Fluctuations in Mass-Action Equilibrium of Protein Binding Networks

Koon-Kiu Yan,[1,2] Dylan Walker,[1,2] and Sergei Maslov[2]

[1]*Department of Physics and Astronomy, Stony Brook University, Stony Brook, New York, 11794, USA*

[2]*Department of Condensed Matter Physics and Material Science, Brookhaven National Laboratory, Upton, New York 11973, USA*

(Received 2 April 2008; published 30 December 2008)

We consider two types of fluctuations in the mass-action equilibrium in protein binding networks. The first type is driven by slow changes in total concentrations of interacting proteins. The second type (spontaneous) is caused by quickly decaying thermodynamic deviations away from equilibrium. We investigate the effects of network connectivity on fluctuations by comparing them to scenarios in which the interacting pair is isolated from the network and analytically derives bounds on fluctuations. Collective effects are shown to sometimes lead to large amplification of spontaneous fluctuations. The strength of both types of fluctuations is positively correlated with the complex connectivity and negatively correlated with complex concentration. Our general findings are illustrated using a curated network of protein interactions and multiprotein complexes in baker's yeast, with empirical protein concentrations.

DOI: 10.1103/PhysRevLett.101.268102 PACS numbers: 87.18.Tt, 05.40.Ca, 87.15.hg, 87.15.km

The study of dynamical fluctuations in complex systems has emerged as a topic of intense interest germane to the fields of biology [1], financial systems [2], traffic in information [3] and transportation [4] networks, and many others. Of particular interest is the nature of collective effects that arise as a consequence of the connectivity of the underlying network. By examining such fluctuations, we can understand when the underlying network plays an important role and when, if possible, it may be ignored. A good candidate arena to study dynamical fluctuations is that of biomolecular processes taking place in cells.

Recently, propagation of biological fluctuations has been studied in the context of genetic regulation [5] and metabolic pathways [6]. These studies are primarily focused on small linear cascades of irreversible interactions. Conversely, we study the related problem of fluctuations in the mass-action equilibrium state of densely connected, reversible protein-protein-interaction (PPI) networks. These networks, in which proteins (nodes) are connected by edges if they bind together, exhibit nontrivial topological properties such as small-world effect, clustering, etc. Ourselves and others have studied the effect of large systematic changes in the levels of just one or a few proteins on the mass-action equilibrium of the entire PPI network [7,8]. Such changes are likely to occur as a consequence of regulated response of the cell, e.g., to changes in the external environment. For the same system, however, there is another type of perturbation that is both different and of significant interest: intracellular noise or small fluctuations in equilibrium (bound and free) concentrations of many proteins. The randomness, smaller magnitude, and sheer number of involved proteins characterize the difference between fluctuations that are the subject of this study and the systematic large changes in the total abundance of single proteins that were the subject of Ref. [8]. These fluctuations come in two varieties. *Spontaneous* fluctuations in equilibrium concentrations occur at constant protein copy numbers and are caused by the intrinsically stochastic nature of binding interactions. These fluctuations are small and change rapidly relative to the characteristic time of changes in copy numbers of individual proteins. They are well described using the machinery of equilibrium statistical physics. In contrast, *driven* fluctuations are induced by changes in proteins' copy numbers due to the stochastic nature of their production and degradation as well as variation in activity of regulators. These driven fluctuations are usually somewhat larger than the spontaneous noise. They also change relatively slowly on time scales (tens of minutes) that are large compared to relaxation times of equilibrium concentrations which are rarely slower than seconds.

To illustrate general principles with a concrete example, in this study, we used a curated genome-wide network of PPI in baker's yeast (*S. cerevisiae*), which, according to the BIOGRID database [9] were independently confirmed in at least two published experiments. We combined this network with a genome-wide data set of protein abundances [10] and the set of curated multiprotein complexes from the MIPS database [11]. After keeping only the interactions between proteins with known concentrations, we were left with 4085 interacting pairs and 81 multiprotein complexes involving 2004 proteins. The same network was previously used by us and others in [7,8]. Another assumption (justified in these earlier studies) is that in the absence of large-scale experimental data on the strength of protein-protein interactions, we use a set of evolutionary-motivated [8] dissociation constants $K_{ij} = \max(C_i, C_j)/20$ for all interactions in our network. The evolutionary-motivated *association* is defined to be the weakest association strength necessary to keep a sizable fraction of the rate-limiting protein in a given interacting pair bound in the dimer. The denominator 20 is chosen to reproduce the average asso-

ciation strength, $\langle 1/K_{ij}\rangle = 1/(5\ \mathrm{n}M)$, in a set of experimentally measured dissociation constants from the PINT database [12], which are assumed to be representative for all biologically functional interactions among yeast proteins. Occasional deviations from the evolutionary-motivated scheme (e.g., due to specific functional requirements) would not undermine our final results since, as was demonstrated in Ref. [8], equilibrium concentrations of proteins and their complexes are not very sensitive to assignment of dissociation constants to individual interactions given their overall strength.

In what follows, we derive a general analytical framework to quantify fluctuations in equilibrium concentrations of two-protein complexes (dimers). To simplify the resulting formulae, we disclude homodimers as well as multiprotein complexes though our formalism can be easily modified to take them into account. We will also assume reactions are occurring in a unit volume, in order to suppress the system volume V in the equations that follow. At any point in time, the system is fully described by $\{C_i\}$, the number of copies of individual proteins, $\{D^*_{ij}\}$ and $\{F^*_i\}$, the sets of integer numbers representing instantaneous abundances of dimers and monomers correspondingly. These three sets of variables are in fact not independent but constrained by mass conservation.

To study *spontaneous* fluctuations, we consider the case where all total copy numbers C_i are held constant over time and variations in free and dimer copy number are driven solely by thermal fluctuations. To this end, we write the free energy for a network of interacting dimers (see, e.g., [13] 3.16, for derivation and discussion):

$$G = \sum_{(ij)\epsilon E} \{\epsilon_{ij} D^*_{ij} + k_B T D^*_{ij} \log(D^*_{ij}/e)\} + k_B T \sum_{i=1}^{N} F^*_i \log(F^*_i/e) \quad (1)$$

where ϵ_{ij} are dimer binding energies. The first sum runs over all E edges (dimers) and the second sum runs over all nodes (proteins) in the network. Free (monomer) copy numbers F^*_i in this expression are not independent variables but rather a shorthand for $C_i - \sum_m D^*_{im}$ derived from mass conservation. The above expression does not include volume-dependent entropy and kinetic terms that we have suppressed as they are not relevant to our discussion here. The requirement of zero first derivative of the free energy with respect to dimer copy number gives the Law of Mass Action (LMA) that relates equilibrium free ($F_i = \langle F^*_i\rangle$) and bound ($D_{ij} = \langle D^*_{ij}\rangle$) concentrations in the system via $D_{ij} = F_i F_j / K_{ij}$, where $K_{ij} = K^{(0)} \exp(-\epsilon_{ij}/k_B T)$ is the dissociation constant and the unstarred variables indicate equilibrium concentrations. The second derivative of the free energy with respect to dimer concentration yields the generalized susceptibility and, in accordance with the Fluctuation-Dissipation Theorem (FDT) [14], the relative magnitude η of *spontaneous* fluctuations $\langle \delta D^2_{ij}\rangle = \langle (D^*_{ij} - D_{ij})^2\rangle$ as

$$\eta \equiv \frac{\langle \delta D^2_{ij}\rangle}{D_{ij}} = (\Gamma^{-1})_{(ij)(ij)} \quad (2)$$

where as derived in [15], the matrix $\hat{\Gamma}$ is

$$\Gamma_{(ij)(km)} = \frac{D_{ij}}{k_B T} \frac{\partial^2 G}{\partial D_{ij} D_{mk}} = \delta_{ik} D_{ij}/F_i + \delta_{jm} D_{ij}/F_j + \delta_{ik}\delta_{jm}.$$

A direct consequence of this result is that spontaneous fluctuations for a dimer linked to the rest of the network involve contributions from other dimers, through nondiagonal elements of Γ contributing to its inverse. To address the impact of propagation of fluctuations over the network, it seems natural to compare the noise of a dimer embedded in the network to the noise for an isolated dimer (isol-F) with the same equilibrium concentrations F_i, F_j, and D_{ij}. Such an isolated dimer corresponds to a matrix Γ that is diagonal so that its inverse is simply

$$\eta^{\text{isol-}F} = [\Gamma_{(ij)(ij)}]^{-1} = \left[\frac{D_{ij}}{F_i} + \frac{D_{ij}}{F_j} + 1\right]^{-1}. \quad (3)$$

Furthermore, $\eta > \eta^{\text{isol-}F}$ is easily shown from convexity arguments [16]. Clearly then, collective effects act to amplify thermal fluctuations. This is related to propagation of static perturbations, studied in [7], as fluctuations from neighboring dimers contribute to a dimer's own noise. We define the amplification factor for a dimer (ij)

$$R = \eta/\eta^{\text{isol-}F}. \quad (4)$$

A cumulative histogram of amplification factors for the PPI network of baker's yeast is examined in Fig. 1. Relative to the isolated case, collective effects can lead to thermal noise that is significantly amplified, as is evident from this histogram.

Collective amplification of thermal noise presents a worrisome theoretical possibility. Can amplification occur without limit? To address this question, it is fruitful to develop an alternative formalism in which the magnitude of fluctuations are calculated directly from the partition function

$$Z(\{C_r\}) = \sum_{\{D^*_{ij}\}} N_S(\{D^*_{ij}\}) \exp\left(-\sum_{i<j} \frac{\epsilon_{ij} D^*_{ij}}{k_B T}\right) \quad (5)$$

where the sum is taken over all possible (integer) copy numbers of individual dimers defining the ''occupation state'' $\{D^*_{ij}\}$. Above the combinatorial factor, $N_S(\{D^*_{ij}\})$ counts the number of microstates of individual labeled proteins resulting in a given occupation state $\{D^*_{ij}\}$. For example, for an isolated dimer (ij), one has

$$N_S(D^*_{ij}) = \binom{C_i}{D^*_{ij}} \binom{C_j}{D^*_{ij}} D^*_{ij}!. \quad (6)$$

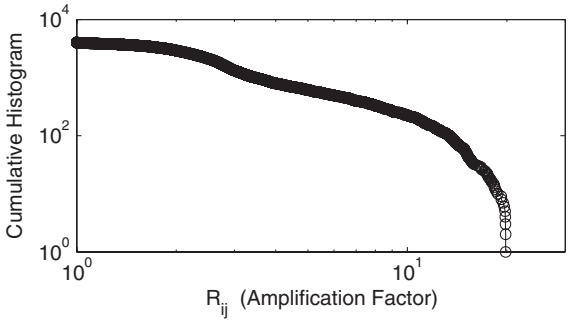

FIG. 1. Cumulative histogram of amplification factors for spontaneous (thermal) noise of equilibrium dimer concentrations D_{ij} in the PPI network of yeast. Collective effects lead to amplification relative to the isolated dimer null model.

In fact, the free energy expression of Eq. (1) can be derived from Eq. (5) using Stirling's approximation for factorials in $N_S(\{D^*_{ij}\})$. Using Eq. (5), it is straightforward to show, by a change of variables, that higher moments of D^*_{ij} can be related to the lower moments evaluated at a reduced system size. Indeed, in calculation of $\langle D^*_{ij}\rangle$, the combinatorial factor containing $C_i!C_j!/D^*_{ij}!$ becomes $C_i!C_j!/(D^*_{ij}-1)! = C_iC_j(C_i-1)!(C_j-1)!/(D^*_{ij}-1)!$. As a result, one has the following *exact* equality:

$$\langle D^*_{ij}\rangle|_{C_i,C_j} = C_iC_j\frac{Z(C_i-1,C_j-1)}{Z(C_i,C_j)}. \tag{7}$$

Here, for the sake of brevity, we omitted the concentrations other than C_i and C_j as parameters of the statistical sum $Z(\{C_k\})$. A similar expression for a higher moment

$$\langle D^*_{ij}(D^*_{ij}-1)\rangle = C_i(C_i-1)C_j(C_j-1) \times \frac{Z(C_i-2,C_j-2)}{Z(C_i,C_j)}$$

may be rewritten as

$$\langle D^*_{ij}(D^*_{ij}-1)\rangle = \langle D^*_{ij}\rangle|_{C_i,C_j}\langle D^*_{ij}\rangle|_{C_i-1,C_j-1}$$

where the latter moment is evaluated in a system for which the copy number of proteins i and j (C_i and C_j) are reduced by exactly one. It follows that apart from Eq. (2), the noise may be alternatively expressed as

$$\eta = 1 + \langle D^*_{ij}\rangle|_{C_i-1,C_j-1} - \langle D^*_{ij}\rangle|_{C_i,C_j}. \tag{8}$$

The above expression for thermal noise hints at an intimate connection between the dynamic and static perturbations of the mass-action equilibrium. This connection can be made even more explicit by expanding the 2nd term to first order in total concentration

$$\eta \simeq 1 - D_{ij}[(\Lambda^{-1})_{ii} + (\Lambda^{-1})_{jj} + 2(\Lambda^{-1})_{ij}] \tag{9}$$

where the matrix

$$\Lambda_{km} = \frac{\partial C_k}{\partial \log F_m} = D_{km} + C_k\delta_{km}$$

characterizes the response of equilibrium concentrations F_m to small static changes in total concentrations C_k [7]. It should be remarked that, despite the approximation used in Eq. (9), this approach is in good agreement with the FDT formalism first introduced. One notes that this expression for noise explicitly depends only on the total and dimer concentrations used to define the matrix Λ. This suggests the definition of a new isolated model (isol-C), consisting of an isolated (ij) dimer formed by proteins with the same C_i, C_j, and D_{ij}. This is only possible through changes in the dissociation constant and free concentrations of constituent proteins i and j. It is important to mention that this model is distinct from the isol-F benchmark defined earlier, in which each isolated dimer has the same equilibrium free and dimer concentrations (yet different C_i and C_j) as the corresponding dimer in the network. For an isol-C dimer, the matrix Λ is 2×2 and trivially invertible. The noise is given by

$$\eta^{\text{isol-}C} = \left(\frac{D_{ij}}{C_i - D_{ij}} + \frac{D_{ij}}{C_j - D_{ij}} + 1\right)^{-1}. \tag{10}$$

A comparison with Eq. (3) for the isol-F model reveals that in the isol-C model, the protein i has an equilibrium free concentration $\tilde{F}_i = C_i - D_{ij} = F_i + \sum_k D_{ik}$ and similarly for protein j. Thus, while the isol-F model completely ignores the effect of neighboring dimers, the isol-C model actually amplifies their contribution to fluctuations. Indeed, in the isol-C model, the concentrations of these dimers are effectively absorbed into the renormalized free concentrations $\tilde{F}_i$, $\tilde{F}_i$ and thus brought one step closer to dimer (ij). Consequently, the noise of a dimer in the isol-C model always exceeds the noise of a corresponding dimer in the real network. The real noise for a dimer in a network falls somewhere between the bounds of these two isolated dimer scenarios. A summary of the lower- and upper-bound models and their noise amplitudes is given in Fig. 2.

Now, we turn our attention to the second type of noise *driven* by stochastic changes in protein copy number C_i. In the living cell, these fluctuations are typically larger than the spontaneous fluctuations. Furthermore, the changes in C_i occur at a relatively slow time scale (tens of minutes) so that the mass-action equilibrium quasistatically responds to these changes. From the definition of Λ, it follows [7,8] that an arbitrary number of small perturbations δC_m add up to

$$\frac{\delta F_i}{F_i} = \sum_m (\Lambda^{-1})_{im}\delta C_m. \tag{11}$$

Because of bilinear dependence of D_{ij} on F_i and F_j, one also has

$$\frac{\delta D_{ij}}{D_{ij}} = \frac{\delta F_i}{F_i} + \frac{\delta F_j}{F_j}. \tag{12}$$

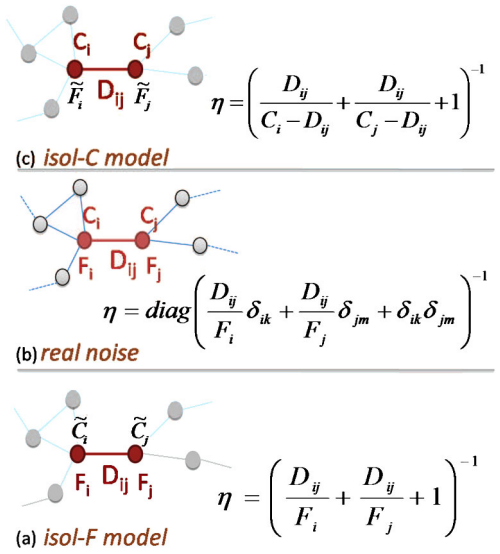

FIG. 2 (color online). A comparison of the noise in a network dimer to two isolated dimer models defined in the text. The real noise is bound below and above by the isolated models $\eta^{\text{isol-}F} \leq \eta \leq \eta^{\text{isol-}C}$.

Thus, in general, the amplitude of driven fluctuations is given by

$$\frac{\langle \delta D_{ij}^2 \rangle}{D_{ij}} = D_{ij} \left\langle \left[\sum_k (\Lambda^{-1})_{ik} \delta C_k + \sum_m (\Lambda^{-1})_{jm} \delta C_m \right]^2 \right\rangle. \tag{13}$$

The evaluation of the above expression requires the full matrix of cross-correlations $\langle \delta C_k \delta C_m \rangle$ which is currently experimentally unknown. For the simplest case $\langle \delta C_k \delta C_m \rangle \propto C_k^2 \delta_{mk}$ of uncorrelated fluctuations (so-called intrinsic noise [1]), the driven response becomes

$$\left(\frac{\langle \delta D_{ij}^2 \rangle}{D_{ij}} \right)_{\text{int}} \propto D_{ij} \sum_k [(\Lambda^{-1})_{ik} + (\Lambda^{-1})_{jk}]^2 C_k^2. \tag{14}$$

In conclusion, we study how the two types of noise (spontaneous and driven) relate to simple properties of a dimer such as its abundance and connectivity (number of connections it has to the rest of network). With high statistical significance, we find that the relative amplitude ($\sqrt{\langle \delta D_{ij}^2 \rangle}/D_{ij}$) of both spontaneous and driven (intrinsic) noise is negatively correlated with dimer abundance D_{ij} (Spearman coefficient of $r = -0.98$, $r = -0.45$, respectively). While in the former case, a strong correlation is expected from the law of large numbers, the significant correlation in the latter case of driven noise is a nontrivial result. Furthermore, we found that relative amplitude of both spontaneous and driven (intrinsic) noise are positively correlated with dimer connectivity ($r = 0.42$, $r = 0.33$). These results are consistent with the overall scenario that we investigated above in which any type of noise propagates throughout the network and where the existence of network connections (both direct and, to some extent, indirect) to noisy partners positively contribute to fluctuations of individual dimers. One should also note that the above two types of correlations represent two independent results since their signs are opposite to what one expects based on a weak ($r = 0.15$) positive correlation between dimer connectivity and its abundance.

Work at Brookhaven National Laboratory was carried out under Contract No. DE-AC02-98CH10886, Division of Material Science, U.S. Department of Energy.

[1] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Science **297**, 1183 (2002).
[2] L. Laloux, P. Cizeau, J. Bouchaud, and M. Potters, Phys. Rev. Lett. **83**, 1467 (1999).
[3] M. Takayusa, H. Takayusa, and T. Sato, Physica A (Amsterdam) **233**, 824 (1996).
[4] K. Nagel and M. Paczuski, Phys. Rev. E **51**, 2909 (1995).
[5] J. M. Pedraza, and A. van Oudenaarden, Science **307**, 1965 (2005).
[6] E. Levine and T. Hwa, Proc. Natl. Acad. Sci. U.S.A. **104**, 9224 (2007).
[7] S. Maslov, K. Sneppen, and I. Ispolatov, New J. Phys. **9**, 273 (2007).
[8] S. Maslov and I. Ispolatov, Proc. Natl. Acad. Sci. U.S.A. **104**, 13655 (2007).
[9] C. Stark *et al.*, Nucleic Acids Res. **34**, D535 (2006).
[10] S. Ghaemmaghami *et al.*, Nature (London) **425**, 737 (2003).
[11] U. Guldener *et al.*, Nucleic Acids Res. **33**, D364 (2004).
[12] M. D. Kumar and M. M. Gromiha, Nucleic Acids Res. **34**, D195 (2006).
[13] B. Widom, *Statistical Mechanics: A Concise Introduction for Chemists* (Cambridge Univ. Press, New York, 2002).
[14] H. B. Callen and T. A. Welton, Phys. Rev. **83**, 34 (1951).
[15] I. Prigogine, Physica (Amsterdam) **16**, 134 (1950).
[16] For brevity, assume the notation $\mu = (ij)$, $\nu = (mk)$. The matrix Γ is symmetrized by the diagonal matrix $Q_{\mu\nu} = \sqrt{D_\mu}\delta_{\mu\nu}$ and is diagonalized by a unitary transformation U so that: $\Gamma = QU\Gamma_D U^{-1} Q^{-1}$. From these considerations and the convexity of $f(x) = x^{-1}$, it can be shown that $\Gamma^{-1}_{\mu\mu} \geq (\Gamma_{\mu\mu})^{-1}$, where $\Gamma^{-1}_{\mu\mu} = \sum_\alpha U^2_{\mu\alpha}(\Gamma_D)^{-1}_{\alpha\alpha}$ and $(\Gamma_{\mu\mu})^{-1} = [\sum_\alpha U^2_{\mu\alpha}(\Gamma_D)_{\alpha\alpha}]^{-1}$.