Neutrality and evolvability of designed protein sequences

Arnab Bhattacherjee and Parbati Biswas*
Department of Chemistry, University of Delhi, Delhi 110007, India

(Received 17 December 2009; revised manuscript received 25 March 2010; published 12 July 2010)

The effect of foldability on protein's evolvability is analyzed by a two-prong approach consisting of a self-consistent mean-field theory and Monte Carlo simulations. Theory and simulation models representing protein sequences with binary patterning of amino acid residues compatible with a particular foldability criteria are used. This generalized foldability criterion is derived using the high temperature cumulant expansion approximating the free energy of folding. The effect of cumulative point mutations on these designed proteins is studied under neutral condition. The robustness, protein's ability to tolerate random point mutations is determined with a selective pressure of stability ($\Delta\Delta G$) for the theory designed sequences, which are found to be more robust than that of Monte Carlo and mean-field-biased Monte Carlo generated sequences. The results show that this foldability criterion selects viable protein sequences more effectively compared to the Monte Carlo method, which has a marked effect on how the selective pressure shapes the evolutionary sequence space. These observations may impact *de novo* sequence design and its applications in protein engineering.

DOI: 10.1103/PhysRevE.82.011906 PACS number(s): 87.15.−v, 87.23.−n, 02.70.−c

I. INTRODUCTION

In-silico insights provide invaluable inputs for comprehending the mechanism of molecular evolution based on the requirements of fitness, stability and function. These models, both on-lattice, off-lattice [1] and spin glass models [2] are often employed to compute suitable measures of foldability, explore the nature of folding energy landscape [3] and how the need to fold can affect the course of molecular evolution [4,5]. Understanding the evolution of proteins based on some optimized foldability provides valuable insight about the mechanism of the selective pressure at the molecular level. Evolution proceeds via mutations and most often these mutations are found to confer marginal or no functional advantage to the protein. This conforms to the theory of neutral evolution [6,7] which implies a set of structurally similar mutated sequences without any significant alteration in their function.

Different aspects of evolution may be simulated, ranging from the scale of the protein/gene to the entire genome to characterize the mapping between sequences and structures. Such models usually define a set of "viable" mutated sequences which is a subset of the larger space of all possible sequences. This set of sequences are connected by single point mutations and encode the same native fold, often termed as neutral network [8,9]. Each mutated sequence in this network act as a node. For an evolving set of sequences, the population dynamics may be viewed as random diffusion [10] over the network. Basic simulation models focus on understanding the variation of the stability of designed sequences only through random point mutations [11] and correlates how the rate of mutations [12–14] affect the protein's functionality and stability. The stability of the designed sequences toward point mutations have been investigated [15] which point out that large energy gaps between the native state and an ensemble of unfolded states makes these sequences more robust toward random point mutations [16,17].

In this work, a self-consistent mean-field theory with binary patterning of amino acid residues is applied to design "wild-type" protein sequences consistent with a generalized foldability criterion ϕ. Selective pressure is explicitly included in the model in terms of $\Delta\Delta G$. The detailed effect of random cumulative mutations is investigated on wild-type protein sequences obtained from theory and simulations. Also the robustness of proteins are assessed in terms of the cumulative (m)-neutrality, which is the fraction of proteins which fold to the wild-type structure but differ from the wild-type sequence exactly at m residues [18,19]. Our theory is based on a three-dimensional lattice model representation of protein structure. Despite the approximations involved, they portray an unique sequence-structure relationship which is observed in real proteins [20,21]. The theory is compared with classic Monte Carlo (MC) and mean-field bias Monte Carlo (MFBMC) results on model lattice proteins. The mutational robustness of the sequences obtained from the theory depict a similar qualitative trend compared to that of the sequences obtained from the simulation results. However, both the results show that foldability (ϕ) of sequences is an important parameter to govern the viable genotypes at a given evolutionary selective pressure.

II. THEORY

Foldability criteria provide a suitable measure for the sequence-structure compatibility in form of a predetermined energy function. The protein is modeled as a cubic lattice polymer consisting of 27 residues on a maximally compact three-dimensional lattice [22–24]. Protein conformations are represented by self-avoiding walks that occupy all lattice vertices and a total of 103 346 compact conformations are possible which are not related by rotational, reflectional, or translational symmetry [25]. The unique target conformation is chosen among these conformations depending on the "designability" [26] of the sequences. The unfolded ensemble comprises of the remaining 103 345 compact conformations,

*pbiswas@chemistry.du.ac.in

which are likely to compete with the target structure. With a binary patterning of the amino acid residues, the number of possible sequences for a 27-mer is large and amounts to $2^{27}=134\,217\,728$. Such models have been extensively studied in the context of protein's neutrality and evolution [15,18,27–31]. Although non-compact conformations are important for understanding the folding phenomenon, these states are likely to have less effect on the qualitative nature of the foldability landscape [32]. However, these lattice models may be limited by the size and topology of the native conformation and the nature of the stabilizing interactions for exhibiting a distinct two-state folding regime [33]. The present study reveals that it is possible to establish the two-state folding pattern by selecting simple foldability and stability criterion, which opts viable protein sequences to explore certain aspects of protein sequence evolution.

Most foldability criteria are based on the energy of different conformations of proteins. E_f denotes the energy of the folded state of the protein and $\langle E_u \rangle$ denotes the average energy of the unfolded ensemble of states. These energy terms are commonly expressed in terms of site-specific monomer probabilities for each sequence position. The energy of the sequence in a particular target conformation, E_f may be expressed as

$$E_f \approx \overline{E_f} = \sum_{i=1}^{N} \sum_{\alpha=1}^{2} \sum_{k} \sigma_{ik}^{(1)} \gamma_{ik}^{(1)}(\alpha) \omega_i(\alpha). \tag{1}$$

The fluctuations in E_f about its mean value due to variation of sequences is assumed to be small. The propensity of the ith monomer to reside in the kth structural context is denoted by $\gamma_k^1(\alpha)$. Such contexts indicate whether the ith site is buried in the interior or accessible to the solvent or the particular type of secondary structures associated with it. The term $\sigma_{ik}^{(1)}$ contains the structural information of the ith monomer to reside in the kth structural context, as given by

$$\sigma_{ik}^{(1)} = \begin{cases} 1 & \text{if site } i \text{ is in structural context } k, \\ 0 & \text{if not.} \end{cases} \tag{2}$$

The sequence averaged energy of an ensemble of the unfolded conformations may be similarly expressed as

$$\overline{\langle E_u \rangle} = \sum_{i=1}^{N} \sum_{\alpha=1}^{2} \sum_{k} \langle \sigma_{ik}^{(1)} \rangle_u \gamma_{ik}^{(1)}(\alpha) \omega_i(\alpha). \tag{3}$$

The difference in the folded state energy (E_f) and average energy of the ensemble of unfolded states ($\langle E_u \rangle$) denotes the stability gap Δ, given by

$$\begin{aligned} \Delta &\equiv E_f - \langle E_u \rangle \approx \overline{E_f} - \overline{\langle E_u \rangle} \\ &= \sum_{i=1}^{N} \sum_{\alpha_i=1}^{2} \omega_i(\alpha) \sum_{k} (\sigma_{ik}^{(1)} - \langle \sigma_{ik}^{(1)} \rangle_u) \gamma_{ik}^{(1)}(\alpha). \end{aligned} \tag{4}$$

For each sequence, the variance of the energy of the unfolded ensemble of states is given by

$$\begin{aligned} \Gamma^2 &= \langle E_u^2 \rangle - \langle E_u \rangle^2 \\ &= \sum_{i,j} \sum_{\alpha,\alpha} \sum_{k,k'} \gamma_k^{(1)}(\alpha) \gamma_{k'}^{(1)}(\alpha) (\langle \sigma_{ik}^{(1)} \sigma_{jk'}^{(1)} \rangle \\ &\quad - \langle \sigma_{ik}^{(1)} \rangle \langle \sigma_{jk'}^{(1)} \rangle) \omega_{ij}(\alpha, \alpha), \end{aligned} \tag{5}$$

where the pairwise monomer probability $\omega_{ij}(\alpha,\alpha)$ is given by

$$\omega_{ij}(\alpha,\alpha) = \begin{cases} \omega_i(\alpha)\omega_j(\alpha) & \text{if } i \neq j, \\ \omega_i(\alpha)\delta_{\alpha,\alpha} & \text{if } i = j, \end{cases} \tag{6}$$

where $\delta_{\alpha,\alpha}$ is the Kronecker delta function.

A generalized foldability criterion ϕ is derived using a high temperature cumulant expansion [34] up to the second order for approximating the free energy of folding [35,36]. The truncation at second order is exact assuming the energy fluctuations in the unfolded state ensemble are Gaussian. This foldability criterion is evaluated as a linear combination of the mean and variance of the energy of the unfolded ensemble.

$$\phi = \Delta + \frac{1}{2}\Gamma^2, \tag{7}$$

where ϕ, Δ, and Γ^2 are dimensionless quantities scaled by appropriate units of thermal energy. The theory relies on the maximization of the sequence entropy S subject to the normalization of the site-specific monomer probabilities

$$\sum_{\alpha=1}^{m} \omega_i(\alpha) = 1 \quad \forall\, i \tag{8}$$

and the energy constraint given by Eq. (7)

$$S = -\sum_{i=1}^{N} \sum_{\alpha=1}^{m} \omega_i(\alpha) ln\ \omega_i(\alpha). \tag{9}$$

Solving the simultaneous equations that define the maximum of the variational functional of the set of monomer probabilities and the constraint equations, a set of coupled nonlinear transcendental equations are obtained.

$$\omega_i(\alpha) = \frac{1}{q_i}[\exp(-\beta_\phi \phi_\omega)],$$

$$\phi = \Delta + \frac{1}{2}\Gamma^2, \tag{10}$$

where $q_i = \Sigma_{\alpha=1}^{m} \exp(-\beta_\phi \phi_\omega)$, $\phi_\omega = \partial\phi / \partial\omega_i(\alpha)$, and β_ϕ is the Lagrange multiplier for Eq. (7). This set of equations are solved numerically to yield the site-specific monomer probabilities and Lagrange multipliers for a given value of ϕ. The individual contribution of Δ and Γ^2 are determined self-consistently from the equations Eqs. (4), (5), and (10), respectively, and thus cannot have any arbitrary values.

III. SIMULATION METHODS

Sequence optimization methods like Monte Carlo (MC) employ an effective temperature to scan the sequence space.

MC methods with simulated annealing are used to identify wild-type sequences for a specified target structure without getting trapped at any local minima. It also provides an ensemble of solutions at a finite temperature following a Boltzmann distribution. A modified form of MC, known as mean-field biased Monte Carlo (MFBMC) often shows a better convergence to a low energy minimum. To identify specific sequences, it would be useful to combine the sampling power of MC methods with the convergence efficacy of the mean-field methods. In MC simulation the site-identity monomer probabilities are predicted by choosing an initial random monomer probability profile $\Psi \equiv \{\omega_1(\alpha) \cdots \omega_N(\alpha)\}$ at each site i according to a uniform distribution. At each MC step, a random site i is chosen where the new monomer probability is predicted randomly and the trial sequence ω' will be generated according to the Metropolis acceptance probability [37,38],

$$a = \min(1, \exp\{-\beta[\phi(\Psi') - \phi(\Psi)]\}) \tag{11}$$

$\beta = 1/T$, where T is the temperature at each MC step. ϕ for each probability profile is calculated from Eq. (7).

In MFBMC simulation, the trial sequences are opted according to a self-consistent theory generated probability profile which greatly reduces the computational time associated with each simulation step. The sequence corresponding to a specified target structure would be accepted according to the acceptance probability a, which is given by [38],

$$a = \min\left(1, \frac{P(\Psi)}{P(\Psi')} \exp\{-\beta[\phi(\Psi') - \phi(\Psi)]\}\right) \tag{12}$$

where $P(\Psi) = \Pi_i \omega_i(\alpha)$ In both simulation methods suitable sequences are sampled for the same target structure the only difference is, unlike MC where the sequences are sampled randomly, the search for sequences in MFBMC is predetermined from the self-consistent mean-field theory.

For both MC and MFBMC, the system is cooled at each simulation step where the temperature decays exponentially $T(t) = T_0 \exp(-t/\tau)$. T_0 is the initial temperature set to a very high value (5000) at $t=0$ so that all sequences would be thermally accessible [38]. The decay constant τ can be tuned to yield different cooling rates and t is the number of MC steps. The system may be equilibrated to a desired temperature to search for optimal sequences at a specified T. Simulations are performed at different τ to yield the probability profile for the target structure at a given T for various values of ϕ. Minimization runs were terminated at $T = 10^{-4}$ to ensure that the calculations were stopped well after the system was effectively frozen.

The generated wild-type sequences resemble the naturally occurring protein sequences corresponding to a given structure/phenotype [20,21]. These sequences are randomly mutated at single site at a time and the viable mutated sequences are selected if they retain the same native fold as of "wild-type" sequence and if the free energy change ($\Delta\Delta G_{mut}$) due to mutations is below some critical cut-off values ($\Delta\Delta G_{cut}$). The cut-off values for $\Delta\Delta G_{cut}$ are 0, −1, −1.2, and −1.5, respectively. The free energy of a sequence in the target state can be calculated with respect to the ensemble of unfolded states [39,18],

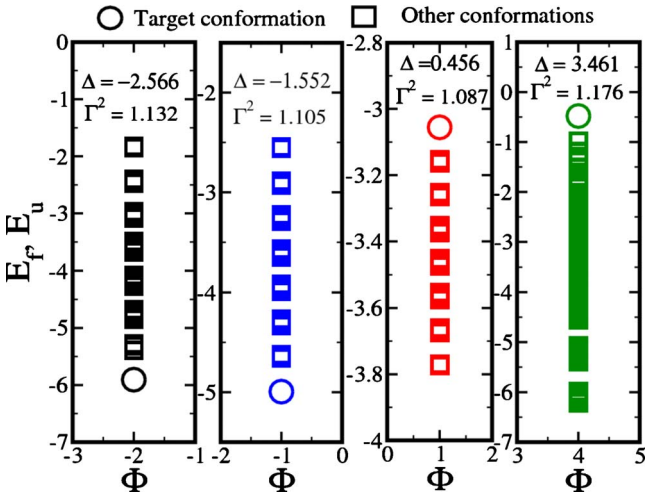

FIG. 1. (Color online) Contribution of Δ and Γ^2 values in determining protein's foldability. The target state is denoted by circles and the other unfolded conformations are represented by squares.

$$\Delta G = E_f + k_B T \, ln[Z - \exp(-E_f/k_B T)], \tag{13}$$

where $k_B T$ is the Boltzmann constant times temperature, Z is the partition function and E_f is the energy of the target state. All 103 345 unfolded conformations are used to calculate the partition function Z. The change in free energy due to mutation is measured by

$$\Delta\Delta G_{mut} = \Delta G_{mutated} - \Delta G_{wild}, \tag{14}$$

where $\Delta\Delta G_{mut} < 0$ implies a thermodynamically favorable mutation.

IV. RESULTS AND DISCUSSIONS

Figure 1 illustrates the applicability of the self-consistent theory in designing protein sequences compatible with a given target structure. Here four different sequences corresponding to $\phi = -2$, $\phi = -1$, $\phi = 1$, $\phi = 4$ are selected. For each sequence the target state energy E_f and the energy of the unfolded ensemble $\langle E_u \rangle$ for all 103 346 conformations are calculated and plotted along the y axis. The x axis represents the range of ϕ values. The plot depicts that for $\phi < 0$, the target conformation is the most stable state as represents the minimum energy conformation. The energies of all unfolded states are higher than the target state. The protein sequences in this region choose the target structure as the unique native state among all conformations. The situation exactly reverses for $\phi > 0$. The target state is destabilized compared to the ensemble of the unfolded conformations and the designed protein sequences are not foldable in spite of the presence of a unique native state. In this regime, a substantial percentage of all possible sequences may not fold to the "proteinlike" target structure, though they have distinct ϕ values. The individual contributions of Δ and Γ^2 are also determined for each value of ϕ. For negative values of ϕ, $\Delta < 0$ which im-

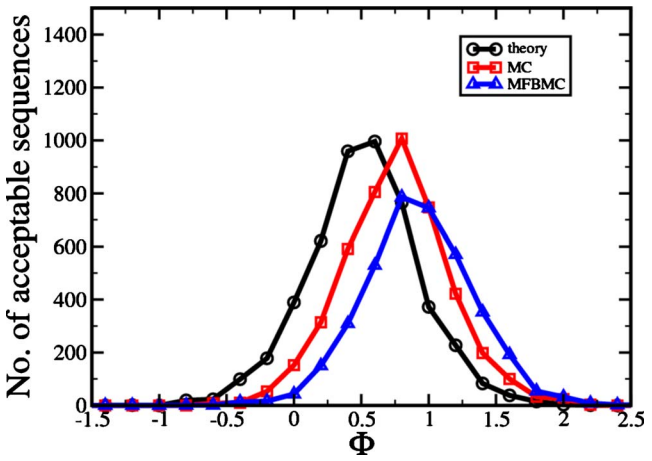

FIG. 2. (Color online) Distribution of number of acceptable sequences at various ϕ after random site mutations.

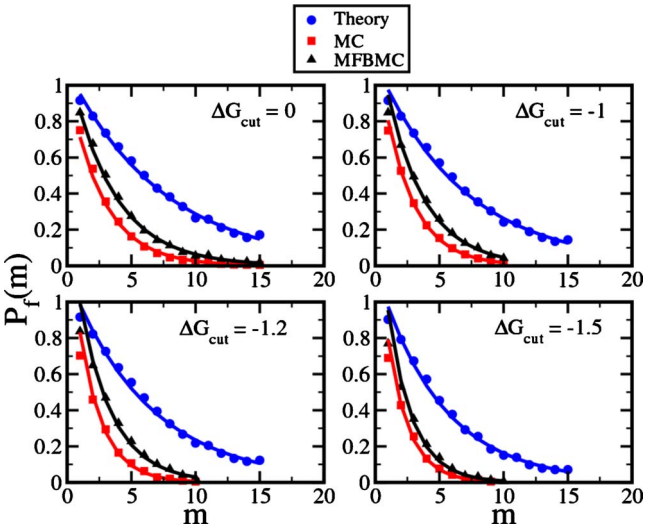

FIG. 3. (Color online) Comparison of neutrality of simulation and theory designed protein sequences at different $\Delta\Delta G_{cut}$ cut-off values.

plies that the target structure is stabilized with respect to the energy of the unfolded conformations. The variance in energy of the unfolded states Γ^2 also differs in the chosen range of ϕ values. Thus sequences sampled from different range of ϕ values differ in terms of their designability in opting the unique native structure.

The effect of cumulative random point mutations is explored by calculating the number distribution of acceptable mutated sequences as a function of the generalized foldability criterion ϕ. The "wild-type" sequences are generated for the theory, MC and MFBMC, respectively, for the ϕ value corresponding to maximum sequence entropy. This ensures that these sequences have a certain degree of similarity in their target structures and fold back to the corresponding native structures. The similarity between different simulated sequences in the maximum entropy region is compared to that of the mean-field theory generated sequences. For a binary representation of the residue identities, a sequence similarity measure $q_{seq}(\alpha,\alpha')$ of two different sequences $\alpha = \alpha_1,\ldots,\alpha_N$ and $\alpha' = \alpha'_1,\ldots,\alpha'_N$ may be defined by

$$q_{seq}(\alpha,\alpha') = \sum_{i=1}^{N} \sigma(\alpha_i)\sigma(\alpha'_i), \tag{15}$$

where the spin-equivalent variables are

$$\sigma(\alpha) = \begin{cases} 1 & \text{for } \alpha = P \\ -1 & \text{for } \alpha = H. \end{cases}$$

The results show that the homology between theory designed sequence and the simulation generated sequences are low.

A total of 5×10^6 cumulative random point mutations are performed on the simulated and theory generated sequences. The viable sequences are selected if they fold back to the desired target structure as their native state and $\Delta\Delta G_{mut}<0$ compared to the respective wild-type sequences. The number of viable sequences are binned at different ϕ values, where $\delta\phi=0.2$ is the width of each bin. Figure 2 represents the number distributions of accepted mutated sequences vs ϕ for random point mutations. The distributions are similar except the shift in maxima which corresponds to the slightly different ϕ values at the entropy maximum region. This suggests that the number distribution of mutated sequences associated with a specific ϕ values has some degree of universality and does not depend on the sequence composition. This also implies that there may be sequences with minimal homology which may adopt the same fold. In course of evolution, mutations of specific amino acids govern the foldability of a particular sequence and this foldability in turn influences the degree of selective pressure in choosing suitable genotypes.

Figure 3 shows protein's neutrality which is defined by the fraction of the single amino acid substitutions that retain the similar native structure as the wild-type sequence. For the specified target structure corresponding to $\phi=-1.2$, theory, MC and MFBMC generated wild-type sequences are randomly mutated 1000 times and the resultant monomer probabilities are generated randomly. In the next generation, each sequence from this set of 1000 single point mutated sequences is further mutated randomly. All mutations are carried out cumulatively and the mutation rate is kept constant to allow only single mutation at each sequence per generation. The mutations are performed for 15 successive generations. After each generation of mutation, the sequences are considered "viable" if they fold back to the same native structure and if the free energy change due to mutations $\Delta\Delta G_{mut}$ are lower than the chosen $\Delta\Delta G_{cut}$ values. To verify whether the mutated sequences fold back to the same target structure, the energies of all possible 103 346 conformations are calculated for each mutated sequence after each generation of mutation. For each generation of mutation, if the target structure is found to possess the lowest energy, then the mutated sequence is able to choose the specified target structure as its unique native state. The fraction of viable sequences $[P_f(m)]$ after mth generation of mutations is calculated and plotted against m. The same procedure is followed for MC and MFBMC generated sequences respectively corresponding to the same specified target structure at $\phi=-1.2$. For large values of m, $P_f(m)$ decays exponentially

TABLE I. Comparison of neutrality of simulation and theory designed protein sequences at different $\Delta\Delta G_{cut}$ cut-off values.

Wild-type sequence	$\Delta\Delta G_{cut}=0$	$\Delta\Delta G_{cut}=-1$	$\Delta\Delta G_{cut}=-1.2$	$\Delta\Delta G_{cut}=-1.5$
Theory	0.8754	0.8640	0.8527	0.8190
MFBMC	0.7489	0.7137	0.6761	0.5908
MC	0.6911	0.6506	0.5787	0.5449

[19] with $\langle\nu\rangle^m$ where $\langle\nu\rangle$ is the protein's neutrality. A linear regression of $ln\ P_f(m)$ vs m yields the neutrality $\langle\nu\rangle=e^s$ where s is the slope of the regression line.

In Table I the comparison of neutrality of theory, MC and MFBMC designed sequences are provided. The neutrality also decreases as the cut-off free energy values are more negative laterally along Table I. As the cut-off value becomes more negative, the number of correctly folded sequences at a given mutational distance m decreases reducing the value of $P_f(m)$. This reflects that the foldability of different sequences for a common target structure varies with degrees of evolutionary selective pressure and hence shape the evolutionary sequence space accordingly. This is also evident from the fact that protein's neutrality decreases with increasing number of cumulative point mutations. With increasing generation of mutations the foldability of protein sequences decrease and thus it becomes difficult to fold back to the specified target structure. Results show that at any cut-off free energy value, the theoretically designed sequence is more neutral compared to the simulation designed sequences in terms of tolerating multiple single point mutations. This may be rationalized by the fact that the theoretically designed sequences are optimized for a particular ϕ value.

The present study highlights the relation between protein's evolvability and foldability. The effect of cumulative mutations on Δ and Γ^2 are separately analyzed. A set of seven theory designed protein sequences are chosen at different ϕ values ranging from −2.5 to 0.5 at an interval of 0.5. Outside this range of ϕ the designed sequences are not optimized to allow for mutations without disrupting the protein's target structure. Each of these wild-type sequences are mutated randomly 1000 times at each generation. The mutated sequences are considered viable if they fold back to the same native structure and have $\Delta\Delta G_{mut}<0$ values. This procedure is repeated up to eighth generation of successive mutations. For the total number of viable sequences the average value of Δ denoted by $\langle\Delta\rangle$ and average Γ^2 denoted by $\langle\Gamma^2\rangle$ are calculated. At each generation of mutations the sequences are successively substituted. Figures 4(a) and 4(b) show how $\langle\Delta\rangle$ and $\langle\Gamma^2\rangle$, respectively, affect the choice of total number of viable sequences until eight generation of mutations.

The graph shows that for each generation of mutation at a certain value of $\langle\Delta\rangle$ and $\langle\Gamma^2\rangle$ the total number of acceptable mutated sequences are maximum and are denoted as Δ_m and Γ^2_m. Results show that with increasing generation of mutations the Δ_m and Γ^2_m shift. Δ_m shifts toward more negative Δ whereas Γ^2_m shifts toward more positive values. At lower generation of mutations this shift is large and decreases with increasing number of generations. These results indicate that with accumulation of mutations in protein sequences, the stability of native fold increases. Also with increasing generation of mutations $\langle\Gamma^2_m\rangle$ increases, which signifies that the width of the energy of unfolded ensemble increases. The asymptotic variation of both $\langle\Delta\rangle$ and $\langle\Gamma^2\rangle$ is justified by the increasing stability of the native state which attains a maximum value for a certain generation of mutations. With increasing generations of mutations, the possibility of obtaining viable mutated sequences with high $\langle\Delta\rangle$ and low $\langle\Gamma^2\rangle$ decreases. Therefore, with accumulation of mutations under neutral conditions, the evolutionary dynamics simultaneously stabilizes the native state and destabilizes the other competing structures to maintain the protein's native structure.

V. CONCLUSIONS

The present results show that at various degrees of selective pressure, foldability parameter decides which mutated

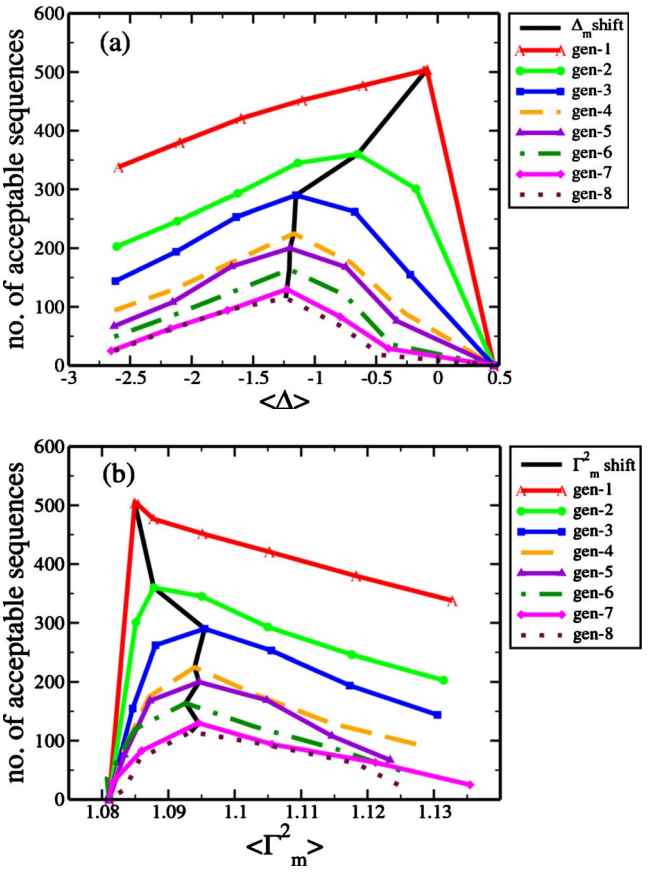

FIG. 4. (Color online) Variation of $\langle\Delta\rangle$ (a) and $\langle\Gamma^2\rangle$ (b) with generation of mutations.

sequence passes the fitness criteria (selective pressure). The theoretically designed sequences are optimally foldable, hence more robust toward random point mutations. The foldability of sequences thus effectively shapes up the evolutionary neutral sequence space by selecting the suitable genotypes for a target structure/phenotype. Also the effect of random point mutations is found to be similar for comparably foldable sequences even if they have low sequence homology. This may account for reinterpreting neutrality and evolvability to design "functionally fit" protein sequences.

ACKNOWLEDGMENTS

This work was financially supported by DST (Grant No. SR/S1/PC-07/06), India and Delhi University Research Grant.

[1] M. Sasai, Proc. Natl. Acad. Sci. U.S.A. **92**, 8438 (1995).
[2] J. D. Bryngelson and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **84**, 7524 (1987).
[3] G. Archontis and M. Karplus, J. Chem. Phys. **105**, 11246 (1996).
[4] H. S. Chan and E. Bornberg-Bauer, Appl. Bioinf. **1**, 121 (2002).
[5] Y. Xia and M. Levitt, Proteins **55**, 107 (2004).
[6] M. Kimura, *The Neutral Theory of Molecular Evolution* (Cambridge University Press, Cambridge, England, 1983).
[7] J. L. King and T. H. Jukes, J. Mol. Biol. **233**, 305 (1969).
[8] *Proceedings of the Sixth International Congress on Genetics*, All ACM Conferences No. 356, edited by D. F. Jones (Brooklyn Botanic Gardens, New York, 1932).
[9] J. M. Smith, Nature (London) **225**, 563 (1970).
[10] U. Bastolla, H. E. Roman, and M. Vendruscolo, J. Theor. Biol. **200**, 49 (1999).
[11] G. G. Tiana, R. A. Broglia, H. E. Roman, E. Vigezzi, and E. Shakhnovich, J. Chem. Phys. **108**, 757 (1998).
[12] C. O. Wilke, BMC Genet. **5**, 25 (2004).
[13] W. Fontana and P. Schuster, Science **280**, 1451 (1998).
[14] R. Forster, C. Adami, and C. O. Wilke, J. Theor. Biol. **243**, 181 (2006).
[15] R. A. Broglia, G. Tiana, H. E. Roman, E. Vigezzi, and E. Shakhnovich, Phys. Rev. Lett. **82**, 4727 (1999).
[16] E. I. Shakhnovich, Folding Des. **3**, R45 (1998).
[17] A. V. Finkelstein, A. Gutin, and A. Badretdinov, Proteins **23**, 142 (1995).
[18] J. D. Bloom, J. J. Silberg, C. O. Wilke, D. A. Drummond, C. Adami, and F. H. Arnold, Proc. Natl. Acad. Sci. U.S.A. **102**, 606 (2005).
[19] C. O. Wilke, J. D. Bloom, D. A. Drummond, and A. Raval, Biophys. J. **89**, 3714 (2005).
[20] A. Bhattacherjee and P. Biswas, J. Phys. Chem. B **113**, 5520 (2009).
[21] A. Bhattacherjee and P. Biswas, J. Chem. Phys. **131**, 125101 (2009).
[22] E. I. Shakhnovich, Phys. Rev. Lett. **72**, 3907 (1994).
[23] N. D. Socci and J. N. Onuchic, J. Chem. Phys. **103**, 4732 (1995).
[24] M.-H. Hao and H. A. Scheraga, Proc. Natl. Acad. Sci. U.S.A. **93**, 4984 (1996).
[25] E. Shaknovich and A. Gutin, J. Chem. Phys. **93**, 5967 (1990).
[26] H. Li, R. Helling, C. Tang, and N. Wingreen, Science **273**, 666 (1996).
[27] G. Tiana, B. E. Shakhnovich, N. V. Dokholyan, and E. I. Shakhnovich, Proc. Natl. Acad. Sci. U.S.A. **101**, 2846 (2004).
[28] Y. Xia and M. Levitt, Proc. Natl. Acad. Sci. U.S.A. **99**, 10382 (2002).
[29] L. C. Oliveira, R. T. H. Silva, V. B. P. Leite, and J. Chahine, J. Chem. Phys. **125**, 084904 (2006).
[30] B. P. Blackburne and J. D. Hirst, J. Chem. Phys. **119**, 3453 (2003).
[31] B. P. Blackburne and J. D. Hirst, J. Chem. Phys. **115**, 1935 (2001).
[32] S. Govindarajan and R. A. Goldstein, Proteins: Struct., Funct., Genet. **29**, 461 (1997).
[33] E. Nelson and N. Grishin, J. Chem. Phys. **118**, 3342 (2003).
[34] M. P. Morrissey and E. I. Shaknovich, Folding Des. **1**, 391 (1996).
[35] J. G. Saven, J. Chem. Phys. **118**, 6133 (2003).
[36] P. Biswas, J. Zou, and J. G. Saven, J. Chem. Phys. **123**, 154908 (2005).
[37] N. Metropolis, A. W. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[38] J. Zou and J. G. Saven, J. Chem. Phys. **118**, 3843 (2003).
[39] D. M. Taverna and R. A. Goldstein, J. Mol. Biol. **315**, 479 (2002).