, Z. N. (2001) *Nature* **411,** 41–42.
4. Dokholyan, N. V., Shakhnovich, B. & Shakhnovich, E. I. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 14132–14136.
5. Doye, J. P. (2002) *Phys. Rev. Lett.* **88,** 238701.
6. Barabasi, A. L. & Oltvai, Z. N. (2004) *Nat. Rev. Genet.* **5,** 101–113.
7. Barabasi, A. L. & Albert, R. (1999) *Science* **286,** 509–512.
8. Sole, R. V., Pastor-Satorras, R., Smith, E. & Kepler, T. B. (2002) *Adv. Complex Syst.* **5,** 43–54.
9. Wagner, A. (2003) *Proc. R. Soc. London Ser. B* **270,** 457–466.
10. Chung, F., Lu, L., Dewey, T. G. & Galas, D. J. (2003) *J. Comput. Biol.* **10,** 677–687.
11. Eisenberg, E. & Levanon, E. Y. (2003) *Phys. Rev. Lett.* **91,** 138701.
12. Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., *et al*. (2000) *Nature* **403,** 623–627.
13. Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M. & Sakaki, Y. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 4569–4574.
14. Li, S., Armstrong, C. M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P. O., Han, J. D., Chesneau, A., Hao, T., *et al*. (2004) *Science* **303,** 540–543.
15. Giot, L., Bader, J. S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y. L., Ooi, C. E., Godwin, B., Vitols, E., *et al*. (2003) *Science* **302,** 1727–1736.
16. Rain, J. C., Selig, L., De Reuse, H., Battaglia, V., Reverdy, C., Simon, S., Lenzen, G., Petel, F., Wojcik, J., Schachter, V., *et al*. (2001) *Nature* **409,** 211–215.
17. Yook, S. H., Oltvai, Z. N. & Barabasi, A. L. (2004) *Proteomics* **4,** 928–942.
18. Wagner, A. (2001) *Mol. Biol. Evol.* **18,** 1283–1292.
19. Aloy, P. & Russell, R. B. (2002) *FEBS Lett.* **530,** 253–254.
20. Bader, J. S., Chaudhuri, A., Rothberg, J. M. & Chant, J. (2004) *Nat. Biotechnol.* **22,** 78–85.
21. Phizicky, E., Bastiaens, P. I., Zhu, H., Snyder, M. & Fields, S. (2003) *Nature* **422,** 208–215.
22. Deane, C. M., Salwinski, L., Xenarios, I. & Eisenberg, D. (2002) *Mol. Cell. Proteomics* **1,** 349–356.
23. Formstecher, E., Aresta, S., Collura, V., Hamburger, A., Meil, A., Trehin, A., Reverdy, C., Betin, V., Maire, S., Brun, C., *et al*. (2005) *Genome Res.* **15,** 376–384.
24. Fraser, H. B., Hirsh, A. E., Steinmetz, L. M., Scharfe, C. & Feldman, M. W. (2002) *Science* **296,** 750–752.
25. Eisenberg, D. & McLachlan, A. D. (1986) *Nature* **319,** 199–203.
26. Noskov, S. Y. & Lim, C. (2001) *Biophys. J.* **81,** 737–750.
27. Baker, B. M. & Murphy, K. P. (1997) *J. Mol. Biol.* **268,** 557–569.
28. Chandler, D. (2005) *Nature* **437,** 640–647.
29. Caldarelli, G., Capocci, A., De Los Rios, P. & Munoz, M. A. (2002) *Phys. Rev. Lett.* **89,** 258702.
30. Oltvai, Z. N. & Barabasi, A. L. (2002) *Science* **298,** 763–764.
31. Maslov, S. & Sneppen, K. (2002) *Science* **296,** 910–913.
32. Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. & Barabasi, A. L. (2002) *Science* **297,** 1551–1555.
33. Spirin, V. & Mirny, L. A. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 12123–12128.
34. Middendorf, M., Ziv, E. & Wiggins, C. H. (2005) *Proc. Natl. Acad. Sci. USA* **102,** 3192–3197.
35. Kim, J., Krapivsky, P. L., Kahng, B. & Redner, S. (2002) *Phys. Rev. E.: Stat., Nonlinear, Soft Matter Phys.* **66,** 055101.