How Perfect Can Protein Interactomes Be?
Emmanuel D. Levy, Christian R. Landry and Stephen W. Michnick (3 March 2009)

Science Signaling **2** (60), pe11. [DOI: 10.1126/scisignal.260pe11]

PROTEOMICS

How Perfect Can Protein Interactomes Be?

Emmanuel D. Levy,* Christian R. Landry,* Stephen W. Michnick†

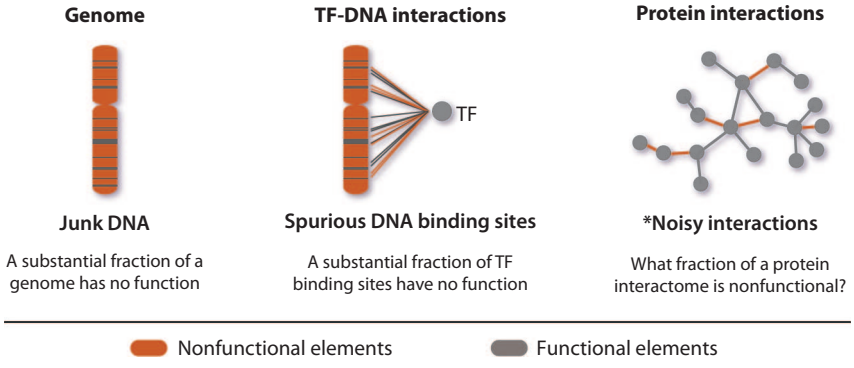
Published 3 March 2009; Volume 2 Issue 60 pe11

Any engineered device should certainly not contain nonfunctional components, for this would be a waste of energy and money. In contrast, evolutionary theory tells us that biological systems need not be optimized and may very well accumulate nonfunctional elements. Mutational and demographic processes contribute to the cluttering of eukaryotic genomes and transcriptional networks with "junk" DNA and spurious DNA binding sites. Here, we question whether such a notion should be applied to protein interactomes—that is, whether these protein interactomes are expected to contain a fraction of nonselected, nonfunctional protein-protein interactions (PPIs), which we term "noisy." We propose a simple relationship between the fraction of noisy interactions expected in a given organism and three parameters: (i) the number of mutations needed to create and destroy interactions, (ii) the size of the proteome, and (iii) the fitness cost of noisy interactions. All three parameters suggest that noisy PPIs are expected to exist. Their existence could help to explain why PPIs determined from large-scale studies often lack functional relationships between interacting proteins, why PPIs are poorly conserved across organisms, and why the PPI space appears to be immensely large. Finally, we propose experimental strategies to estimate the fraction of evolutionary noise in PPI networks.

Among the "omics" networks recently unveiled, transcriptional regulatory networks and protein-protein interaction (PPI) networks have received a lot of attention. How these networks reflect the reality of cellular interactomes, however, remains unclear (*1*, *2*). These protein interactomes in particular have been susceptible to questions about their biological meaning, and two opposite views have emerged. A skeptical view postulates that, despite being experimentally reproducible, many interactions have no apparent function and may not exist in vivo. These spurious interactions may result from technical limitations of PPI assays, where proteins may not be expressed at the appropriate level, within a relevant time frame, or at the correct location in the cell (*3*, *4*). This view also interprets the poor overlap between experiments conducted by different research groups as a consequence of such "false positive" interactions. In contrast, an optimistic view notes the ongoing discovery of novel interactions when PPI screens are performed with proteins at the appropriate physiologically relevant abundance (*5*, *6*), at the appropriate time and place (*7*), or when expressed from single-copy plasmids (*8*). The optimistic view thus implies that PPIs observed in large-scale screens should occur in vivo and that the poor overlap between different experimental technologies would stem from low sensitivities of the methods; that is, each PPI screen would report only a small fraction of the entire interaction space, which is estimated as ~20,000 unique pairs in yeast (*8*) and ~130,000 in humans (*9*). Nonetheless, many of these interactions make little functional sense. Thus, if we accept that so many PPIs are indeed taking place in vivo, what could their biological meaning be? Trying to answer this question brings this debate beyond technical aspects of methodologies.

The issue of false positives is less controversial for transcriptional regulatory networks, despite experiments that also commonly report hundreds of binding events with no apparent functional meaning (*10*). The existence of such imperfection in transcriptional regulatory networks is perhaps easier to conceptualize, because it is possible to estimate the number of expected nonfunctional binding sites for a given transcription factor. Put simply, a nucleotide motif of length n is expected to appear $\sim L/4^n$ times in a random DNA sequence of length L. Probably because interactions between proteins are far more difficult to formalize and predict than transcription factor–DNA interactions (*11*), no such rationale has yet been applied to PPIs. This may be why appreciation of imperfection in PPI networks is unacknowledged, and why it is often assumed that interactomes are optimally specific (*12*–*14*). However, the skeptical and optimistic views could be reconciled substantially if we accept that a large number of PPIs may be the product of evolutionary noise—that is, that they do exist in vivo and can be detected by a PPI assay, but have not been selected for a specific function (Fig. 1). Rather, they would result from neutral or nearly neutral mutations, or as a correlation to another, independent function. The existence of such "noisy" PPIs would help to explain why their space appears remarkably large, why their function can be diffi-

Fig. 1. Nonfunctional elements are present at different levels of cellular organization. Nonfunctional elements, such as junk DNA (*30*) and spurious DNA binding sites (*31*), are present in genomes. Nonfunctional protein-protein interactions may also exist and represent "noisy" interactions. TF, transcription factor.



*Noisy interactions: Interaction that exists in vivo, with an affinity above typical assay detection thresholds (μM), but that has not been selected for a specific function. It can rather be considered as evolutionary noise.

Université de Montréal, C.P. 6128, Succ. Centre-Ville, Montreal, Quebec H3C 3J7, Canada.

*These authors contributed equally to this work.

†Corresponding author. E-mail, stephen.michnick@umontreal.ca

cult to interpret, and why they exhibit little conservation among species (*15*). If such noise exists, what is its extent?

What biophysical and evolutionary parameters shall we consider? In transcription factor networks, the number of noisy interactions decreases as the average size of transcription factor binding sites increases—this defines how often they appear through random mutations—and grows with genome size. Ultimately, their number depends on the selection pressure acting against their fixation in populations. A relationship between these three parameters and noise was previously described in the context of transcription factor networks (*16*). Here, we extend the relationship to PPI networks in Eq. 1 in Fig. 2A, where the amount of noise depends on three equivalent parameters. We discuss each parameter below in order to gauge whether noise is expected to exist in protein interactomes.

A basal level of noise to expect is set by the ratio of probabilities that a random mutation abolishes a noisy interaction or creates one (p_{loss}/p_{gain}). Each probability implicitly accounts for any change that can alter a protein's ability to bind another protein, including changes in surface-exposed shape or hydrophobicity, or changes in abundance or localization. This ratio means that if gaining a noisy interaction is as likely as losing one by random mutation, noise should be present. Although we know that point mutations may affect the stability of an interaction (*17*), it is more difficult to imagine how they could result in the formation of a new PPI. Yet, like protein-DNA interactions (*18*), PPIs can appear through only a few amino acid substitutions (*19*, *20*) or even from single point mutations (*21*, *22*). Although these examples do not give precise estimates of p_{loss} and p_{gain}, they illustrate that these probabilities should not differ by more than a few orders of magnitude.

Moreover, the larger the proteome size (N), the larger the number of potential noisy interaction pairs (N^2) (*23*), which is reflected by the multiplication of p_{gain} by N^2 in Eq. 1 in Fig. 2A. This illustrates that strategies such as compartmentalization and differential expression contribute to minimize the effective number of proteins that may interact with each other within a cell (*23*) [~1800 in yeast cytoplasm (*24*)]. As a result, even if p_{gain} were much smaller than p_{loss}, $1800^2 \times p_{gain}$ is unlikely to be much smaller than p_{loss}. This should also hold true for other eukaryotes where N is comparable or larger, and for prokaryotes where there are no cell compartments.

A third parameter is the tolerance of biological systems to noisy interactions, which is presented here as γ, the fixation probability of a mutation relative to a neutral mutation (Fig. 2B). The assumption that natural selection has optimized protein interactomes (*12*–*14*) considers that nonfunctional PPIs have either an important cost or deleterious effects, such that when they occur, their probability to be fixed in a population is null or extremely small. However, the deleterious effects of noisy interactions remain largely unknown. Furthermore, the probability of elimination of a deleterious mutation by natural selection depends on population genetics parameters, with a lower probability of elimination in small populations relative to larger ones (*25*). Typically, for organisms with small population sizes, such as mammals, a substantial fraction of the genomic characteristics are likely to have no function or even to be far from optimal (*26*, *27*). For example, the large number of transposable elements in the human genome might represent a cost, but even so, they were never eliminated.

All three parameters thus indicate that noisy interactions may exist, much as in transcriptional regulatory networks or, as suggested, in kinase-substrate networks (*28*). A better interpretation of PPI networks will thus require estimating the parameters of Eq. 1—if not at the network level, then possibly at the level of single proteins, where N^2 should be replaced by N in the same equation. For this purpose, a protein mutagenesis experiment combined with a large-scale PPI detection strategy could be envisioned. Also, nonfunctional PPIs could be engineered in order to estimate their fitness cost. At the same time, it appears crucial to assess the functionality of known PPIs. An approach to this could be the systematic use of orthogonal information, such as PPI conservation among closely related species, because noisy interactions are not functionally constrained and should not be conserved. Indeed, this approach was successfully used to delineate the yeast regulatory code (*29*). As a matter of practice, the existence of evolutionary noise implies that all PPIs reported by large-scale projects, with the exception of provable technical artifacts, should be considered without prejudice by the scientific community. As in

A

Assuming that n is at equilibrium and that $n << N^2$, then:

$$n \sim \frac{p_{gain}}{p_{loss}} \cdot N^2 \cdot \gamma \quad (1)$$

n	Number of protein pairs that exhibit a noisy interaction
N	Number of proteins in the proteome
N^2	Number of possible interaction pairs in a proteome of size N
p_{loss}	Probability that a random mutation disrupts a noisy interaction
p_{gain}	Probability that a random mutation creates a noisy interaction
γ	Fixation probability of a mutation that creates a noisy interaction relative to the fixation probability of a neutral mutation

B

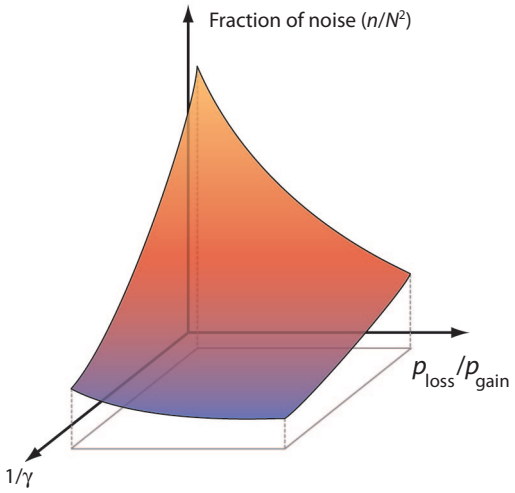


Fig. 2. Noisy interactions are expected in protein interactomes. (**A**) The extent of evolutionary noise in protein interactomes depends on three parameters: N, the ratio p_{gain}/p_{loss}, and γ. (**B**) Graphing the relationship between these three parameters shows that the smaller the value of p_{loss}/p_{gain} or 1/γ, the more noise exists in a given interactome. The graph depicts a qualitative picture, and experiments are needed to estimate the values of these parameters.

transcriptional networks, cells withstand or even make use of evolutionary noise, and we must learn to tolerate it as an essential signature of dynamic evolution.

References and Notes

1. L. J. Jensen, P. Bork, Not comparable, but complementary. *Science* **322**, 56–57 (2008).
2. S. Balaji, L. M. Iyer, M. M. Babu, L. Aravind, Comparison of transcription regulatory interactions inferred from high-throughput methods: What do they reveal? *Trends Genet.* **24**, 319–323 (2008).
3. J. P. Mackay, M. Sunde, J. A. Lowry, M. Crossley, J. M. Matthews, Protein interactions: Is seeing believing? *Trends Biochem. Sci.* **32**, 530–531 (2007).
4. C. von Mering, R. Krause, B. Snel, M. Cornell, S. G. Oliver, S. Fields, P. Bork, Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* **417**, 399–403 (2002).
5. N. J. Krogan, G. Cagney, H. Yu, G. Zhong, X. Guo, A. Ignatchenko, J. Li, S. Pu, N. Datta, A. P. Tikuisis, T. Punna, J. M. Peregrin-Alvarez, M. Shales, X. Zhang, M. Davey, M. D. Robinson, A. Paccanaro, J. E. Bray, A. Sheung, B. Beattie, D. P. Richards, V. Canadien, A. Lalev, F. Mena, P. Wong, A. Starostine, M. M. Canete, J. Vlasblom, S. Wu, C. Orsi, S. R. Collins, S. Chandran, R. Haw, J. J. Rilstone, K. Gandi, N. J. Thompson, G. Musso, P. St. Onge, S. Ghanny, M. H. Lam, G. Butland, A. M. Altaf-Ul, S. Kanaya, A. Shilatifard, E. O'Shea, J. S. Weissman, C. J. Ingles, T. R. Hughes, J. Parkinson, M. Gerstein, S. J. Wodak, A. Emili, J. F. Greenblatt, Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**, 637–643 (2006).
6. A. C. Gavin, P. Aloy, P. Grandi, R. Krause, M. Boesche, M. Marzioch, C. Rau, L. J. Jensen, S. Bastuck, B. Dumpelfeld, A. Edelmann, M. A. Heurtier, V. Hoffman, C. Hoefert, K. Klein, M. Hudak, A. M. Michon, M. Schelder, M. Schirle, M. Remor, T. Rudi, S. Hooper, A. Bauer, T. Bouwmeester, G. Casari, G. Drewes, G. Neubauer, J. M. Rick, B. Kuster, P. Bork, R. B. Russell, G. Superti-Furga, Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631–636 (2006).
7. K. Tarassov, V. Messier, C. R. Landry, S. Radinovic, M. M. Serna Molina, I. Shames, Y. Malitskaya, J. Vogel, H. Bussey, S. W. Michnick, An in vivo map of the yeast protein interactome. *Science* **320**, 1465–1470 (2008); published online 7 May 2008 (10.1126/science.1153878).
8. H. Yu, P. Braun, M. A. Yildirim, I. Lemmens, K. Venkatesan, J. Sahalie, T. Hirozane-Kishikawa, F. Gebreab, N. Li, N. Simonis, T. Hao, J. F. Rual, A. Dricot, A. Vazquez, R. R. Murray, C. Simon, L. Tardivo, S. Tam, N. Svrzikapa, C. Fan, A. S. de Smet, A. Motyl, M. E. Hudson, J. Park, X. Xin, M. E. Cusick, T. Moore, C. Boone, M. Snyder, F. P. Roth, A. L. Barabási, J. Tavernier, D. E. Hill, M. Vidal, High-quality binary protein interaction map of the yeast interactome network. *Science* **322**, 104–110 (2008); published online 21 August 2008 (10.1126/science.1158684).
9. K. Venkatesan, J.-F. Rual, A. Vazquez, U. Stelzl, I. Lemmens, T. Hirozane-Kishikawa, T. Hao, M. Zenkner, X. Xin, K.-I. Goh, M. A. Yildirim, N. Simonis, K. Heinzmann, F. Gebreab, J. M. Sahalie, S. Cevik, C. Simon, A.-S. de Smet, E. Dann, A. Smolyar, A. Vinayagam, H. Yu, D. Szeto, H. Borick, A. Dricot, N. Klitgord, R. R. Murray, C. Lin, M. Lalowski, J. Timm, K. Rau, C. Boone, P. Braun, M. E. Cusick, F. P. Roth, D. E. Hill, J. Tavernier, E. E. Wanker, A.-L. Barabasi, M. Vidal, An empirical framework for binary interactome mapping. *Nat. Methods* **6**, 83–90 (2008).
10. G. Euskirchen, M. Snyder, A plethora of sites. *Nat. Genet.* **36**, 325–326 (2004).
11. J. Janin, S. Wodak, The Third CAPRI Assessment Meeting Toronto, Canada, April 20-21, 2007. *Structure* **15**, 755–759 (2007).
12. M. A. Stiffler, J. R. Chen, V. P. Grantcharova, Y. Lei, D. Fuchs, J. E. Allen, L. A. Zaslavskaia, G. MacBeath, PDZ domain binding selectivity is optimized across the mouse proteome. *Science* **317**, 364–369 (2007).
13. A. Zarrinpar, S. H. Park, W. A. Lim, Optimization of specificity in a cellular protein interaction network by negative selection. *Nature* **426**, 676–680 (2003).
14. D. Endy, M. B. Yaffe, Signal transduction: Molecular monogamy. *Nature* **426**, 614–615 (2003).
15. P. Beltrao, L. Serrano, Specificity and evolvability in eukaryotic protein interaction networks. *PLOS Comput. Biol.* **3**, e25 (2007).
16. M. W. Hahn, J. E. Stajich, G. A. Wray, The effects of selection against spurious transcription factor binding sites. *Mol. Biol. Evol.* **20**, 901–906 (2003).
17. T. Clackson, J. A. Wells, A hot spot of binding energy in a hormone-receptor interface. *Science* **267**, 383–386 (1995).
18. D. Menendez, A. Inga, J. Snipe, O. Krysiak, G. Schonfelder, M. A. Resnick, A single-nucleotide polymorphism in a half-binding site creates p53 and estrogen receptor control of vascular endothelial growth factor receptor 1. *Mol. Cell. Biol.* **27**, 2590–2600 (2007).
19. J. M. Skerker, B. S. Perchuk, A. Siryaporn, E. A. Lubin, O. Ashenberg, M. Goulian, M. T. Laub, Rewiring the specificity of two-component signal transduction systems. *Cell* **133**, 1043–1054 (2008).
20. J. Kuriyan, D. Eisenberg, The origin of protein interactions and allostery in colocalization. *Nature* **450**, 983–990 (2007).
21. D. Grueninger, N. Treiber, M. O. Ziegler, J. W. Koetter, M.-S. Schulze, G. E. Schulz, Designed protein-protein association. *Science* **319**, 206–209 (2008).
22. C. T. Rollins, V. M. Rivera, D. N. Woolfson, T. Keenan, M. Hatada, S. E. Adams, L. J. Andrade, D. Yaeger, M. R. van Schravendijk, D. A. Holt, M. Gilman, T. Clackson, A ligand-reversible dimerization system for controlling protein-protein interactions. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7096–7101 (2000).
23. J. Zhang, S. Maslov, E. I. Shakhnovich, Constraints imposed by non-functional protein-protein interactions on gene expression and proteome size. *Mol. Syst. Biol.* **4**, 10.1038/msb.2008.48 (2008).
24. W. K. Huh, J. V. Falvo, L. C. Gerke, A. S. Carroll, R. W. Howson, J. S. Weissman, E. K. O'Shea, Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
25. D. Hartl, A. Clark, *Principles of Population Genetics* (Sinauer, Sunderland, MA, ed. 4, 2007).
26. M. Lynch, The evolution of genetic networks by non-adaptive processes. *Nat. Rev. Genet.* **8**, 803–813 (2007).
27. M. Lynch, The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc. Natl. Acad. Sci. U.S.A.* **104** (suppl. 1), 8597–8604 (2007).
28. G. E. Lienhard, Non-functional phosphorylations? *Trends Biochem. Sci.* **33**, 351–352 (2008).
29. C. T. Harbison, D. B. Gordon, T. I. Lee, N. J. Rinaldi, K. D. Macisaac, T. W. Danford, N. M. Hannett, J. B. Tagne, D. B. Reynolds, J. Yoo, E. G. Jennings, J. Zeitlinger, D. K. Pokholok, M. Kellis, P. A. Rolfe, K. T. Takusagawa, E. S. Lander, D. K. Gifford, E. Fraenkel, R. A. Young, Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**, 99–104 (2004).
30. M. A. Nobrega, Y. Zhu, I. Plajzer-Frick, V. Afzal, E. M. Rubin, Megabase deletions of gene deserts result in viable mice. *Nature* **431**, 988–993 (2004).
31. Q. Li, S. A. Johnston, Are all DNA binding and transcription regulation by an activator physiologically relevant? *Mol. Cell. Biol.* **21**, 2467–2474 (2001).

32. We thank N. Lartillot, M. Babu, S. Janga, and D. Hartl for useful comments on the manuscript. Supported by Canadian Institutes of Health Research (CIHR) grant MOP-GMX-152556 (S.W.M.), the EMBO postdoctoral fellowship program (E.D.L.), and the CIHR postdoctoral fellowship program (C.R.L.).

10.1126/scisignal.260pe11

Citation: E. D. Levy, C. R. Landry, S. W. Michnick, How perfect can protein interactomes be? *Sci. Signal.* **2**, pe11 (2009).