Propagation of large concentration changes in reversible protein-binding networks

Sergei Maslov*† and I. Ispolatov†‡

*Department of Condensed Matter Physics and Materials Science, Brookhaven National Laboratory, Upton, NY 11973; and ‡Ariadne Genomics, Inc., 9430 Key West Avenue, Suite 113, Rockville, MD 20850

Edited by Peter G. Wolynes, University of California at San Diego, La Jolla, CA, and approved June 21, 2007 (received for review March 28, 2007)

We study how the dynamic equilibrium of the reversible protein–protein-binding network in yeast *Saccharomyces cerevisiae* responds to large changes in abundances of individual proteins. The magnitude of shifts between free and bound concentrations of their immediate and more distant neighbors in the network is influenced by such factors as the network topology, the distribution of protein concentrations among its nodes, and the average binding strength. Our primary conclusion is that, on average, the effects of a perturbation are strongly localized and exponentially decay with the network distance away from the perturbed node, which explains why, despite globally connected topology, individual functional modules in such networks are able to operate fairly independently. We also found that under specific favorable conditions, realized in a significant number of paths in the yeast network, concentration perturbations can selectively propagate over considerable network distances (up to four steps). Such "action-at-a-distance" requires high concentrations of heterodimers along the path as well as low free (unbound) concentration of intermediate proteins.

dissociation constant | genetic interactions | law of mass action | small-world networks | binding equilibrium

Recent high-throughput experiments performed in a wide variety of organisms revealed networks of protein–protein physical interactions (PPI) that are interconnected on a genome-wide scale. In such "small-world" PPI networks, most pairs of nodes can be linked to each other by relatively short chains of interactions involving just a few intermediate proteins (1). Although globally connected architecture facilitates biological signaling and possibly ensures a robust functioning of the cell after a random failure of its components (2), it also presents a potential problem by providing a conduit for propagation of undesirable cross-talk between individual functional modules and pathways. Indeed, large (severalfold) changes in proteins' levels in the course of activation or repression of a certain functional module affect bound concentrations of their immediate interaction partners. These changes have a potential to cascade down a small-world PPI network affecting the equilibrium between bound and unbound concentrations of progressively more distant neighbors, including those in other functional modules. Most often such indiscriminate propagation would represent an undesirable effect that has to be either tolerated or corrected by the cell. On the other hand, a controlled transduction of reversible concentration changes along specific conduits may be used for biologically meaningful signaling and regulation. A routine and well known example of such regulation is inactivation of a protein by sequestration with its strong binding partner.

In this study, we quantitatively investigate how large concentration changes propagate in the PPI network of yeast *S. cerevisiae*. We focus on the noncatalytic or reversible binding interactions whose equilibrium is governed by the law of mass action (LMA) and do not consider irreversible, catalytic processes such as protein phosphorylation and dephosphorylation, proteolytic cleavage, etc. Although such catalytic interactions constitute the most common and best studied mechanism of intracellular signaling, they represent only a rather small minority of all PPI (for example, only ≈5% of links in the yeast network used in our study involve a kinase).

Furthermore, the balance between free and bound concentrations of proteins matters even for irreversible (catalytic) interactions. For example, the rate of a phosphorylation reaction depends on the availability of free kinases and substrate proteins, which are both controlled by the LMA equilibrium calculated here. Thus, perturbations of equilibrium concentrations considered in this study could be spread even further by other mechanisms such as transcriptional and translational regulation and irreversible post-translational protein modifications.

Further information is available in supporting information (SI) *Appendix*, SI Figs. 7–10, and SI Tables 2–6.

Results

To illustrate general principles on a concrete example, in this study, we used a highly curated genome-wide network of PPI in yeast (*S. cerevisiae*), which, according to the BIOGRID database (3), were independently confirmed in at least two publications. We combined this network with a genome-wide data set of protein abundances in the log-phase growth in rich medium, measured by the TAP-tagged Western blotting technique (4). Average protein concentrations in this data set range between 50 and 1,000,000 molecules per cell with the median value ≈3,000 molecules per cell. After keeping only the interactions between proteins with known concentrations, we were left with 4,185 binding interactions among 1,740 proteins (Table 2). The BIOGRID database (3) lists all interactions as pairwise and thus lacks information about multiprotein complexes larger than dimers. Thus, in the main part of this study, we consider only homodimers and heterodimers and ignore the formation of higher-order complexes. In *SI Appendix*, we show that the reliable data on multiprotein complexes can easily be incorporated into our analysis. Furthermore, we demonstrate that taking into account such complexes leaves our results virtually unchanged (see SI Table 3 and SI Fig. 9).

The state-of-the-art genome-wide PPI data sets lack information on dissociation constants K_{ij} of individual interactions. The only implicit assumption is that the binding is sufficiently strong to be detectable by a particular experimental technique [some tentative bounds on dissociation constants detectable by different techniques were reported recently (5)]. A rough estimate of the average binding strength in functional protein–protein interactions could be obtained from the PINT database (6). This database contains ≈400 experimentally measured dissociation constants between wild-type proteins from a variety of organisms. In agreement with the

Author contributions: S.M. designed research; S.M. and I.I. performed research; S.M. and I.I. analyzed data; and S.M. and I.I. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Abbreviations: PPI, protein–protein physical interactions; LMA, law of mass action.

†To whom correspondence may be addressed. E-mail: maslov@bnl.gov or slava@ariadnegenomics.com.

This article contains supporting information online at www.pnas.org/cgi/content/full/0702905104/DC1.

© 2007 by The National Academy of Sciences of the USA

BIOPHYSICS

predictions of refs. 7 and 8, the histogram of these dissociation constants has an approximately log-normal shape. The average relevant for our calculations is that of the association constant $\langle 1/K_{ij} \rangle = 1/(5\ \text{nM})$. Common sense dictates that the dissociation constant of a functional binding between a pair of proteins should increase with their abundances. The majority of specific physical interactions between proteins are neither too weak (to ensure a considerable number of bound complexes) nor unnecessarily strong. Indeed, there is little evolutionary sense in increasing the binding strength between a pair of proteins beyond the point when both proteins (or at least the rate-limiting one) spend most of their time in the bound state. The balance between these two opposing requirements is achieved by the value of dissociation constant K_{ij} equal to a fixed fraction of the largest of the two abundances C_i and C_j of interacting proteins. In some of our simulations, we used $K_{ij} = \max(C_i, C_j)/20$ in which case the average association constant nicely agrees with its empirical value [1/(5 nM)] observed in the PINT database (6). In addition to this, perhaps, more realistic assignment of dissociation constants, we also simulated binding networks in which dissociation constants of all 4,185 edges in our network are equal to each other and given by 1 nM, 10 nM, 100 nM, and 1 μM.

Numerical Calculation of Bound and Free (Unbound) Equilibrium Concentrations. The LMA relates the free (unbound) concentration F_i of a protein to its experimentally known (4) total (bound and unbound) concentration C_i as

$$F_i = \frac{C_i}{1 + \sum_j F_j/K_{ij}}. \qquad [1]$$

Here the sum is over all specific binding partners of the protein i with free concentrations F_j and dissociation constants K_{ij}. Although in the general case these nonlinear equations do not allow for an analytical solution for F_i, they readily are solved numerically, e.g., by successive iterations.

Concentration-Coupled Proteins. To investigate how large changes in abundances of individual protein affect the equilibrium throughout the PPI network, we performed a systematic numerical study in which we recalculated the equilibrium free concentrations of all protein nodes after a 2-fold increase in the total concentration of just one of them: $C_i \rightarrow 2C_i$. This process was repeated for the source of 2-fold perturbation spanning the set of all 1,740 of proteins in our network.[§]

The magnitude of the initial perturbation was selected to be representative of a typical shift in gene expression levels or protein abundances after a change in external or internal conditions. Thus, here we simulate the propagation of functionally relevant changes in protein concentrations and not that of background stochastic fluctuations. A change in the free concentration F_j of another protein was deemed to be significant if it exceeded the 20% level, which according to ref. 9 is the average magnitude of cell-to-cell variability of protein abundances in yeast. We refer to such protein pairs $i \rightarrow j$ as "concentration-coupled." The detection threshold could be raised simultaneously with the magnitude of the initial perturbation. For example, we found that the list of concentration-coupled pairs changes very little if instead of 2-fold (+100%) perturbation and the 20% detection threshold one applies a 6-fold (+500%) initial perturbation and 2-fold (100%) detection threshold.

In general, we found that lists of concentration-coupled proteins calculated for different assignments of dissociation constants strongly overlap with each other. For example, more than 80% of concentration-coupled pairs observed for the variable $K_{ij} = \max(C_i, C_j)/20$ assignment described above also were detected for the uniform assignment $K_{ij} = \text{const} = 10$ nM (for more details, see SI Table 4) This relative robustness of our results allowed us to use the latter conceptually simplest case to illustrate our findings in the rest of the article.

The complete list of concentration-coupled pairs is included in SI Table 2. Given the incompleteness and uncertainty in our knowledge of the network topology, protein abundances, and values of dissociation constants, these lists provide only a rough estimate of the actual magnitude of perturbations that could be measured experimentally.

Central Observations. We found that

- On average, the magnitude of cascading changes in equilibrium free concentrations exponentially decays with the distance from the source of a perturbation, which explains why, despite a globally connected topology, individual modules in such networks are able to function fairly independently.
- Nevertheless, specific favorable conditions identified in our study cause perturbations to selectively affect proteins at considerable network distances (sometimes as far as four steps away from the source). This finding indicates that, in general, such cascading changes could not be neglected when evaluating the consequences of systematic changes in protein levels, e.g., in response to environmental factors or in gene knockout experiments. Conditions favorable for propagation of perturbations combine high yet monotonically decreasing concentrations of all heterodimers along the path with low free (unbound) concentrations of intermediate proteins. Although reversible protein-binding links are symmetric, the propagation of concentration changes usually is asymmetric with the preferential direction pointing down the gradient in the total concentrations of proteins.

Examples of Multistep Cascading Changes. In Fig. 1, we illustrate these observations by using two examples. In each of these cases, the 2-fold increase in the abundance of just one protein (marked with the yellow circle in the center of both Fig. 1 *A* and *B*) has significantly (>20%) affected equilibrium free concentrations of a whole cluster of proteins some as far as four steps away from the source of the perturbation. However, the propagation beyond immediate neighbors is rather specific. For example, in the case of SUP35 (Fig. 1*A*), only 1 of 169 of its third nearest neighbors were affected above the 20% level. Note that changes in free concentrations generally sign-alternate with the network distance from the source. Indeed, free concentrations of immediate binding partners of the perturbed protein usually drop as more of them become bound in heterodimers with it. This, in turn, lowers concentrations of the next-nearest heterodimers and thus increases free concentrations of proteins at distance two from the source of perturbation, and so on.

Exponential Decay with the Network Distance. The results of our quantitative network-wide analysis of these effects are summarized in Fig. 2 and Table 1. From Fig. 2, it can be concluded that the fraction of proteins with significantly affected free concentrations rapidly (exponentially) decays with the length L of the shortest path (network distance) from the perturbed protein. The same statement holds true for bound concentrations if the distance is measured as the shortest path from the perturbed protein to any of the two proteins forming a heterodimer. Thus, on average, the propagation of concentration changes along the PPI network indeed is considerably dampened. On the other hand, from Table 1 it can be concluded that the total number of multistep chains along which

[§]As an alternative to this computationally expensive approach we also tried the linear response matrix formalism (11), relating small changes in F_j to the ones in C_i. We found the linear response algorithm to be much less computationally expensive, although still providing remarkably good approximation to directly computed results even for large changes in protein levels.

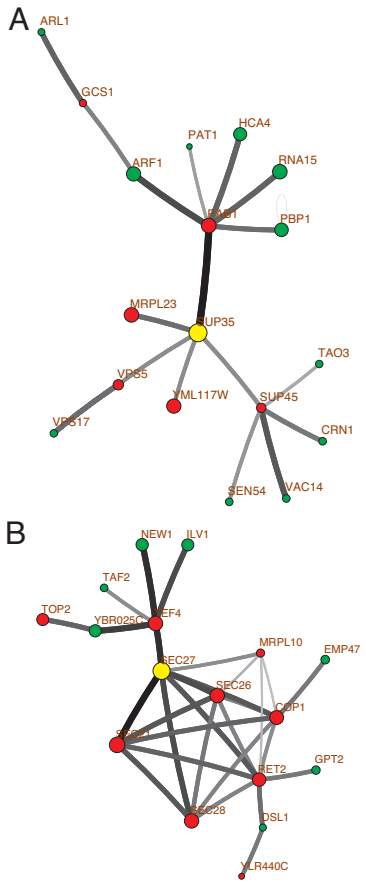

Fig. 1. Two cases of propagation of large concentration changes in the yeast protein-binding network. The total (bound and unbound) concentration of the protein marked with the yellow circle [the SUP35 protein (*A*); the SEC27 protein (*B*)] was increased 2-fold from its wild-type value in the rich growth medium (4). Red and green circles mark all other proteins whose equilibrium free (unbound) concentrations have increased (green) or decreased (red) by >20%. The area of each circle is proportional to the logarithm of the change in free concentration. Edges show all physical interactions among this group of proteins with the shade of gray proportional to the logarithm of the equilibrium concentration of the corresponding dimer calculated for K_{ij} = const = 10 nM.

concentration changes propagate with little attenuation remains significant for all but the largest values of the dissociation constant. These two observations do not contradict each other because the number of proteins separated by distance L (the last column in Table 1) rapidly grows with L.

Conditions Favoring the Multistep Propagation of Perturbations. What conditions favor the multistep propagation of perturbations along particular channels? In Fig. 3*A*, we show a group of highly abundant proteins along with all binding interactions between them. Then in Fig. 3*B*, we show only those interactions that according to our LMA calculation give rise to highly abundant heterodimers (equilibrium concentration >1,000 per cell), which breaks the densely interconnected subnetwork drawn in Fig. 3*A* into 10 mutually isolated clusters. Some of these clusters contain pronounced linear chains that serve as conduits for propagation of concentration perturbations. The fact that perturbations indeed tend to propagate via highly abundant heterodimers is illustrated in Fig. 3*C*, where arrows correspond to concentration-coupled nearest neighbors A → B. Evidently, the edges in Fig. 3 *B* and *C* largely (but not completely) coincide. Additionally, Fig. 3*C* defines the preferred direction of propagation of perturbations from a more abundant protein to its less abundant binding partners.

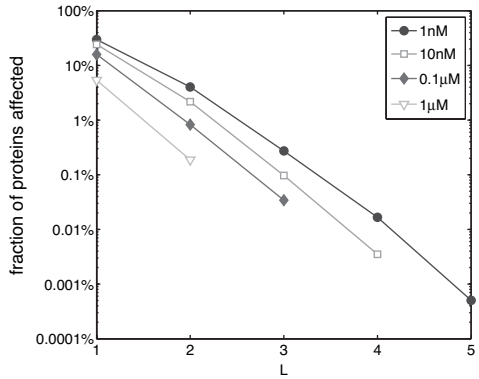

Fig. 2. Indiscriminate propagation of concentration perturbations is suppressed exponentially. The fraction of proteins with free concentrations affected by >20% among all proteins at network distance L from the perturbed protein. Different curves correspond to simulations with K_{ij} = const = 1 nM (filled circles), 10 nM (open squares), 0.1 μM (filled diamonds), and 1 μM (open triangles).

To further investigate what causes concentration changes to propagate along particular channels, we took a closer look at eight three-step chains $A \rightarrow A_1 \rightarrow A_2 \rightarrow B$ with the largest magnitude of perturbation of the last protein B (2-fold detection threshold after a 2-fold initial perturbation). The identification of intermediate proteins A_1 and A_2 was made by a simple optimization algorithm searching for the largest overall magnitude of intermediate perturbations along all possible paths connecting A and B.

Inspection of the parameters of these chains shown in Fig. 4 allows one to conjecture that for a successful transduction of concentration changes, the following conditions should be satisfied:

- Heterodimers along the whole path have to be of sufficiently high concentration D_{ij}.
- Intermediate proteins have to be highly sequestered. That is to say, to reduce buffering effects, free-to-total concentration ratios F_i/C_i should be sufficiently low for all but the last protein in the chain.
- Total concentrations C_i should decrease gradually in the direction of propagation. Thus, propagation of perturbations along virtually all of these long conduits is unidirectional and follows the gradient of concentration changes (a related concept of a "gradient network" was proposed for technological networks in ref. 10).
- Free concentrations F_i should alternate between relatively high and relatively low values in such a way that free concentrations of proteins at steps 2 and 4 have enough "room" to go down. The two apparent exceptions to this rule visible in Fig. 4 may be optimized to respond to a drop (instead of increase) in the level of the first protein.

Table 1. The number of concentration-coupled pairs of yeast proteins separated by network distance *L*

L	Var. 5 nM	1 nM	10 nM	0.1 μM	1 μM	All
1	2,003	2,469	1,915	1,184	387	8,168
2	415	1,195	653	206	71	29,880
3	15	159	49	8	0	87,772
4	2	60	19	0	0	228,026
5	0	3	0	0	0	396,608

Numerical simulations (2-fold initial perturbation, 20% detection threshold) were performed for different assignment of dissociation constants: K_{ij} = max(C_i, C_j) = 20 (column 2), K_{ij} = const = 1 nM, 10 nM, 0.1 μM, and 1μM (columns 3–6). Column 7 lists the total number of protein pairs at distance L.

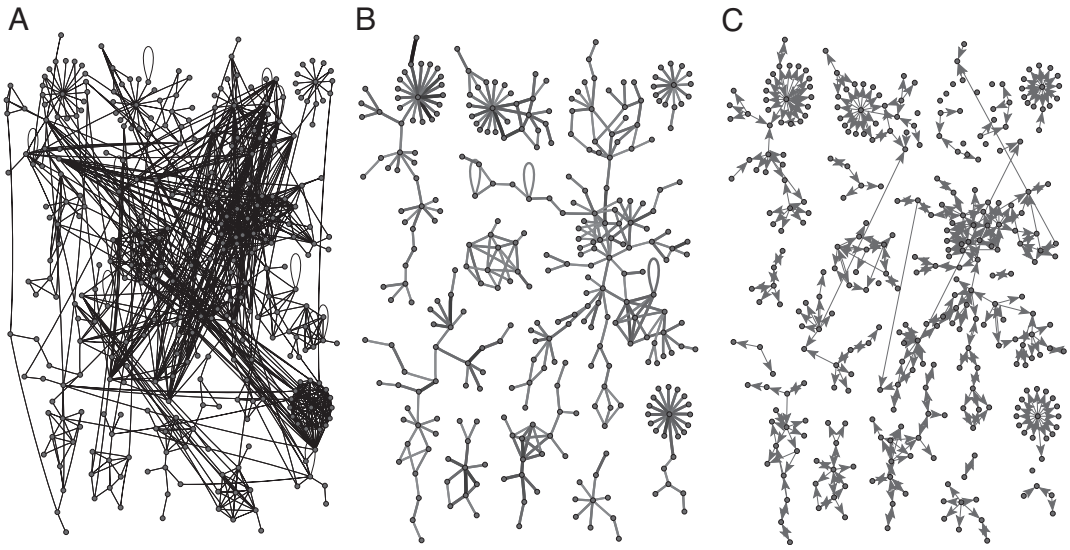

Fig. 3. Three views of a subset of 312 highly abundant nodes in a protein-binding network. (*A*) All binding links between these nodes. (*B*) Binding links characterized by high concentration of heterodimers (>1,000 molecules per cell). (*C*) Concentration-coupled proteins A → B with the property that a 2-fold increase in the abundance A reduces free concentration of its immediate binding partner B by 20% or more. Note that links roughly coincide with highly abundant dimers shown in *B*. Arrows reveal the preferential direction of propagation of perturbations.

These findings are in agreement with our more detailed numerical and analytical analysis of propagation of fluctuations presented in ref. 11 and illustrated for simple networks in *SI Appendix*. In ref. 11, we demonstrated that the linear response of the LMA equilibrium to small changes in protein abundances could be mapped approximately to a current flow in the resistor network in which heterodimer concentrations play the role of conductivities (which need to be large for a good transmission), whereas high F_i/C_i ratios result in the net loss of the perturbation "current" on such nodes and thus need to be minimized.

Discussion

Robustness with Respect to Assignment of Dissociation Constants. It often has been conjectured that the qualitative dynamical properties of biological networks to a large extent are determined by their topology rather than by quantitative parameters of individual interactions such as their kinetic or equilibrium constants (for a classic success story see, e.g., ref. 12). Our results generally support this conjecture, yet go one step further: we observe that the response of reversible protein–protein-binding networks to large changes in concentrations strongly depends not only on topology but also on abundances of participating proteins. Indeed, perturbations tend to preferentially propagate via paths in the network in which abundances of intermediate proteins monotonically decrease along the path (see Fig. 3). Thus, by varying protein abundances while strictly preserving the topology of the underlying network, one can select different conduits for propagation of perturbations.

On the other hand, our results indicate that these conduits are to a certain degree insensitive to the choice of dissociation constants. In particular, we found (see Fig. 5) that equilibrium concentrations of dimers and the remaining free (unbound) concentrations of individual proteins calculated for two different K_{ij} assignments [K_{ij} = const = 5 nM and $K_{ij} = \max(C_i, C_j)/20$ with the inverse mean of 5 nM] had a high Spearman rank correlation coefficient of 0.89 and even higher linear Pearson correlation coefficient of 0.98. The agreement was especially impressive in the upper part of the range of dimer concentrations (see Fig. 5). For example, the typical difference between dimer concentrations above 1,000 molecules per cell was measured to be as low as 40%. As we demonstrated above, it is exactly these highly abundant heterodimers that form the backbone for propagation of concentration perturbations. Thus, it should come as no surprise that sets of concentration-coupled protein pairs observed for different K_{ij} assignments also have a large (≈70–80%) overlap with each other (see SI Table 4).

Such degree of robustness with respect quantitative parameters of interactions can be explained partially by the following observation: proteins whose abundance is higher than the sum of abundances of all of their binding partners cannot be fully sequestered into heterodimers for any assignment of dissociation constants. As we argued above, such proteins with substantial unbound concentrations considerably dampen the propagation of perturbations and thus cannot participate in highly conductive chains. Another argument in favor of this apparent robustness is based on extreme heterogeneity of wild-type protein abundances (in the data set of ref. 4 they span 5 orders of magnitude). In this case, concentrations of heterodimers depend more on relative abundances of two constituent proteins than on the corresponding dissociation constant (within a certain range).

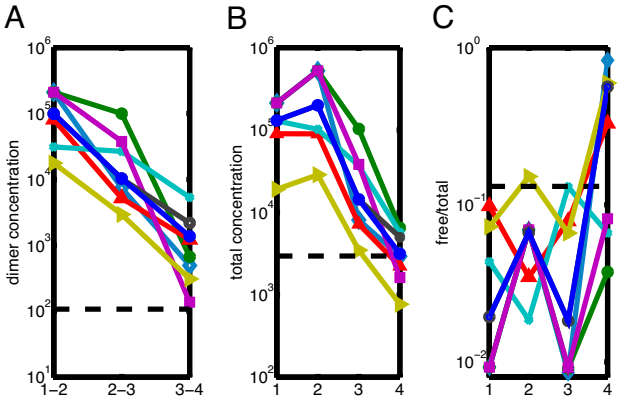

Fig. 4. Parameters of the eight three-step chains that exhibit the best transduction of concentration changes. Heterodimer concentrations D_{ij} (*A*) for three binding links along the chain. Total concentrations C_i (*B*) and free-to-total concentration ratios F_i/C_i (*C*) of the four proteins involved in these chains. Dashed lines correspond to network-wide geometric averages of the corresponding quantities: $\langle D_{ij}\rangle \approx 100$ copies per cell, $\langle C_i\rangle \approx 3{,}000$ copies per cell, and $\langle F_i/C_i\rangle = 13\%$.

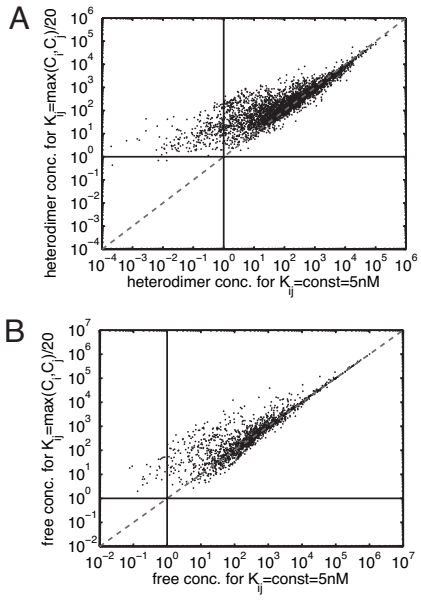

Fig. 5. The scatter plot of 4,185 bound concentrations D_{ij} (*A*) and 1,740 free concentrations F_i (*B*) calculated for two different assignments of dissociation constants to links in the PPI network. The *x* axis was computed for the homogeneous assignment K_{ij} = const = 5 nM, whereas the *y* axis was computed for the heterogeneous assignment $K_{ij} = \max(C_i, C_j)/20$ with the same average strength. The dashed lines along the diagonals are drawn at $x = y$, whereas the horizontal and vertical solid lines denote the concentration of one molecule per cell. Note that equilibrium concentrations in the upper part of their range (e.g., above 1,000 molecules per cell) are nearly independent of the choice of K_{ij}. Also, our choice of heterogeneous assignment nearly eliminates free or bound concentrations in a biologically unreasonable range <1 molecule per cell.

In a separate numerical control experiment, we verified that the main results of this study are not particularly sensitive to false positives and false negatives in the network topology inevitably present even in the best curated large-scale data. The percentage of concentration-coupled pairs surviving a random removal or addition of 20% of links in the network generally ranges between 60% and 80% (see SI Table 5).

Genetic Interactions. The effects of concentration perturbations discussed above could explain some of the genetic interactions between proteins. Consider for example a "dosage rescue" of a protein A by a protein B or the correction of an abnormal phenotype caused by deletion or other type of inactivation of A by overexpression of B. One possible mechanism behind this effect is that the knockout of A and overexpression of B affect the LMA equilibrium in opposite directions and to some extent cancel one another. In order for this mechanism to be applicable (albeit tentatively), concentrations of both A and B must be coupled simultaneously (in the sense used throughout this work) to at least one crucial protein *C* whose free or bound concentration has to be maintained at or close to wild-type levels. To assess this hypothesis, we analyzed the set of 772 dosage rescue pairs (3) involving proteins from the PPI network used in this study. For 136 pairs (or 18% of all dosage rescue pairs), we were able to identify one or more putative "rescued" protein whose free concentration was considerably (by >20%) affected by changes in abundances of both A and B (see SI Table 6). This overlap is highly statistically significant, having the Fisher's exact test *P* value of $\approx 10^{-216}$. Even more convincing evidence that perturbations to the LMA equilibrium state cause some of genetic interactions is presented in Fig. 6. It plots the fraction of protein pairs at distance *L* from each other in the PPI network that are known to dosage rescue each other. From Fig. 6, it can be concluded that proteins separated by distances 1, 2, and 3 are significantly more likely to genetically interact with each other than one expects by pure chance alone [the expected background level is marked with a dashed line ($772/1{,}740^2$) or, better yet, visible as a plateau for large values of *L*]. Furthermore, the slope of the exponential decay in the fraction of dosage rescue pairs as a function of *L* is roughly consistent with that shown in Fig. 2 for the fraction of concentration-coupled pairs.

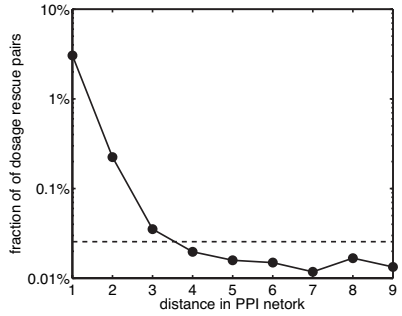

Fig. 6. The fraction of dosage rescue protein pairs separated by distance *L* in the PPI network. Note that pairs at distances 1, 2, and 3 are significantly overrepresented over the background level marked with dashed line or visible as a plateau at large distances *L*. The exponential decay constant at low values of *L* is consistent with that in Fig. 2.

Possibility of Functional Signaling and Regulation Mediated by Multistep Reversible Protein Interactions. Another intriguing possibility raised by our findings is that multistep chains of reversible protein–protein bindings in principle might be involved in meaningful intracellular signaling and regulation.

There are many well documented cases in which one-step "chains" are used to reversibly deactivate individual proteins by the virtue of sequestration with their binding partner(s). An example involving a longer regulatory chain of this type is the control of activity of condition-specific sigma factors in bacteria. In its biologically active state, a given sigma factor is bound to the RNA polymerase complex. Under normal conditions, it commonly is kept in an inactive form by the virtue of a strong binding with its specific anti-sigma factor (anti-sigma factors are reviewed in ref. 13). In several known cases, the concentration of the anti-sigma factor in turn is controlled by its binding with the specific anti-anti-sigma factor (13). The existence of such experimentally confirmed three-step regulatory chains in bacteria hints at the possibility that at least some of the longer conduits we detected in yeast could be used in a similar way.

Application to Microarray Data Analysis. To unequivocally detect cascading perturbations, in our simulations we always modified the total concentration of just one protein at a time. In more realistic situations, expression levels of a whole cluster of genes change, for example, in response to a shift in environmental conditions. Our general methods easily could be extended to incorporate this scenario. With the caveat that changes in expression levels of genes reflect changes in overall abundances of corresponding proteins, our algorithm allows one to calculate the impact of an external or internal stimulus measured in a microarray on free and bound concentrations of all proteins in the cell. Including such indirectly perturbed targets could considerably extend the list of proteins affected by a given shift in environmental conditions. Simultaneous shifts in expression levels of several genes may amplify changes of free concentrations of some proteins and/or mutually inhibit changes of others.

Effects of Intracellular Noise. Another implication of our findings is for intracellular noise, or small random changes in total concentrations C_i of a large number of proteins. The randomness, smaller

magnitude, and sheer number of the involved proteins characterize the differences between such noise and systematic severalfold changes in the total concentration of one or several proteins considered above. Our methods allow one to decompose the experimentally measured (9) noise in total abundances of proteins into biologically meaningful components (free concentrations and bound concentrations within individual protein complexes). Given a fairly small magnitude of fluctuations in protein abundances [on average ≈20% (9)], one could safely employ a computationally efficient linear response algorithm (see ref. 11). Several recent studies (9, 14, 15) distinguish between the so-called extrinsic and intrinsic noise. The extrinsic noise corresponds to synchronous or correlated shifts in abundance of multiple proteins, which, among other things, could be attributed to variation in cell sizes and their overall mRNA and protein production or degradation rates. Conversely, the intrinsic noise is attributable to stochastic fluctuations in production and degradation and thus lacks correlation between different proteins. We found that extrinsic and intrinsic noise affect equilibrium concentrations of proteins in profoundly different ways. In particular, although multiple sources of the extrinsic noise partially (yet not completely) cancel each other, intrinsic noise contributions from several sources sometimes can add up and cause considerable fluctuations in equilibrium free and bound concentrations of particular proteins (see SI Fig. 10).

Limitations of the Current Approach and Directions for Further Studies. In our study, we used a number of fundamental approximations and idealizations including the assumption of spatially uniform concentrations of proteins, the neglect of temporal dynamics, or, equivalently, the assumption that all concentrations have sufficient time to reach their equilibrium values, the continuum approximation neglecting the discrete nature of proteins and their bound complexes, etc. Another set of approximations was mostly attributable to the lack of reliable large-scale data quantifying these effects. They include not taking into account the effects of cooperative binding within multiprotein complexes, using a relatively small number (81) of well curated multiprotein complexes used in our study (see *SI Appendix*), neglecting systematic changes in protein abundances in the course of the cell cycle, etc. We do not expect these effects to significantly alter our main qualitative conclusions, namely, the exponential decay of the amplitude of changes in equilibrium concentrations, the existence of 3- to 4-step chains that nevertheless successfully propagate concentration changes, and the general conditions that enhance or inhibit such propagation.

In the future, we plan to extend our study of fluctuations in equilibrium concentrations by incorporating the effects of protein diffusion (nonuniform spatial concentration) and kinetic effects. Another interesting avenue for further research is to apply the concept of "potential energy landscape" (for definitions see ref. 16 and references therein) to reversible processes governed by the LMA, such as, for example, the equilibrium in protein-binding networks. In the past, this concept was applied to processes involving catalytic, irreversible protein–protein interactions such as, for example, phosphorylation by kinases or regulation by transcription factors. In this case, it helped to reveal the robustness of regulatory networks in the cell cycle (17) and in a simple two-protein toggle switch (18).

Methods

Source of Interaction and Concentration Data. The curated PPI network data used in our study is based on the 2.020 release of the BIOGRID database (3). We kept only pairs of physically interacting proteins that were reported in at least two publications using the following experimental techniques: affinity capture-MS (28,172 pairs), affinity capture-RNA (55 pairs), affinity capture-Western blotting (5,710 pairs), cocrystal structure (107 pairs), FRET (43 pairs), far Western blotting (41 pair), and two-hybrid (11,935 pairs). That left us with 5,798 nonredundant interacting pairs. Further restriction for both proteins to have experimentally measured total abundance (4) narrowed it down to 4,185 distinct interactions among 1,740 yeast proteins.

The list of manually curated yeast protein complexes was obtained from the latest release (May 2006) of the MIPS CYGD database (19, 20). This database contains 1,205 putative protein complexes, 326 of which are not coming from systemic analysis studies (high-throughput MS experiments). In the spirit of using only the confirmed PPI data, we limited our study to these curated complexes. For 99 of these complexes, the MIPS database lists three or more constituent proteins. After elimination of proteins with unknown total concentrations, we were left with 81 multiprotein complexes.

Genetic interactions of dosage rescue type also were obtained from the BIOGRID database. There are 772 pairs of dosage rescue interactions among 1,740 proteins participating in our PPI network (the full list contains 2,531 dosage rescue pairs).

Numerical Algorithms. The numerical algorithm calculating all free concentrations F_i given the set of total concentrations C_i and the matrix of dissociation constants K_{ij} was implemented in MATLAB 7.1 and is available from S.M. on request. It consists of iterating Eq. **1** starting with $F_i = C_i$. Iterations stop once relative change of free concentration on every node in the course of one iteration step becomes smaller than 10^{-8}, which for networks used in our study takes less than a minute on a desktop computer. When necessary, multiprotein complexes are incorporated into this algorithm as described in *SI Appendix*.

The effects of large concentration perturbations was calculated by recalculating free concentrations after a 2-fold increase in abundance of a given perturbed protein. The effects of small perturbations such as those of concentration fluctuations were calculated by using the faster linear response matrix formalism described elsewhere (11).

We thank Kim Sneppen for valuable discussions and contributions in early phases of this project. This work was supported by National Institute of General Medical Sciences Grant 1 R01 GM068954-01. Work at Brookhaven National Laboratory was carried out under Division of Material Science, U.S. Department of Energy Contract DE-AC02-98CH10886. S.M.'s visit to the Kavli Institute for Theoretical Physics, where part of this work was accomplished, was supported by National Science Foundation Grant PHY05-51164.

1. Wagner A (2001) *Mol Biol Evol* 18:1283–1292.
2. Jeong H, Mason SP, Barabasi AL, Oltvai ZN (2001) *Nature* 411:41–42.
3. Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M (2006) *Nucleic Acids Res* 34(Suppl 1):D535–D539.
4. Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK (2003) *Nature* 425:737–741.
5. Piehler J (2005) *Curr Opin Struct Biol* 15:4–14.
6. Kumar MD, Gromiha MM (2006) *Nucleic Acids Res* 34:D195–D198.
7. Lancet D, Sadovsky E, Seidemann E (1993) *Proc Natl Acad Sci USA* 90:3715–3719.
8. Deeds EJ, Ashenberg O, Shakhnovich EI (2006) *Proc Natl Acad Sci USA* 103:311–316.
9. Newman JRS, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, Weissman JS (2006) *Nature* 441:840–846.
10. Toroczkai Z, Bassler KE (2004) *Nature* 428:716.
11. Maslov S, Sneppen K, Ispolatov I (2006) arXiv:q-bio/0611026v1.
12. vonDassow G, Meir E, Munro EM, Odell GM (2000) *Nature* 406:188–192.
13. Hughes KT, Mathee K (1998) *Annu Rev Microbiol* 52:231–286.
14. Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) *Science* 297:1183–1186.
15. Raser JM, O'Shea EK (2005) *Science* 309:2010–2013.
16. Ao P (2004) *J Phys A Math Gen* 37:L25–L30.
17. Wang J, Huang B, Xia X, Sun Z (2006) *PLoS Comput Biol* 2:e147.
18. Kim K, Wang J (2007) *PLoS Comput Biol* 3:e60.
19. Guldener U, Munsterkotter M, Kastenmuller G, Strack N, van Helden J, Lemer C, Richelles J, Wodak SJ, Garcia-Martinez J, Perez-Ortin JE, *et al*. (2005) *Nucleic Acids Res* 33:D364–D368.
20. Mewes HW, Heumann K, Kaps A, Mayer K, Pfeiffer F, Stocker S, Frishman D (1999) *Nucleic Acids Res* 27:44–48.