Proteinlike behavior of a spin system near the transition between a ferromagnet and a spin glass

Chai-Yu Lin,[1] Chin-Kun Hu,[1] and Ulrich H. E. Hansmann[2]
[1]*Institute of Physics, Academia Sinica, Nankang, Taipei 11529, Taiwan*
[2]*Department of Physics, Michigan Technological University, Houghton, Michigan 49931-1295*
(Received 14 February 2001; revised manuscript received 8 August 2001; published 25 October 2001)

A simple spin system is studied as an analog for proteins. We investigate how the introduction of randomness and frustration into the system affects the designability and stability of ground-state configurations. We observe that the spin system exhibits proteinlike behavior in the vicinity of the transition between a ferromagnet and a spin glass. Our results illuminate some guiding principles in protein evolution.

DOI: 10.1103/PhysRevE.64.052903 PACS number(s): 87.10.+e, 05.50.+q, 75.10.Hk, 87.15.−v

The folding of a protein into a specific three-dimensional (3D) biologically active structure is now often described by the funnel concept [1]. It is assumed that the energy landscape of a protein is rugged but with a sufficient overall slope towards the native structure [2]. Folding occurs by multipathway kinetics and the particulars of the folding funnel determine the transitions between the different thermodynamic states [2,3]. While originally derived from studies of minimalistic protein models, evidence for the validity of the funnel concept was subsequently presented for real proteins [4].

A funnellike energy landscape guarantees thermodynamic stability and kinetic accessibility for the biologically active structure of proteins. Both are necessary conditions for proteins to perform their biological functions. Hence, the funnel concept suggests that the optimal state of a protein is one of minimal frustration [5]. This is because a smoother energy landscape and a steeper slope leads to faster folding and greater stability. However, proteins are in general only marginally stable [6], and both stability and speed of folding can often be increased in protein engineering [7]. Hence, it appears that the sequence of amino acids in a protein is in general not optimized for the smoothness of its energy landscape. The question arises then on why is this the case and why are proteins only marginally stable. Or, what factors constraint the amount of frustration (and the ruggedness of the funnel landscape) in the evolution of proteins?

When studying the above questions, one encounters the problem that the amount of frustration is difficult to control in protein models. For this reason, we propose to use the frustrated 3D-Ising model on a simple cubic lattice [8] with periodic boundary conditions as an analogy of proteins, and to study the above questions for this much simpler system in which the frustration may be easily measured. Unlike in earlier work [9,10], we interpolate continuously between the ferromagnet and the spin glass by varying the density of antiferromagnetic bonds. Our choice of the system is motivated by the observation that proteins are similar to spin glasses in that their energy landscape is characterized by a huge number of local minima separated by high-energy barriers [11]. On the other hand, the global funnellike topology of protein energy landscapes, leading to a unique ground state, resembles more a ferromagnet. Hence, it seems that proteins show behavior between that of a ferromagnet and a spin glass. However, the limitations of the analogy between the frustrated Ising model and proteins should be kept in mind. Spin systems do not fold, only the process by which the system finds its ground state may be regarded as analogous to folding. We can study only how, for Ising models, this process depends on the frustration and under which conditions there are similarities to proteins.

Our model is described by the Hamiltonian

$$H = -\sum_{\langle lm \rangle}^{3N} J_{lm}\sigma_l\sigma_m, \tag{1}$$

where the sum goes over all $3N$ (N the number of spins) pairs $\langle lm \rangle$ of nearest-neighbor spins $\sigma_i = \pm 1$. A certain number M of randomly chosen bond variables, J_{lm}, are set to $J_{lm} = -1$ while the remaining $3N - M$ bonds are assigned the value $J_{lm} = 1$. The ratio $R = M/3N$ is a measure for the randomness in our Ising system and leads to the frustration in the systems that is, as usual, defined through

$$F = \frac{1}{3N}\sum_{\square_i} \delta(F_{\square_i}, -1) \quad \text{with} \quad F_{\square_i} = J_{12}J_{23}J_{34}J_{14}. \tag{2}$$

Here, $J_{12}, J_{23}, J_{34}, J_{14}$ are the four bond variables of the ith elementary plaquette $\square_i$ of the lattice, and the sum goes over all $3N$ elementary plaquettes.

Our simulations are done on a $4\times4\times4$ lattice, which is small enough that simulated annealing will find the ground state. An even smaller lattice size may have allowed exhaustive enumeration, but would have introduced severe finite-size effects. For a given value F of frustration, 2000 realizations of bond variables $\{J_{lm}\}$ are generated in random. For each realization, N_1 simulated annealing runs are used to search for the global minimum. In each run, we cool down the system with step size $\Delta T = 0.1$ from temperature $T = 3$ to $T = 0.3$ performing 40 Monte Carlo sweeps (one update for each spin) at each temperature. We define as ground-state C_g of one realization the configuration with the lowest-energy E_g obtained in the N_1 runs. To ensure reasonable statistics, we require that this energy is found in at least N_2 simulated annealing runs. The total number N_1 of runs is adjusted accordingly and the failure rate $N_F = (N_1 - N_2)/N_1$ defines an index for the difficulty to find the global minimum. In the

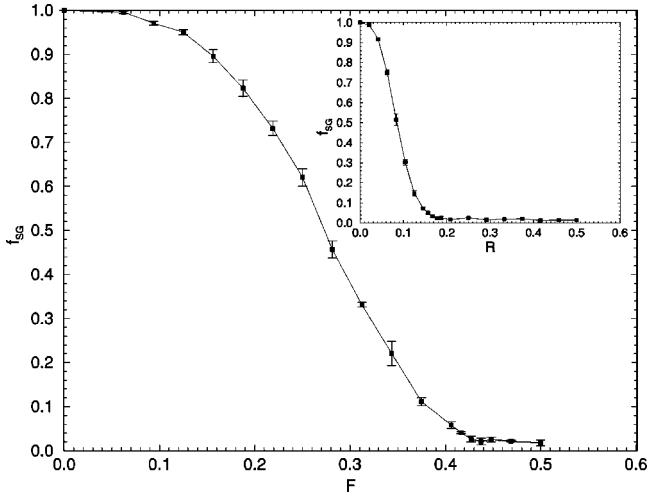

FIG. 1. The frequency $f_{SG}=N_{SG}/N_T$ of realizations with a single ground state as a function of F and (inset) R.

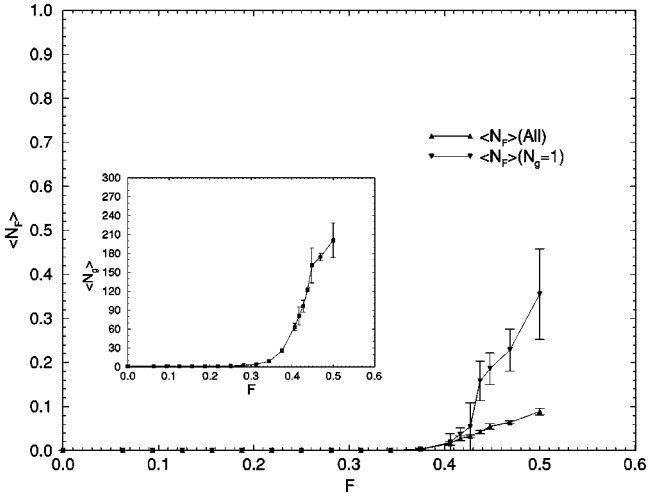

FIG. 2. The average failure rate $\langle N_F\rangle$ as a function of F. In the inset, we display the average number $\langle N_g\rangle$ of ground states as a function of F.

next step, we check the N_2 ground-state configurations for rotational and translational symmetries, and identify in this way the number N_g of *distinct* ground-state configurations found for the given realization. For small values of R we set $N_2=10\,000$. As the system approaches the spin glass, N_g increases rapidly. Therefore, if $N_g>1000$, we repeat the simulation with $N_2=100\,000$ to obtain more accurate values for N_g.

By altering the frustration F, we can tune our system between a ferromagnet ($F=0$) and a spin glass ($\langle F\rangle_{av}=0.5$) and investigate the relation between F in the system and the occurrence of proteinlike behavior. Since the native state of a protein is unique and commonly assumed to be its ground state, we define a realization $\{J_{lm}\}$ as proteinlike if it has a single ground state. The number of protein-like realization $\{J_{lm}\}$ among 2000 realizations is denoted by N_{SG}. We display the frequency $f_{SG}=N_{SG}/2000$ of such realizations as a function of F in Fig. 1, which shows that f_{SG} decreases with growing F and is almost constant for $F\geqslant 0.44$. The inset of Fig. 1 shows the same quantity as a function of R and here flattening occurs for $R\geqslant 0.23$. Hence, the probability to find protein-like realizations decreases as a function of F (or R). However, the total number of realizations is given by $N_{Realiz.}=(3N)!/[[3N(1-R)]!(3NR)!]$, i.e., grows much faster with increasing R. It follows that the total number of proteinlike realizations that may be designed (a randomly chosen realization has vanishing small probability for a single ground state) is also an *increasing* function of F since the bond randomness R and the average frustration over realizations $\langle F\rangle_{av}$ are related through $\langle F(R)\rangle_{av}=4[(1-R)^3R+(1-R)R^3]$ [8].

We know that with growing F, the energy landscape becomes more and more rugged. The number of local minima separated by high-energy barriers will grow, and the probability will increase that our simulated annealing runs get trapped in one of them and do not find the global minimum. This may be seen in Fig. 2 where we display the average failure rate $\langle N_F\rangle$ as a function of F for the case of all 2000 samples and for only these realizations with single ground-state $N_g=1$. In this plot, we observe a steep increase of $\langle N_F\rangle$ at $F_g=0.44\pm 0.02$ for the curve corresponding to the ''all samples'' case. Note that this value corresponds to $R_g=0.23\pm 0.02$, which is consistent with that for the transition between a ferromagnetic and spin-glass order found in [12]. The transition between the ferromagnet and the spin glass may also be observed in the average number of ground states per realization $\langle N_g\rangle$ as a function of F that we display in the inset of Fig. 2. The location of the steep increase in this quantity, $F_g=0.44\pm 0.02$ (which corresponds to $R_g=0.23\pm 0.02$), is the same as for the failure rate and agrees with the point in [12].

The failure rate N_F in Fig. 2 measures how often a simulation did *not* find the ground state and is therefore related to the ''folding time,'' i.e., the time that would be necessary to find the ground state in a simulation. The ''folding time'' itself is a measure for the kinetic accessibility of the ground states. For the frustrated Ising model, we see from Fig. 2 that the failure rate (and consequently the ''folding time'') is small for small F and differs little from the time needed for the ferromagnet $F=0$. This changes once we reach values of F where the system behaves as a spin glass. At that point, the failure rate and the ''folding time'' increases by orders of magnitude, and even for realizations $\{J_{lm}\}$ that have a single ground state that state may no longer be kinetically accessible. Such a situation is not desirable for real proteins, which have only limited time to fold, and therefore, must have kinetically accessible native states. Hence, we may not assume that realizations $\{J_{lm}\}$ with $F\geqslant 0.44\pm 0.02$ are proteinlike even if they have a unique ground state. If the analogy between proteins and spin systems holds, then we may expect for proteins also an interplay between the increasing entropy of sequences, which lead to unique ground-state structures, and the requirement that this state has to be kinetically accessible. On one hand, the entropy of sequences increases with frustration, while on the other hand the folding times become prohibitively large once the frustration exceeds a certain value. In the Ising model, the transition to this

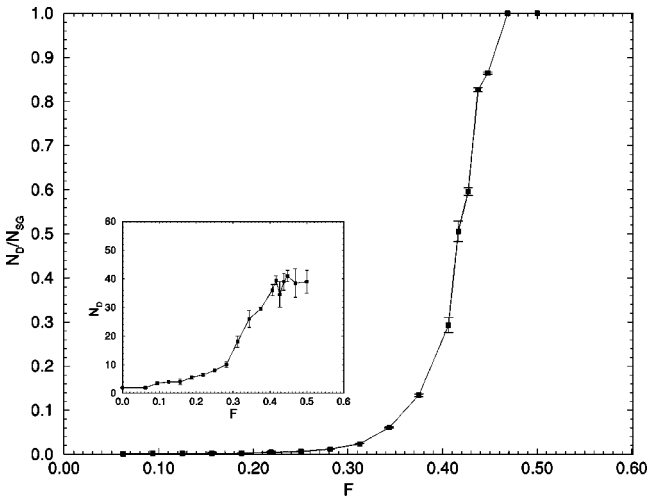

FIG. 3. The ratio N_D/N_{SG} as a function of F. In the inset, we show the number N_D of *truly different* single ground-state configurations, as a function of F.

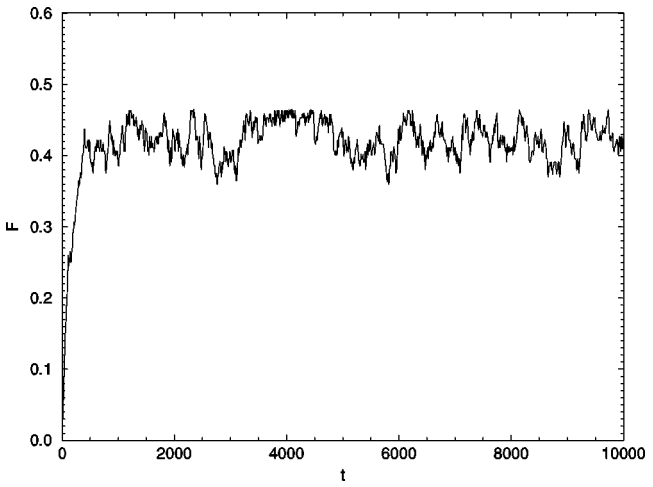

FIG. 4. Time series of the bond randomness F from a dynamic simulation described in the text.

spin-glass behavior is pronounced and located at $F_g=0.44\pm0.02$ ($R_g=0.23\pm0.02$). The above conjecture may explain why proteins are marginally stable: the entropy of marginally stable proteins is much higher than that of sequences optimized for thermodynamic stability and fast folding. However, a limiting minimal amount of thermodynamic stability is necessary to guarantee function of the protein.

The above conjecture implies that the ''optimal'' amount of frustration in proteins is where the system is ''almost'' at the point of becoming a spin glass. This is because in such a case, the entropy of sequences that lead to a single *and* accessible ground state is maximal. However, a protein should also be stable in the sense that a mutation will not lead to an amino sequence with a *different* native structure or no unique ground state at all. Hence, such protein structures are preferred that can be realized by a maximal number of different amino acid sequences [13]. In the language of our spin system the above statement implies that these spin configurations are most proteinlike that are single ground state for the largest number of realizations $\{J_{lm}\}$. For this reason, we have further checked the N_{SG} proteinlike ground-state configurations on translational and rotational symmetries. This procedure leads to a much smaller number N_D of distinct single ground-state configurations. N_D is displayed as a function of F in the inset of Fig. 3. N_D is an increasing function over the whole ferromagnetic range and more or less constant in the spin-glass range. Hence, with increasing value of F, not only the total number of proteinlike realizations grows but also the variety of proteinlike states.

From the inset of Fig. 3, we would expect that the situation in proteins would correspond to small values of frustration F in the Ising model where one single ground-state configuration dominates, which may be realized by many sets of bond variables $\{J_{lm}\}$. However, proteins have to change over the course of evolution. The requirement of evolutionary flexibility suggests that larger values of randomness and frustration should be preferred that increase the number of distinct ground-state structures and enhance the chance that a mutation will lead from one structure to a different one. Hence, we expect for proteins an interplay between the requirement that the native structure is stable under mutations, and the need for structural changes over the course of evolution.

In order to study this interplay, we plot in Fig. 3 the ratio N_D/N_{SG}. Note that this ratio corresponds to the inverse of the (averaged) ''designability'' [13] and is a measure for the degeneracy of the various proteinlike states (i.e., spin configurations that are unique ground states for some realizations $\{J_{lm}\}$) of our spin system. We see that this ratio has a steplike behavior at $F_p=0.41\pm0.02$ (which corresponds to $R_p=0.17\pm0.02$). For smaller values of F, the N_D types of ground-state configurations are realized by many sets $\{J_{lm}\}$, while for larger values of F, each spin configuration is realized by only one realization $\{J_{lm}\}$. Hence, we conclude that in our spin system, the ''optimal'' frustration is at F_p where both a variety of different proteinlike configurations may be realized, but at the same time, these structures may be designed by more than one set of $\{J_{lm}\}$, and therefore are stable under mutations [14]. Note that this point is close to, but smaller than, the glass transition point ($F_g=0.44\pm0.02$). Our value of F_p also corresponds to the point where in Fig. 1, the failure rate of realizations with single ground state diverges from the corresponding plot for *all* realizations: $F=0.41\pm0.03$.

The above results suggest that in proteinlike systems, randomness and frustration is necessary to increase the designability of proteins. In our spin system, the absolute number of realizations with a single ground state will increase with frustration. On the other hand, once the frustration exceeds a certain value, the system becomes a spin glass. The resulting rugged energy landscape implies now that the single ground state, if existing, is no longer kinetically accessible. This would be biologically not desirable, and the frustration in proteins has to be below this critical value. In a similar fashion, the evolutionarily favorable increase in diversity of protein-like states with frustration is counteracted by the growing probability that a given configuration becomes unstable under mutations. If the frustration exceed a certain

value, any mutation would lead to a different structure that is again biologically not desirable. We conjecture that proteins are not minimally frustrated but that in proteinlike systems, the competition between these factors leads to a maximal value of F where the number of different kinetically accessible structures, which can be realized as single ground states by *many* sequences, is largest. For our spin system, this point is $F_p = 0.41 \pm 0.02$, which is close to, but below the point $F_g = 0.44 \pm 0.02$, where the system starts to behave as a spin glass.

In order to demonstrate how the interplay of the above-outlined factors may lead to an optimal value of F, we have made up the following game. Our starting point is the ferromagnet, i.e., $J_{lm} = 1$. The game consists of a series of Monte Carlo steps that simulate ''evolution.'' At each Monte Carlo step, our system has two offspring before it dies. One of the offspring is a copy of the parent, the other carries a mutation. We simulate mutations by chosing at random one-bond variable J_{lm} and switching its sign. Only one of the offspring is allowed to survive, and the survival rate of the ''mutant'' is given by $P(F_N)/[P(F_N)+P(F_0)]$. Here, F_N and F_0 are the frustration of the ''mutant'' and the ''unchanged system,'' respectively, with $P(F) = f_{SG}(F)[1 - \langle N_F(F) \rangle][1 - N_D(F)/N_{SG}(F)]$, where $f_{SG}(F)$, $\langle N_F(F) \rangle$, and N_D/N_{SG} are taken from our previous simulations and $\langle N_F(F) \rangle$ corresponds to the curve $\langle N_F(F) \rangle$ with $N_g = 1$ in Fig. 1. With these rules, our system performs a random walk in F shown in Fig. 4. The average value of F throughout this random walk gives $F = 0.42 \pm 0.03$, which is consistent with $F_p = 0.41 \pm 0.02$ and supports our assumption that the evolution of proteinlike systems leads to a optimal point of F in the system.

In summary, we have studied the simple frustrated Ising model as an analog for proteins. Investigating this system as a function of frustration, we found that the spin system exhibits proteinlike behavior at or slightly below the point at which a system changes from an ordered (ferromagnet) to a random system (spin glass). Whether this observation (which questions the common belief that proteins are minimally frustrated systems) holds for realistic protein models remains to be investigated. As a next step in this direction, we have started simulations of a bond-diluted and site-diluted frustrated Ising model. In such a model, it may be possible to generate more realistic proteinlike structures with backbone and side chains.

We thank Jonathan Dushoff and C. Tang for a critical reading of the paper. U.H. acknowledges support by a research grant (Grant No. CHE-9981874) of the National Science Foundation. This work is also supported by the National Science Council of the Republic of China (Taiwan) under Grant No. 89-2112-M001-084.

[1] J.N. Onuchic, Z. Luthey-Schulten, P.G. Wolynes, Annu. Rev. Phys. Chem. **48**, 545 (1997).
[2] K.A. Dill and H.S. Chan, Nat. Struct. Biol. **4**, 10 (1997).
[3] J.D. Bryngelson and P.G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **84**, 7524 (1987).
[4] U.H.E. Hansmann, M. Masuya, and Y. Okamoto, Proc. Natl. Acad. Sci. U.S.A. **94**, 10 652 (1997); U.H.E. Hansmann, Y. Okamoto, and J.N. Onuchic, Proteins **34**, 472 (1999).
[5] N. Go, Annu. Rev. Biophys. Bioeng. **12**, 183 (1983).
[6] P.L. Privalov and S.J. Gill, Adv. Protein Chem. **39**, 191 (1988).
[7] See, for instance, S.M. Malakauskas and S.L. Mayo, Nat. Struct. Biol. **5**, 470 (1998).
[8] S. Kirkpatrick, Phys. Rev. B **16**, 4630 (1977).
[9] P. Garstecki, T.X. Hoang, and M. Cieplak, Phys. Rev. E **60**, 3219 (1999).
[10] M. Cieplak, T.X. Hoang, and M.S. Li, Phys. Rev. Lett. **83**, 1684 (1999).
[11] C. Dasgupta, S.K. Ma, and C.-K. Hu, Phys. Rev. B **20**, 3837 (1979).
[12] A.K. Hartmann, Phys. Rev. B **59**, 3617 (1999).
[13] H. Li, R. Helling, C. Tang, and N. Wingreen, Science **273**, 666 (1996).
[14] We have tested this by changing a randomly chosen bond J_{lm} in realizations with a single ground state. At the transition point, 70% of the ''mutations'' have the same ground state. This number reduces to less than 40% in the spin-glass phase. Realizations with multiple ground states are less stable under mutations. In the spin-glass phase, only $\approx$ 3% of the mutations have the same set of ground states. This observed chaotic behavior of multiple ground-state realizations is consistent with the results observed by A.J. Bray and M.A. Moore, Phys. Rev. Lett. **58**, 57 (1987).