PHYSICAL REVIEW E **77**, 020901(R) (2008)

Statistically enhanced promiscuity of structurally correlated patterns

D. B. Lukatsky* and E. I. Shakhnovich
Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA
(Received 13 October 2007; published 1 February 2008)

We predict that patterns with correlated surface density of atoms have statistically higher promiscuity (ability to bind stronger to an arbitrary pattern) as compared with noncorrelated patterns with the same average surface density. We suggest that this constitutes a generic design principle for highly connected proteins (hubs) in protein interaction networks. We develop an analytical theory for this effect. We show that our key predictions are generic and independent, qualitatively, on the specific form of the interatomic interaction potential, provided it has a finite range.

DOI: 10.1103/PhysRevE.77.020901 PACS number(s): 87.15.A−

Any living cell is densely crowded with proteins. The average density of proteins in a cell is strikingly high. For example, in a typical bacterial cell, *E. Coli*, with a volume of 1 μm^3 there are approximately 4×10^6 proteins at any moment of time [1]. With an average protein size being 2 nm, this translates to a volume fraction of about 0.02, which is a high fraction! A key, open question of molecular biology is how proteins maintain the necessary level of specificity at such high protein concentrations. What are the generic principles that evolution utilizes in designing protein surfaces that provide the specificity? Conversely, are there generic design signatures that make proteins more *promiscuous* (i.e., able to interact with multiple partners)? It has been recognized rather recently that understanding protein promiscuity is a key step towards understanding the evolvability and plasticity of protein-protein interaction (PPI) networks [2]. Another central issue concerns the competition between promiscuity and robustness in PPI networks [3,4]. Motivated by these questions, here we predict one generic, structural design principle responsible for the enhanced promiscuity of proteinlike and other types of patterned surfaces. In particular, we predict that patterns with the enhanced atomic surface density correlations are more promiscuous than noncorrelated patterns (the entirely random pattern is an example of a noncorrelated pattern).

We suggest that such promiscuous proteins play a role of "hubs" in PPI networks (i.e., proteins with many interacting partners; see, e.g., Refs. [5,6] for a review on hub proteins in the human and yeast PPI). We will provide a more precise definition of "promiscuity" below.

We begin by introducing a simple, coarse-grained model of random [7,8] and "designed" proteinlike surfaces. A random surface is obtained by distributing N particles at random on a planar, circular surface with the surface area A, Fig. 1. The average density of particles is thus $\phi_0=N/A$. Particles are not allowed to overlap, as they have the hard-core diameter d_0. After the surface pattern is generated, the particles are fixed and not allowed to move. A designed surface is obtained using the following Monte Carlo (MC) design procedure. First, we generate a random surface pattern, as described above. Second, we allow particles to anneal at a given design temperature T_d. To this end, we impose that particles within the surface interact through the pairwise design potential, $U_d(\rho)$, where ρ is the distance between two particles within the surface. For a given surface configuration, the total interaction energy of the pattern, E_d, is

$$E_d=\frac{1}{2}\int \varphi_d(\vec{\rho})U_d(|\vec{\rho}-\vec{\rho}'|)\varphi_d(\vec{\rho}')d^2\vec{\rho}d^2\vec{\rho}', \qquad (1)$$

where $\varphi_d(\vec{\rho})=\phi_0+\phi_d(\vec{\rho})$ is the local pattern density at the point $\vec{\rho}\equiv(x,y)$, with $\phi_d(\vec{\rho})$ being the deviation (fluctuation) of the local density from its average value ϕ_0. The only

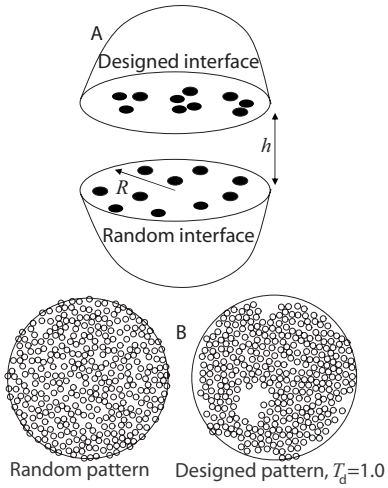

FIG. 1. (A) Schematic representation of a model interacting pair of proteins, where one protein interface is random and another one is designed. Both interfaces have identical *average* surface density of particles (amino acids). Qualitatively, a designed interface contains dense clusters (patches) of amino acids. (B) Snapshots of random and designed patterns as obtained from simulations. The designed pattern is obtained at the design temperature, $T_d=1$. There are 350 particles in each snapshot. The diameter of the surface is $D=140$ Å, the particle size is $d_0=5$ Å. The average surface fraction of particles in each surface pattern is thus $\tilde{\phi}_0=Nd_0^2/D^2\simeq0.45$. The MC design procedure performs 10^4 MC sweeps (for each designed pattern) at a given design temperature.

*Present address: Department of Molecular Genetics, Weizmann Institute of Science, 76100 Rehovot, Israel; and Department of Chemistry, Ben-Gurion University of the Negev, Beer Sheva, 84105 Israel.

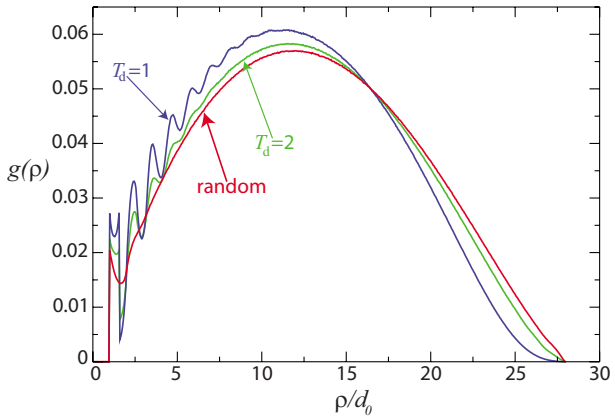

FIG. 2. (Color online) Computed density correlation function $g(\rho)$ for random and designed surface patterns. $g(\rho)$ is normalized in such a way that $\int_0^D g(\rho)d\rho=1$. The MC design procedure performs 10^4 MC sweeps at a given design temperature for each designed pattern. Each curve is obtained as the ensemble average over 10^4 patterns. The error bars are not shown, as they are comparable with the linewidth.

assumption about the potential is that it is a pairwise additive, finite-range potential characterized by the interaction length scale ξ.

We note that in the analytical analysis described below, we assume that the potential $U_d(\rho)$ in Eq. (1) is a continuous potential, and correspondingly, the density function $\varphi_d(\rho)$ is also continuous. To speed up the numerical simulations, however, we use hard-core particles with the additional, square-well potential (see below).

We next perform the standard MC annealing procedure [9] with the Boltzmann weight, $\sim\exp(-E_d/T_d)$, where T_d is measured in units of k_BT, where k_B is the Boltzmann constant and T is the absolute temperature. The MC procedure is stopped after a certain number of MC moves, and the resulting annealed configuration is accepted as the final, designed configuration. An example of one such designed configuration is shown in Fig. 1. In this example the adopted design potential $U_d(\rho)$ is an attractive, square-well potential: $U_d(\rho)=-2k_BT$, if 5 Å$<\rho\leqslant$8 Å, and $U_d(\rho)=0$, if $\rho>8$ Å, and particles have the hard-core diameter of 5 Å.

Quantitatively, the density correlation function, $g(\rho)\sim\langle\varphi_d(0)\varphi_d(\rho)\rangle$, characterizes how strongly the surface is designed, where the averaging here is performed with respect to different realizations of patterns at a given design temperature. Apparently, $g(\rho)$ is proportional to the probability density to find a particle separated by the distance ρ from a randomly selected particle. Figure 2 shows the computed $g(\rho)$ for random and designed surfaces. The lower the design temperature T_d, the higher is the degree of the particle ordering (correlation) and clusterization on the surface. The averaging is performed over the ensemble of surface patterns generated at each given design temperature, Fig. 2.

Our first task is to analyze the probability distribution for the interaction energies $P(E)$, and the corresponding extreme value distribution (EVD), $P_{\min}(E)$, for pairs of interacting surface patterns. Every such pair consists of a random pattern and a designed pattern, respectively, Fig. 1. The EVD is the probability distribution of the *minimum* intersurface energies with respect to mutual rotation of surfaces. This probability distribution is practically the most important one, as proteins bind approximately according to this distribution.

We begin by computing the properties of $P(E)$. The interaction energy E between the random surface and the designed surface (in a random mutual orientation between the two surfaces) is

$$E=\int \varphi_d(\vec{\rho})U(|\vec{r}-\vec{r}'|)\varphi(\vec{\rho}')d^2\vec{\rho}d^2\vec{\rho}', \tag{2}$$

where $\varphi(\vec{\rho})=\phi_0+\phi(\vec{\rho})$ is the density of particles in the random surface pattern, and $\phi(\vec{\rho})$ is the deviation of the density from its average value ϕ_0; and $\varphi_d(\vec{\rho}')$ is the corresponding density in the designed pattern [it is defined after Eq. (1)]. Here we assume for the sake of simplicity that random and designed patterns have the same average density ϕ_0. $U(|\vec{r}-\vec{r}'|)$ is the intersurface interaction potential (it does not have to coincide with the design potential U_d), where $\vec{r}=(\vec{\rho},-h/2)$ and $\vec{r}'=(\vec{\rho}',+h/2)$, with h being the intersurface separation. We aim to compute the probability distribution $P(E)$ which is the Gaussian probability distribution entirely specified by its mean $\langle E\rangle$ and by the dispersion $\sigma^2=\langle(E-\langle E\rangle)^2\rangle$. The probability distribution for the density fluctuations in the designed patterns has the form

$$P_d[\phi_d(\vec{\rho})]=C_1\exp\left[-\int \phi_d^2(\vec{\rho})d^2\vec{\rho}/2\phi_0\right]\exp(-E_d/T_d), \tag{3}$$

where E_d is given by Eq. (1), and C_1 is the normalization constant. The first exponential in Eq. (3) takes into account the configurational entropy of patterns, and the second exponential describes the contribution to the intrapattern energy due to the local density fluctuations. The corresponding probability distribution for random patterns has only the entropic contribution [7], $P[\phi(\vec{\rho})]=C_2\exp[-\int\phi^2(\vec{\rho})d^2\vec{\rho}/2\phi_0]$. We note that this entropic contribution can be derived by expanding the ideal gas entropy to the second order in the density fluctuations, $\phi(\rho)$.

Using these probability distributions we can now perform the ensemble averaging of the interpattern interaction energy, Eq. (2), in the Fourier space [7]. The resulting dispersion σ, of the energy fluctuations has the form

$$\sigma^2=4\phi_0A\int|\hat{U}(\vec{q})|^2\frac{1}{[1/\phi_0+\hat{U}_d(\vec{q})/T_d]}\frac{d^2\vec{q}}{(2\pi)^2}, \tag{4}$$

where $\hat{U}_d(\vec{q})=\int U_d(\rho)e^{-i\vec{q}\vec{\rho}}d^2\vec{\rho}$, and $\hat{U}(\vec{q})$ is defined analogously. We note that in the limit of the vanishing magnitude of the design potential, $\hat{U}_d(\vec{q})/T_d=0$, Eq. (4) takes the correct form of the dispersion for *random* heterodimers [7], $\sigma^2_{r,\text{hetero}}$. Our key, qualitative result, following from the analysis of Eq. (4), is that the *lower* the design temperature T_d in the ensemble of designed patterns, the *larger* is the dispersion σ. This conclusion holds true for any *attractive* design potential $U_d(\rho)$. We stress that the latter result is robust with respect to the sign of the intersurface potential $U(r)$. Enhanced density

correlations within the designed surface patterns are responsible for the predicted effect. In particular, due to the presence of correlated, inhomogeneous clusters within designed surface patterns, it is more probable to find both the lower and the higher energy states upon the interaction with random patterns as compared with the case of the interaction between two entirely random patterns. We note that the singularity in the integrand of Eq. (4) indicates the onset of the order-disorder phase transition at sufficiently high strength of the attractive design potential. Our Gaussian fluctuation theory, however, is breaking down in the vicinity of this transition, and higher-order terms in density fluctuations must be considered to account for the phase transition properly.

Adopting a simple form for both $U(r)=U_0\exp(-r^2/\xi^2)$ and $U_d(\rho)=U_{0d}\exp(-\rho^2/\xi^2)$, where $r^2=\rho^2+h^2$, we obtain the following scaling relationship in the limit $\hat{U}_d(\vec{q})/T_d \ll 1$:

$$\sigma^2 - \sigma^2_{r,\text{hetero}} \sim -|U_0|^2 \frac{U_{d0}}{T_d}\phi_0^3\xi^4 A. \tag{5}$$

For any attractive design potential, $U_{d0}<0$, and therefore $\sigma^2>\sigma^2_{r,\text{hetero}}$. This implies that the EVD, $P_{\min}(E)$, for interacting surface pairs in which one pattern is designed, Fig. 1, is always shifted towards lower energies as compared with the EVD for pairs consisting of both entirely random patterns. Intuitively, this means that designed patterns are statistically more promiscuous than random patterns.

We emphasize that not only the dispersion σ, but also the average energy $\langle E\rangle$, depends on T_d. The lower T_d is, the lower is $\langle E\rangle$ for any attractive U_d. This is due to the following finite-size effect (this result does not follow from the straightforward averaging of Eq. (2) in the thermodynamic limit). As a result of the design procedure, any finite, designed pattern has a higher local density than a random pattern; this is visually seen in Fig. 1. As the average surface density ϕ_0 is fixed, designed patterns are shrunk within the finite surface area as compared with random patterns, Fig. 1. It is easy to derive the scaling relationship for the difference between the average energies $\langle E\rangle$ in the limits of very low (extremely correlated patterns) and very high (entirely random patterns) design temperatures:

$$\langle E(T_d \ll 1)\rangle - \langle E(T_d \gg 1)\rangle \sim U_0\phi_0^2\xi^2 A, \tag{6}$$

where we again assumed the Gaussian form of the intersurface interaction potential $U(r)$. It is interesting to note that this result is sensitive to the sign (and not only to the amplitude) of $U(r)$. While for any attractive $U(r)$, the lower T_d is, the lower is $\langle E\rangle$; for a repulsive $U(r)$ the trend would be the opposite.

The numerical results illustrating the principles described above are presented in Fig. 3. Both probability distributions, $P(E)$ and $P_{\min}(E)$, are computed in Fig. 3 for different values of T_d. In order to compute $P_{\min}(E)$ numerically, we rotate each pair of surfaces about their common axis, until the lowest possible intersurface energy for this pair is found.

The key, qualitative conclusion following from the above analysis is that designed proteinlike patterns *always* have a larger number of interacting partners in a PPI network as compared with random patterns. The magnitude of the energy shift ΔE between the corresponding $P_{\min}(E)$ in Fig. 3 provides a quantitative estimate for the strength of the predicted effect. In particular, $\Delta E=\langle E_{\min}(T_d=1)\rangle - \langle E_{\min}(\text{random})\rangle \simeq -0.6$ kT, per one interface residue. For a typical protein interface with $N_i=50$ interface residues, this translates to a large ratio of the concentrations, $n_{d-r}/n_{r-r}=\exp(-\Delta E N_i/kT)\simeq 20$, where n_{d-r} and n_{r-r} are the concentrations of the interacting (designed-random) and (random-random) protein pairs, respectively. Even at considerably higher design temperature (corresponding to a much weaker design strength), $T_d=2$, this ratio is still as large as 4.5! The key message here is that the PPI network connectivity (i.e., the average number of interacting partners) for a designed protein is about an order of magnitude larger than for a random protein even for moderate design strength. In other words, structural design via enhancing surface density correlations is a generic design principle for hub proteins in PPI networks. The latter result constitutes our key finding in this paper.

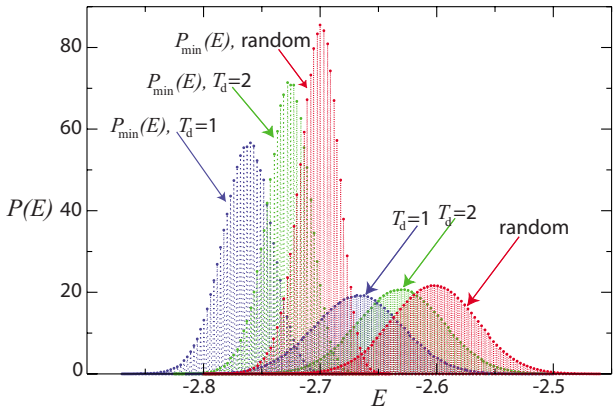

FIG. 3. (Color online) Computed probability distributions for the interaction energies within the pairs of interacting random and designed surface patterns, $P(E)$, and the extreme value distributions (EVD) $P_{min}(E)$, respectively, at different values of the design temperature T_d. Each generated pair of patterns consists of one entirely random pattern and one designed pattern (see the caption of Fig. 1 for the numerical surface parameters). The lower is the design temperature, the lower is the corresponding average energies, and the larger is the dispersion of the corresponding probability distributions. This is compared with the corresponding probability distributions for pairs of entirely random patterns. The separation h between the surfaces within each interacting pair is $h=5.01$ Å. The intersurface interaction potential $U(r)$ was chosen to be identical to the design potential $U_d(\rho)$ specified in the main text. The energy is plotted in the units of T_d and normalized per one particle.

Our final result is the prediction that hub-hub interactions are always statistically *stronger* than interactions between pairs consisting of a hub and a random protein [described by Eq. (4)]. This finding may help to resolve an open and controversial experimental question about the strength of hub-hub interactions in PPI networks of different organisms [6].

In this case *both* patterns within each interacting pair are designed at a given design temperature T_d. Following the steps described earlier, and averaging E with the probability

distribution, Eq. (3), we obtain the following dispersion $\sigma_{d,\text{hetero}}$ for the case of designed heterodimers (i.e., two *distinct*, designed, interacting patterns):

$$\sigma_{d,\text{hetero}}^2 = 4A \int |\hat{U}(\vec{q})|^2 \frac{1}{[1/\phi_0 + \hat{U}_d(q)/T_d]^2} \frac{d^2\vec{q}}{(2\pi)^2}. \quad (7)$$

The principal message here is the fact that $\sigma_{d,\text{hetero}} > \sigma$, where σ is given by Eq. (4). This implies that interacting pairs of two designed patterns are statistically more strongly bound than pairs consisting of a random pattern and a designed pattern.

We note, finally, that the universal principle of statistically enhanced self-interaction of *random* patterns predicted earlier [7] holds true for the case of *designed* patterns, as well. In particular, the pairs of randomly superimposed, *identical* (i.e., homodimeric) designed patterns have always twice as large magnitude of the energy fluctuations with respect to their mutual orientation, as compared with pairs of *distinct* (heterodimeric) designed patterns $\sigma_{d,\text{homo}}^2/\sigma_{d,\text{hetero}}^2 = 2$, where $\sigma_{d,\text{homo}}^2$ is the corresponding dispersion for homodimers. The latter law implies that the EVD of the interaction energies for designed homodimers is always shifted towards lower energies as compared with the EVD for designed heterodimers. Therefore in the PPI network consisting of random and designed proteins, the attractive hub-hub interactions (both homo- and heterodimeric) are statistically stronger than interactions between random proteins and hubs, and of course, stronger than interactions between pairs of entirely random proteins.

In summary, our key result is the prediction that protein-like surface patterns with enhanced correlations of the density are, statistically, more promiscuous as compared with entirely random patterns (with the same size and the same average surface density). In particular, we suggest that such highly promiscuous proteins play a role of hubs in PPI networks. The key, experimental verification of our prediction requires the computation of the density correlation functions $g(\rho)$ at real protein interfaces. Such a key test, using crystallographic PPI structural data, is currently under way.

We are grateful to B. Shakhnovich, K. Zeldovich, D. Tawfik, G. Schreiber, and P. M. Kim for helpful discussions.

[1] U. Alon, *An Introduction to Systems Biology* (Chapman & Hall/CRC, Taylor & Francis Group, London, 2007).
[2] O. Khersonsky, C. Roodveldt, and D. S. Tawfik, Curr. Opin. Chem. Biol. **10**, 498 (2006).
[3] S. Maslov and I. Ispolatov, Proc. Natl. Acad. Sci. U.S.A. **104**, 13655 (2007).
[4] E. J. Deeds *et al.*, Proc. Natl. Acad. Sci. U.S.A. **104**, 14952 (2007).
[5] J. F. Rual *et al.*, Nature (London) **437**, 1173 (2005).
[6] N. N. Batada, L. D. Hurst, and M. Tyers, PLOS Comput. Biol. **2**, e88 (2006).
[7] D. B. Lukatsky, K. B. Zeldovich, and E. I. Shakhnovich, Phys. Rev. Lett. **97**, 178101 (2006).
[8] S. Panyukov and Y. Rabin, Phys. Rev. E **56**, 7053 (1997).
[9] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Academic Press, San Diego, 2002).