Evolvability and Single-Genotype Fluctuation in Phenotypic Properties: A Simple Heteropolymer Model

Tao Chen,[†] David Vernazobres,[‡] Tetsuya Yomo,[§¶] Erich Bornberg-Bauer,[‡] and Hue Sun Chan[†*]

[†]Departments of Biochemistry and of Molecular Genetics, Faculty of Medicine, and Department of Physics, University of Toronto, Toronto, Ontario, Canada; [‡]Institute for Evolution and Biodiversity, School of Biological Sciences, University of Münster, Münster, Germany; [§]Department of Bioinformatic Engineering, Graduate School of Information Science and Technology, and the Graduate School of Frontier Bioscience, Osaka University, Osaka, Japan; and [¶]Exploratory Research for Advanced Technology, Japan Science and Technology Agency, Osaka, Japan

ABSTRACT Experiment showed that the response of a genotype to mutation, i.e., the magnitude of mutational change in a phenotypic property, can be correlated with the extent of phenotypic fluctuation among genetic clones. To address a possible statistical mechanical basis for such phenomena at the protein level, we consider a simple hydrophobic-polar lattice protein-chain model with an exhaustive mapping between sequence (genotype) and conformational (phenotype) spaces. Using squared end-to-end distance, R_N^2, as an example conformational property, we study how the thermal fluctuation of a sequence's R_N^2 may be predictive of the changes in the Boltzmann average $\langle R_N^2 \rangle$ caused by single-point mutations on that sequence. We found that sequences with the same ground-state $(R_N^2)_0$ exhibit a funnel-like organization under conditions favorable to chain collapse or folding: fluctuation (standard deviation σ) of R_N^2 tends to increase with mutational distance from a prototype sequence whose $\langle R_N^2 \rangle$ deviates little from its $(R_N^2)_0$. In general, large mutational decreases in $\langle R_N^2 \rangle$ or in σ are only possible for some, though not all, sequences with large σ values. This finding suggests that single-genotype phenotypic fluctuation is a necessary, though not sufficient, indicator of evolvability toward genotypes with less phenotypic fluctuations.

INTRODUCTION

The study of protein evolution entails ascertaining how changes in a protein's amino acid sequence lead to changes in its biological functions (1). Biological functions of proteins, in turn, are often intimately related to their conformational structures. Thus, to address principles of protein evolution, the mappings between sequences and structures in various simplified heteropolymer models have been investigated. Using explicit—albeit highly coarse-grained—representations of the protein chain with physics-inspired interactions (see, e.g., (2–16)), these modeling efforts have led to significant advances (reviewed in (17,18)). Because the folded nature of many proteins are crucial for their functions, explicit-chain models of protein evolution to date have focused primarily on the mapping from sequences to their ground-state (lowest-energy) conformations that represent the folded native structures of globular proteins. In these analyses, the role of excited-state conformations (those with energies higher than that of the ground state) is only subsidiary and a structure is seen as encodable (5) or designable (6) only if it is the unique ground-state conformation of a sequence.

Recent developments, however, have revealed a more central evolutionary role for excited-state conformations that are accessible by thermal fluctuation. In the study of RNA evolution, it is well known that a sequence can assume a variety of energetically favorable shapes (19). Likewise, it is recognized that excited-state populations of proteins might serve ''promiscuous'' biological functions (20,21) and thus be subject to a selection process that can speed up evolution (22–24) and reviewed in (25,26)). This perspective is in line (20) with the energy landscape picture of protein folding (27–30), the observation of in vivo phenotypic variations modulated by molecular chaperones (31), as well as the notion that broadened phenotypic fluctuations among individuals with the same genotype can be a successful evolutionary strategy under severe selective environments (32). Accordingly, it is more appropriate to view sequence-to-structure mapping as one that takes each sequence to a structural distribution that encompasses all conformations (23), rather than one that matches each sequence solely to its ground-state conformation(s). Depending on the sequence and its environment, a sequence's structural distribution can be dominated by one conformation, as for globular proteins under folding conditions, or it can favor many conformations simultaneously, as for intrinsically disordered proteins (33–35).

To what extent, then, is a single genotype's evolvability predetermined by its phenotypic fluctuation (26,36)? At the level of protein molecules, evolvability is the propensity of an amino acid sequence or a population of sequences to develop new structural/functional features by perturbatively changing the original sequence(s). Here, genotype is identified with the amino acid sequence and phenotype corresponds to the structural or functional properties of the sequence. Sato et al. (37) and Yomo et al. (38) addressed single-sequence evolvability by using artificial evolution (39,40) of mutants of a green fluorescent protein (GFP) in bacteria. By monitoring the variance of fluorescence intensity among clones (bacteria with the same amino acid sequence for the GFP)

Submitted December 9, 2009, and accepted for publication February 26, 2010.

*Correspondence: chan@arrhenius.med.utoronto.ca

Editor: Ruth Nussinov.

0006-3495/10/06/2487/10 $2.00

doi: 10.1016/j.bpj.2010.02.046

and selecting for large increases in this trait (41), they found that the largest possible increase in fluorescence intensity caused by mutation on a genotype (a particular amino acid sequence for the GFP) is well correlated with the variance of fluorescence intensity exhibited by that genotype (37). This observation suggests that the largest achievable evolutionary change on a phenotypic property by mutation may be governed by the extent of fluctuation of that property in the original parent genotype. Inspired by this discovery, here we explore biophysical principles that may underlie such a phenomenon.

Fig. 1 illustrates the question we aim to address. Starting with two sequences (represented by the *black* and *red solid circles*) that have the same average value for a conformational (structural) property, Fig. 1 shows a hypothetical relationship between variations in sequence space (*top* and *bottom*) and variations in conformational space (*middle*). Variation in the structural distribution encoded by a sequence is expected to be smooth—at least on average—with respect to change in sequence (23). Based on this premise, Fig. 1 stipulates, hypothetically, that the sequence depicted as a red solid circle and has a broad distribution in a conformational property is likely to have a single-point mutant (e.g., *red triangle*) with a larger shift in the average value of that property than the corresponding shifts achieved by single-point mutants (e.g., *blue square*) of the sequence with a narrower distribution to begin with (*solid circle*). As commented by Sato et al. (37), such a correlation between increased single-genotype phenotypic fluctuation and increased evolvability is reminiscent of the fundamental relationship between fluctuation and dissipation in statistical physics (42) because, by analogy, enhanced evolvability may be viewed as reduced resistance to phenotypic change.

The scenario in Fig. 1 is intuitive, but the extent to which it is physically viable for protein chains remains to be assessed. Approximate analytical formulations have provided insights into the fluctuation-evolvability question (37,43) (see below) as well as whether evolvability is a selectable trait (44). However, as has been demonstrated in the study of protein folding (45), explicit-chain modeling is indispensable for evaluating whether assumptions made in analytical formulations of protein evolution are physically plausible. We adopt the simple exact HP (hydrophobic-polar) lattice model (28) for this task, as in our previous studies of evolutionary questions (9,11,15,23). Using this extremely coarse-grained and thus computationally tractable construct (17), the present effort is an essential complement to analytical approaches.

Here, we choose to study the squared end-to-end distance, R_N^2, as an example of a conformational property that exhibits thermal fluctuations. Although not directly related to biological function in general, R_N^2 is useful for a first test of principle because its average, $\langle R_N^2 \rangle$, is experimentally accessible by fluorescence resonance energy transfer (see, e.g., (46)). Other measures of conformational geometry such as chain compactness and radius of gyration may also be studied,

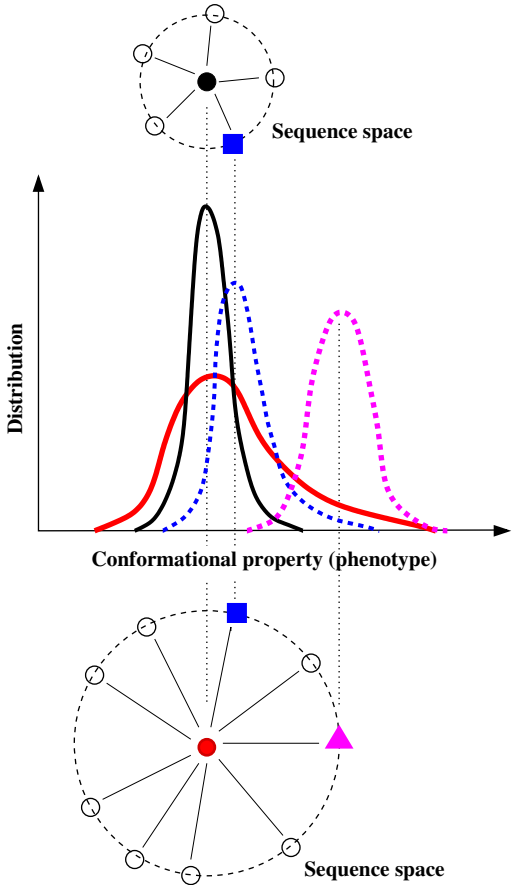

FIGURE 1 Schematic of how single-genotype phenotypic fluctuation might correlate with evolvability. Top and bottom are parts of the sequence space, wherein single-point substitutive mutations are represented by solid lines and a dashed circle is used to indicate unit Hamming distance (a single substitution) from a center sequence (*black* or *red circle*). The middle plot shows distributions of a hypothetical conformational property for four of the sequences (marked by *vertical dotted lines* and *color coding*). Other sequences are depicted as black open circles.

but we do not pursue them here. In view of the experimental advances on artificial evolution of random amino acid sequences (47), our consideration is not restricted to unique or low-degeneracy sequences (9). By extending our attention to all random sequences, we also explore how selection on $\langle R_N^2 \rangle$ might, as a side effect, lead to changes in ground-state degeneracy. Our analysis demonstrates that evolvability of $\langle R_N^2 \rangle$ indeed correlates with fluctuation in R_N^2, in a manner somewhat similar to that envisioned by analytical theory (37). However, our results also reveal richer, unanticipated features. These include a marked difference between selecting for increasing $\langle R_N^2 \rangle$ versus selecting for decreasing $\langle R_N^2 \rangle$, the detail of which will be described below.

MODEL AND METHOD

A reduced two-letter alphabet (5,48) is used in the HP model to mimic the attractive interactions among nonpolar amino acid residues (28) (Fig. 2). We adopt the HP model for evolutionary studies because it captures general

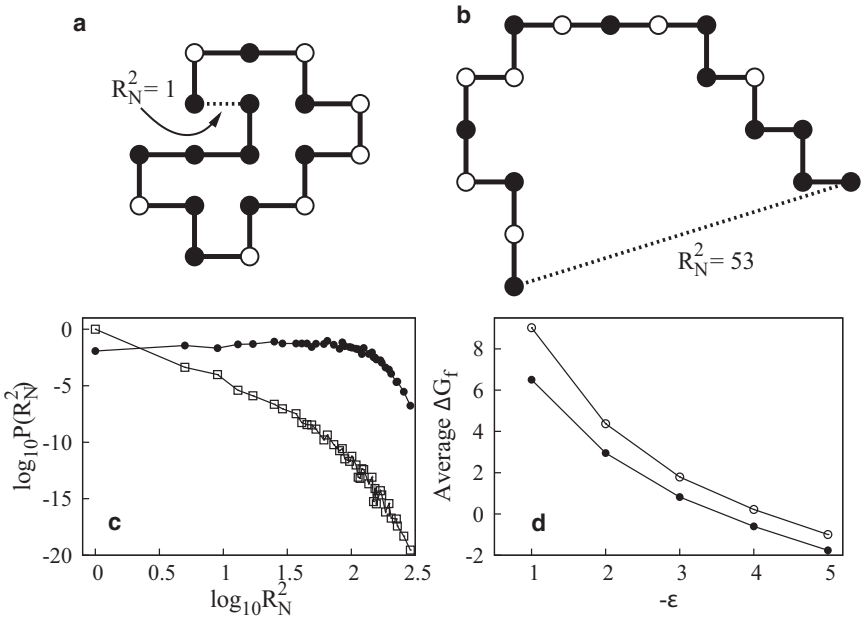

FIGURE 2 Squared end-to-end distance R_N^2 of HP model chains on square lattices. The example in panels *a* and *b* shows a $g = 2$ sequence (H residues: *solid circles*; P residues: *open circles*) in two conformations with different R_N^2 values. (*a*) One of this sequence's two ground-state conformations. The other ground-state conformation (not shown) also has $(R_N^2)_0 = 1$. (*b*) This is an unfolded conformation. (*c*) Normalized distribution of R_N^2 of an example $g = 1$ sequence HPHPHPPHPHPPHPPHHH at $\varepsilon = 0$ (*solid circles*) and $\varepsilon = -5$ (*open squares*). Note that only discrete R_N^2 values of 1, 5, 9, … are allowed by the lattice model (*circles* and *squares*); lines through the symbols are merely a guide for the eye. (*d*) ΔG_f (in units of k_BT) as a function of ε. Solid circles show ΔG_f averaged over all sequences that can make at least one HH contact (261,088 sequences with $h_N > 0$); open circles show ΔG_f averaged over the 6,349 $g = 1$ sequences in our model.

trends of the sequence-to-structure mapping for real proteins (49,50), notwithstanding the model's insufficiency for detailed protein energetics (45,51). HP and other lattice models with reduced alphabets are useful for rationalizing properties of disordered protein conformations as well (4,52–54).

As in our previous studies (9,11,15,23), model protein chains are configured as self-avoiding walks on the two-dimensional square lattice. A favorable energy ε (< 0) is assigned to each hydrophobic-hydrophobic (HH) contact; thus a conformation with h HH contacts has energy εh. The ground state of a sequence is the collection of conformations each of which has the largest number (denoted as $h = h_N$) of HH contacts that the sequence can achieve. The number of such conformations is the ground-state degeneracy g. We denote the R_N^2 value of a ground-state conformation by $(R_N^2)_0$. On the square lattice, the possible R_N^2 values (see illustrations in Fig. 2, *a* and *b*) are given by $R_N^2 = x^2 + y^2$ where x and y are nonnegative integers, $x + y$ is restricted to be odd (even) when n is even (odd), and $0 < x + y \leq n - 1$. As before (9,11,15,23), here we use chains with length (number of residues) $n = 18$, for which there are 39 possible R_N^2 values (1,5,9,…289) among the 5,808,335 conformations that are not related by rigid rotations and mirror reflections. Short two-dimensional HP model chains are appropriate for mimicking the ratio between the numbers of surface and interior residues in real globular proteins (28) and are apparently adequate for rationalizing their hydrophobic patterns (49).

For each of the 2^{18} HP sequences in our model, we obtain by exact enumeration the number of conformations as a joint function, $g(h, R_N^2)$, of h and R_N^2

$$P\left(R_N^2\right) = \frac{\sum_{h=0}^{h_N} g\left(h, R_N^2\right)\exp\left(-\varepsilon h/k_BT\right)}{\sum_{R_N^2}\sum_{h=0}^{h_N} g\left(h, R_N^2\right)\exp\left(-\varepsilon h/k_BT\right)}, \qquad (1)$$

which is the distribution of R_N^2, where k_B is the Boltzmann constant and T is absolute temperature. To simplify notation, ε is given in units of k_BT below. When ε becomes more negative, $P(R_N^2)$ becomes more dominated by the average value, $\overline{(R_N^2)_0}$, of $(R_N^2)_0$ among its ground-state conformation(s). For $g = 1$, $\overline{(R_N^2)_0}$ reduces to a single value of $(R_N^2)_0$. An example showing the variation of $P(R_N^2)$ with ε is provided in Fig. 2 *c*. In this example, because ground-state $(R_N^2)_0 = 1$, $P(R_N^2)$ for $\varepsilon = -5$ is sharply peaked at $R_N^2 = 1$. The distribution of $\overline{(R_N^2)_0}$ over all 2^{18} sequences is shown in Fig. S1 in the Supporting Material.

Information about ground-state thermodynamic stability in our model is provided in Fig. 2 *d*, showing averages, over sequences, of the free energy of folding,

$$\Delta G_f = -k_BT\ln\left[\frac{g(h_N)\exp(-\varepsilon h_N)}{\sum_{h=0}^{h_N-1} g(h)\exp(-\varepsilon h)}\right], \qquad (2)$$

where

$$g(h) = \sum_{R_N^2} g\left(h, R_N^2\right)$$

is the density of states (9). The data indicate that $\varepsilon \leq -5$ is required, on average, for ground-state dominance ($\Delta G_f < 0$). The Boltzmann-weighted average of squared end-to-end distance

$$\left\langle R_N^2\right\rangle = \sum_{R_N^2} R_N^2 P\left(R_N^2\right) \qquad (3)$$

is computed for each sequence by using Eq. 1. Averages for other functions of h and R_N^2 are similarly defined. Fluctuation of R_N^2 is characterized by its standard deviation

$$\sigma = \sqrt{\left\langle\left(R_N^2 - \left\langle R_N^2\right\rangle\right)^2\right\rangle}. \qquad (4)$$

By definition, σ has the same dimension ([length]^2) as R_N^2, both of which are expressed in units of squared lattice bond length. Fig. S2 provides the ε-dependence of $\langle R_N^2\rangle$ and σ for two sets of sequences to be analyzed below. Each is a net of unique ($g = 1$) sequences interconnected by single-point H → P or P → H substitutive mutations. Results in Fig. S2, *a* and *c*, are for a neutral net in which all sequences encode for the same $(R_N^2)_0 = 1$ ground-state conformation (9,55), whereas those in Fig. S2, *b* and *d*, are for a net in which all sequences encode for ground-state conformations with $(R_N^2)_0 = 9$. The latter $(R_N^2)_0 = 9$ net is not a neutral net in the original definition (9) because sequences in this net can encode for different

ground-state conformations, although this net may be viewed as neutral, insofar as $(R_N^2)_0$ is concerned.

RESULTS AND DISCUSSION

To address the relationship between evolvability and single-genotype phenotypic fluctuation in the context of our model, we ask: To what extent can σ of a sequence predict the change in $\langle R_N^2 \rangle$ among the sequence's single-point mutants?

We first inspect the relationship between σ and $\langle R_N^2 \rangle$ for all sequences. When the HH attraction is weak (Fig. 3, *left column*), a single correlation covers all sequences. This trend follows from the fact that more-compact conformations tend to have smaller R_N^2 values (e.g., whereas the compact conformations that can be uniquely encoded by $g = 1$ sequences have $R_N^2 \leq 29$, the maximum possible R_N^2 among all conformations is 289). Therefore, sequences with thermodynamically more stable ground states tend to have somewhat less open conformations and thus smaller $\langle R_N^2 \rangle$ values even under a weakly favorable ε. Conformational fluctuations of these sequences tend to be less because of their higher thermodynamic stability; hence a general correlation between $\langle R_N^2 \rangle$ and σ in Fig. 3, *a* and *c*.

The situation under stronger folding conditions is quite different (Fig. 3, *right column*). Whereas a decrease in σ with a more negative ε value is expected because stronger HH attractions reduce conformational fluctuations, Fig. 3, *b* and *d*, show that the relationship between σ and R_N^2 is complex under conditions favorable to folding. Two noteworthy features emerge:

1. Instead of dispersing widely under weakly folding conditions at $\varepsilon = -1$, sequences with the same uniform ground-state $(R_N^2)_0$ (plotted in the *same color*) now cluster together for $\varepsilon = -5$ in Fig. 3, *b* and *d*.
2. At $\varepsilon = -5$, a funnel-like variation of σ with $\langle R_N^2 \rangle$ develops for each set of sequences with $(R_N^2)_0 = 1, 5, 9, 13, 17$ (each set shown in a different color) such that as σ decreases toward ≈ 0, deviations of $\langle R_N^2 \rangle$ from $(R_N^2)_0$ also decreases toward ≈ 0, with $(R_N^2)_0$ acting like an attractor. This pattern applies to the $g = 1$ sequences (Fig. 3 *b*) that have been used to model natural globular proteins (9) as well as sequences that have multiple ground-state conformations sharing the same $(R_N^2)_0$ (plotted in *black* and in *color* in Fig. 3 *d*).

The emergence of these features suggest that the general formula

$$\langle x \rangle_{a+\Delta a} - \langle x \rangle_a = b(\Delta a)\left(\sigma_x^2\right)_a \tag{5}$$

proposed by Sato et al. (37) to relate change in the average of a variable x with variance $(\sigma_x^2)_a$ of x may apply to each of the funnel-like clusters in Fig. 3, *b* and *d*. Following Sato et al. (37), the left-hand side in Eq. 5 is the change in the average of x induced by a change $a \rightarrow a + \Delta a$ in a parameter a related to x. On the right-hand side of Eq. 5, b is a constant independent of a and $(\sigma_x^2)_a$ is the variance of x before the $a \rightarrow a + \Delta a$ change (37). We test the applicability of Eq. 5 to our model system by setting the variable x to our R_N^2 and identifying Δa as a unit change in mutational (Hamming) distance (from a reference sequence) caused by a single-point H $\rightarrow$ P or P $\rightarrow$ H mutation. Using this formulation, we aim to ascertain the extent to which changes in $\langle R_N^2 \rangle$ caused by single-point mutations in our model can be determined by σ.

Such an analysis requires information on the model sequences' mutational connections, which we will investigate below. Although mutational connectivity is not included in Fig. 3, the appearance of a curved-funnel pattern for each

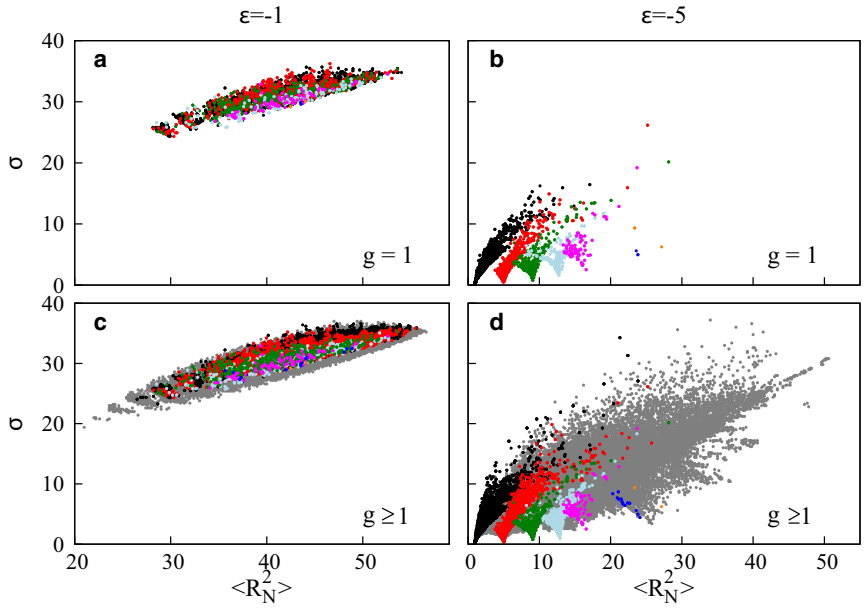

FIGURE 3 Boltzmann average of a conformational property and its thermal fluctuation. Shown here are scatter plots of $\langle R_N^2 \rangle$ and σ at $\varepsilon = -1$ (*a* and *c*) and $\varepsilon = -5$ (*b* and *d*) for various sets of sequences, as follows. Data points in panels *a* and *b* for 6349 unique ($g = 1$) sequences are plotted in black, red, green, light blue, magenta, blue, and orange, respectively, for $(R_N^2)_0 = 1$, 5, 9, 13, 17, 25, and 29. Data points in *c* and *d* for 19,309 $g \geq 1$ sequences each with only one uniform $(R_N^2)_0$ value for its ground-state conformation(s) are plotted using the same color code for $(R_N^2)_0$ as that in panels *a* and *b*. This set of sequences includes those in panels *a* and *b*. Plotted in gray in panels *c* and *d* are data points for the other 2^{18}–19,309 = 242,835 sequences in the model, each with more than one $(R_N^2)_0$ value among its $g > 1$ ground-state conformations.

set of sequences with the same $(R_N^2)_0$ in Fig. 3, *b* and *d* (*black* and *color dots*) already suggests that their behaviors might, to an extent, conform to Eq. 5. The curved-funnel shapes indicate that the magnitude of horizontal change in $\langle R_N^2 \rangle$ from one sequence to the next is larger for sequences with larger σ-values located higher up in the funnels. This behavior would be similar to that described by Eq. 5 if we assume that changing from one sequence to a neighboring sequence in Fig. 3, *b* and *d*, corresponds roughly to an $a \rightarrow a + \Delta a$ process, with $b(\Delta a) > 0$ or $b(\Delta a) < 0$ depending on whether the change in sequence results in a positive or negative change in $\langle R_N^2 \rangle$. Fig. 3, *b* and *d*, show two classes of behaviors in this regard. For sequences with uniform $(R_N^2)_0 = 1$ (*black dots*), the funnel is one-sided because $\langle R_N^2 \rangle < 1$ is impossible in the model. For sequences with uniform $(R_N^2)_0$ in the range $5 \leq (R_N^2)_0 \leq 17$, the funnels are two-sided because for these cases, $R_N^2 < (R_N^2)_0$ is possible for some excited-state conformations. Funnel-like organization for sequences with uniform $(R_N^2)_0 > 17$ is not easily discerned because the number of such sequences is small: There are 156 $g = 1$ sequences for $(R_N^2)_0 = 17$ but only 3 $g = 1$ sequences each for ground-state $(R_N^2)_0 = 25$ and $(R_N^2)_0 = 29$. The scatter of gray dots in Fig. 3 *d* shows that funnel-like organization is not apparent for $g > 1$ sequences with nonuniform $(R_N^2)_0$.

In view of the suggestive trends in Fig. 3, we now address directly our model's conformity, or lack thereof, with Eq. 5. We first focus on two networks of sequences interconnected

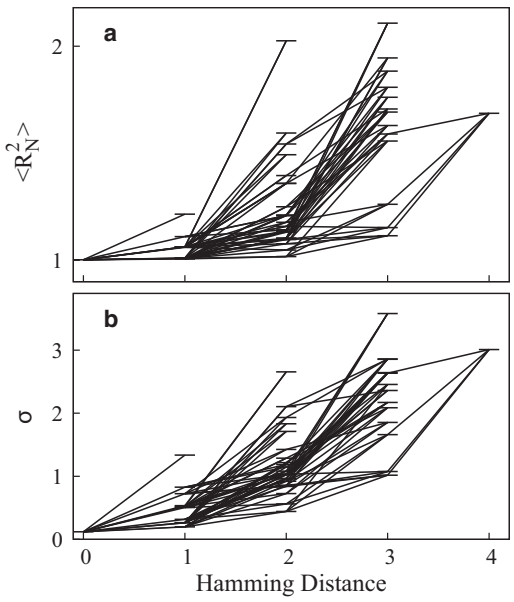

FIGURE 4 Generalization of the superfunnel paradigm. Funnel-like organization of variation of $\langle R_N^2 \rangle$ (*a*) and σ (*b*) with Hamming distance (number of single-point mutations) from the prototype sequence. Results shown are for the 48 $g = 1$ sequences in the HP model neutral net described in the text. Horizontal lines indicate the sequences' $\langle R_N^2 \rangle$ and σ-values computed at $\varepsilon = -5$. Inclined lines connect pairs of sequences that are single-point mutants of each other, as in the original superfunnel drawing in Fig. 2a of Bornberg-Bauer and Chan (9). The relationship between $\langle R_N^2 \rangle$ and σ for the sequences here and those in Fig. 5 are provided by Fig. S3.

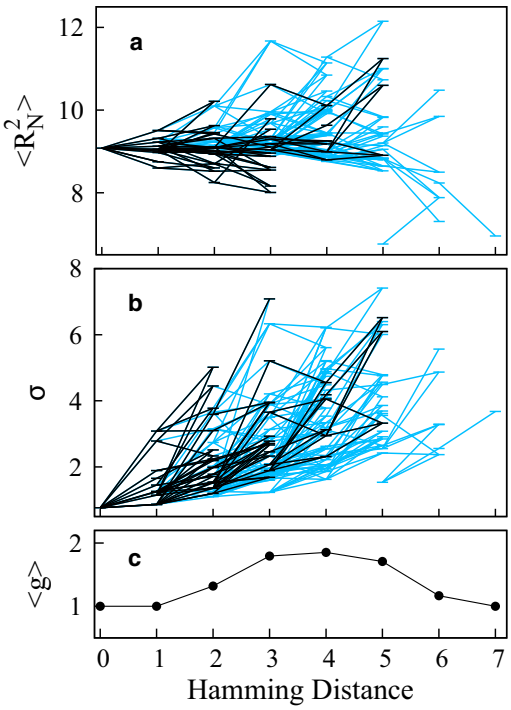

FIGURE 5 Funnel-like organization of variation of $\langle R_N^2 \rangle$ and σ in an extended net for $(R_N^2)_0 = 9$ ($\varepsilon = -5$). Hamming distance is from a prototype sequence with the maximum number of 10 single-point mutants in the net. Data plotted in black in panels *a* and *b* are for the 52 $g = 1$ sequences studied by Fig. S2, *b* and *d*; those plotted in light blue are for 83 $g > 1$ sequences in an extended net containing a total of 135 sequences for which every ground-state conformation has $(R_N^2)_0 = 9$. Mutational connections between $g = 1$ sequences are in black, those involving $g > 1$ sequences are in light blue. The plotting convention in panels *a* and *b* is otherwise the same as that in Fig. 4. (*c*) Average ground-state degeneracy as a function of Hamming distance.

by single-point H $\rightarrow$ P or P $\rightarrow$ H mutations (Figs. 4 and 5). Such networks are of interest as models for studying evolution within a sequence subspace whereby mutations that take sequences outside the network are lethal (9) or, in Maynard Smith's terminology, not "meaningful" (1). In other words, the restrictive conditions for defining the protein network in such models, e.g., the $g = 1$ requirement, are seen as necessary for the survival of the organism in which the protein operates. As an example, Fig. 4 studies the same neutral net of 48 $g = 1$ sequences as that in Fig. S2, *a* and *c*. Biologically, the situation for this net may correspond to one in which the presence of a specific protein structure in sufficiently high concentration (which would not be possible if $g > 1$) is necessary for survival. As another example, Fig. 5 studies an extended net of sequences with a uniform ground-state $(R_N^2)_0 = 9$. This situation may correspond to one in which a high concentration of structures possessing $\langle R_N^2 \rangle$ values within a narrow range around $(R_N^2)_0$ is necessary for survival. For each net, we identify a prototype sequence as the sequence with maximum mutational stability in that it has the maximum number of single-point mutants within the given net (9).

Under conditions favoring folding, the organizations of both $\langle R_N^2 \rangle$ and σ in Fig. 4 resemble the superfunnel pattern of native stabilities for the same neutral net (Fig. 2a of Bornberg-Bauer and Chan (9)). The role of the prototype sequence in a sequence-space superfunnel for evolution is analogous (9) to that of the native structure in a conformational-space funnel for protein folding (27,29,56). Here, we find that the prototype sequence is also the sequence that has the minimum $\langle R_N^2 \rangle$ as well as the minimum fluctuation σ. This trend means that selecting for a mutant with a smaller $\langle R_N^2 \rangle$ would most likely lead to a mutant with a reduced σ as well, and vice versa. Essentially all 99 mutational connections in Fig. 6 have positive slopes and thus are funnel-like (9,13,15): The number of connections accompanying a decrease in Hamming distance with a decrease in $\langle R_N^2 \rangle$ and σ are, respectively, 97 and 99.

Fig. 5 shows the largest net of $g = 1$ sequences with $(R_N^2)_0 = 9$ and its extension to include $g > 1$ sequences with the same uniform $(R_N^2)_0 = 9$. The $g = 1$ net covers two different ground-state conformations, whereas the entire extended net covers a total of 11 different ground-state conformations. The prototype sequence has an $\langle R_N^2 \rangle$ value very close to 9 (Fig. 5 *a*). It also has the minimum σ among the sequences in this net (Fig. 5 *b*). The population of $g > 1$ sequences is concentrated in the middle range of Hamming distances from the prototype sequence (Fig. 5 *c*). As expected from the two-sided funnel patterns in Fig. 3, *b* and *d*, Fig. 5 *a* shows that as Hamming distance decreases, the $\langle R_N^2 \rangle$ value for the prototype sequence is approached both from above and from below. Both of these tendencies are concomitant with a unified trend of decreasing σ (Fig. 5 *b*). There are a total of 272 mutational connections in the extended net in Fig. 5, 86 of which are between $g = 1$ sequences. We define a funnel-like connection for $\langle R_N^2 \rangle$ as one that accompanies a decrease in Hamming distance with a decrease in the absolute value of the difference between the sequence's $\langle R_N^2 \rangle$ and that of the prototype sequence. Such a connection can have either a positive or a negative slope. A funnel-like connection for σ is defined as for Fig. 4 above and always has a positive slope. Most of the connections in Fig. 5 are funnel-like in this regard: In Fig. 5 *a*, 219:272 = 80.5% of the connections for the

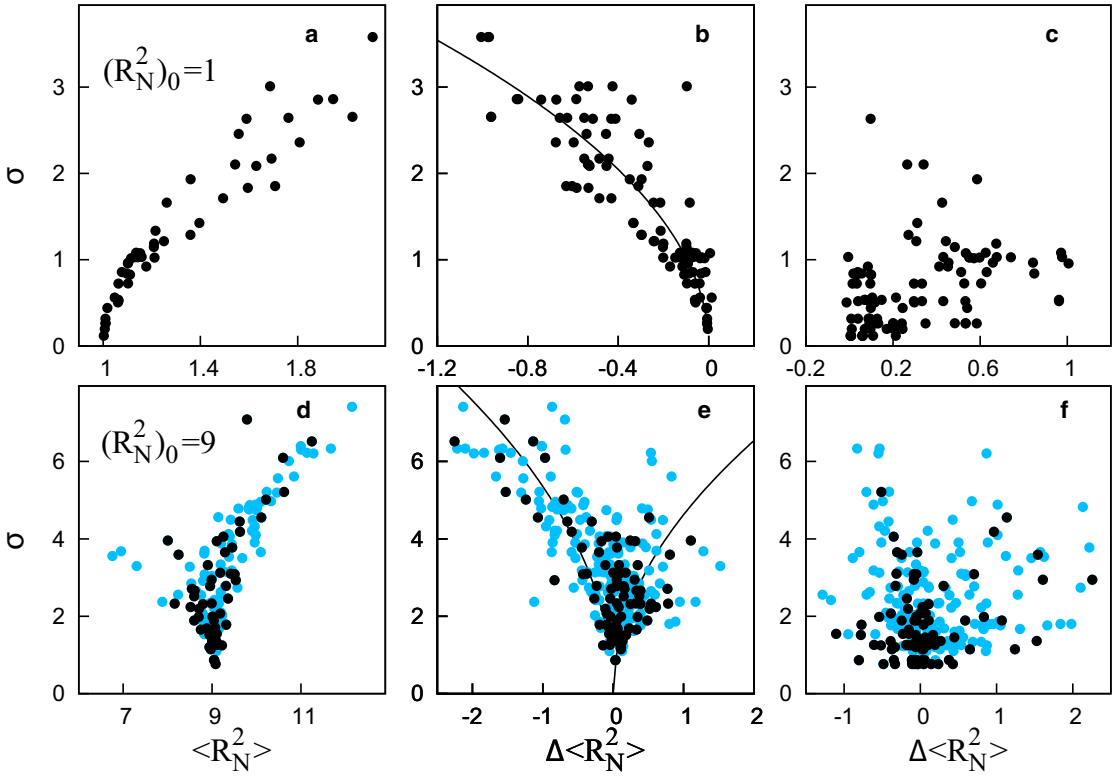

FIGURE 6 Correlation between conformational fluctuation and mutational effect. All results shown are for $\varepsilon = -5$. Panels *a–c* are for the $g = 1$, $(R_N^2)_0 = 1$ sequences in Fig. 4. Panels *d–f* are for the $(R_N^2)_0 = 9$ sequences in Fig. 5, with data involving $g > 1$ sequences plotted in light blue. Each data point in panels *a* and *d* represents a sequence. Each data point in panels *b*, *c*, *e*, and *f* represents a mutation, showing the σ-value of a given sequence and the change, $\Delta\langle R_N^2 \rangle$, in the Boltzmann average $\langle R_N^2 \rangle$ resulting from a mutation on that sequence. Scatter plots in *b* and *e* (*middle column*) and in *c* and *f* (*right column*) are for mutations that change the Hamming distance, respectively, by –1 and by +1 in Fig. 4 or Fig. 5. The curves in panels *b* and *e* show the theoretical expression $\sigma = C\sqrt{|\Delta\langle R_N^2 \rangle|}$ obtained by least-square fitting our model data to $\sigma^2 = C^2|\Delta\langle R_N^2 \rangle|$, with $C = 3.26$ for panel *b* and $C = 5.0$ and 4.2 (in units of lattice bond length), respectively, for the $\Delta\langle R_N^2 \rangle < 0$ and $\Delta\langle R_N^2 \rangle > 0$ data points in panel *e*. The Pearson correlation coefficients are, respectively, $r = 0.85$, 0.82, and 0.28.

extended net, and 70:86 = 81.4% of the connections between $g = 1$ sequences satisfy the above funnel-like criterion. Likewise, in Fig. 5 *b*, 241 (88.6%) of all connections and 82 (95.3%) of the $g = 1$ connections are funnel-like. These percentages of funnel-like connections in Fig. 5 are high but not as high as the 98.0% or 100% for the neutral net in Fig. 4, indicating that there is more sequence-space ruggedness (9) in Fig. 5. Nonetheless, the general superfunnel-like organization in Fig. 5, *a* and *b*, implies that selecting for a mutant with a smaller σ in this net would most likely shorten the Hamming distance from, and reduce the difference in $\langle R_N^2 \rangle$ with the prototype sequence.

Each of the nets in Figs. 4 and 5 thus represents a funnel-like organization centered around a prototype sequence with the least fluctuation σ in R_N^2. It follows from previous analyses of the effect of sequence-space topologies on evolutionary dynamics (9,17,57) that such an organization entails a tendency for the minimum-σ prototype sequence to achieve a higher steady-state evolutionary population than any other sequence in the same net. Are the mutational changes in $\langle R_N^2 \rangle$ in these nets governed by σ as in Eq. 5? We address this question in Fig. 6. Mirroring the overview in Fig. 3, *b* and *d*, the σ-versus-$\langle R_N^2 \rangle$ scatter plot in Fig. 6 *a* for the ground-state $(R_N^2)_0 = 1$ neutral net shapes like a one-sided funnel, whereas that in Fig. 6 *d* for the $(R_N^2)_0 = 9$ net shapes like a two-sided funnel. We next consider mutational changes, $\Delta\langle R_N^2 \rangle$, in the two nets. Here, $\Delta\langle R_N^2 \rangle$ is the $\langle R_N^2 \rangle$ value of the mutant sequence after the mutation minus that of the original sequence before the mutation. Following the formulation of Sato et al. (37), $\Delta\langle R_N^2 \rangle$ is plotted against the σ-value of the original sequence before the mutation.

Fig. 6 separates the mutations into two classes: Those that move toward (namely *b* and *e*); and those that move away (namely, *c* and *f*) from the prototype sequence. Our results show a marked difference between them. Whereas mutations toward the prototype sequence exhibit a reasonable conformity to Eq. 5 (see *fitted curves*), no similarity with Eq. 5 is discernible for mutations that move away from the prototype sequence (largely random scatter) in Fig. 6, *c* and *f*. In Fig. 6, *b* and *e*, the $C \sim b(\Delta a)$ values for the fitted curves are similar although they are not identical. Moreover, in Fig. 6 *e*, the HP model data for the $\Delta\langle R_N^2 \rangle < 0$ mutations fit significantly better with Eq. 5 than those for the $\Delta\langle R_N^2 \rangle > 0$ mutations. Because an overwhelming majority of the mutations that move away from the prototype sequences in Figs. 4 and 5 increase fluctuation σ, the different behaviors for the two classes of mutations in Fig. 6 suggest that the relation in Eq. 5 is more likely to be viable for protein mutations toward more ordered conformations (with smaller σ-values) but less likely to hold for mutations toward more disordered conformations (with larger σ-values). In light of this asymmetry, it is noteworthy that in the original artificial evolution experiment on mutants of a GFP in bacteria (37,38), Eq. 5 was verified for mutations that increase fluorescence intensity but was not tested for mutations that decrease fluorescence intensity.

Thus, although the

$$\sigma = C\sqrt{|\Delta\langle R_N^2 \rangle|}$$

relation stipulated by Eq. 5 fits reasonably with mutations toward the prototype sequences in a net, the data in Fig. 6 also exhibit substantial scatter. To further assess the viability of the idea behind Eq. 5, we now take a global view by considering all 18×2^{18} single-point mutations in our model (Fig. 7). These mutations include, but are not restricted to, those in the relatively small networks in Figs. 4–6. Now, we assess the evolvability of $\langle R_N^2 \rangle$ of every sequence by

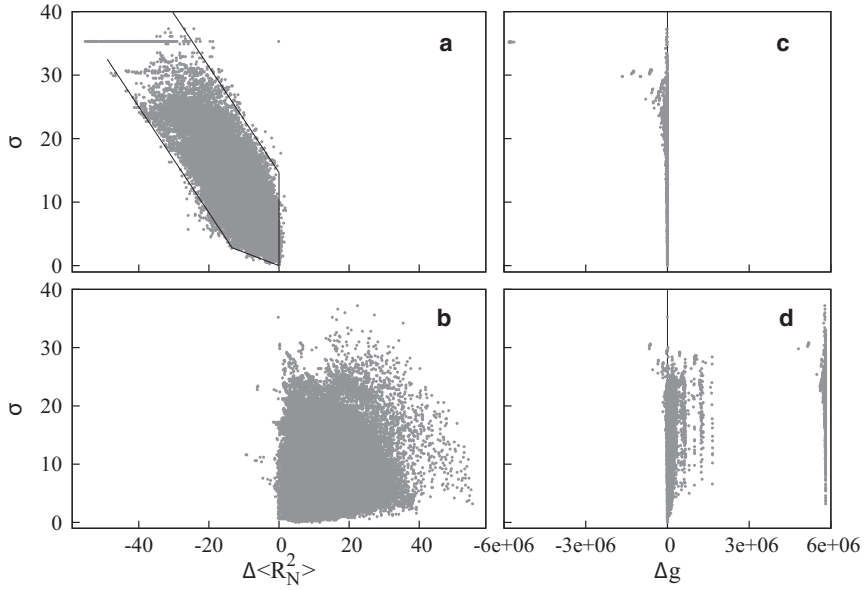

FIGURE 7 Phenotypic fluctuation and evolvability. $\Delta\langle R_N^2 \rangle$ is the change in a sequence's $\langle R_N^2 \rangle$ as a result of a single-point mutation, and Δg is the corresponding change in ground-state degeneracy. Data shown are computed at $\varepsilon = -5$ for all 2^{18} sequences in the model. The scatter plots *a* and *c* are for mutations that achieve the largest possible decreases (steepest descent) in $\langle R_N^2 \rangle$; scatter plots *b* and *c* are for mutations that achieve the largest possible increases (steepest ascent) in $\langle R_N^2 \rangle$. The lines in panel *a* are the approximate boundaries of the distribution discussed in the text. In panels *c* and *d*, mutations that cause no change in ground-state degeneracy are marked by the vertical line at $\Delta g = 0$. In panel *a*, the data points lining up horizontally at $\sigma = 35.4$ are for the $h_N = 0$ sequences. Note that some $h_N = 1$ sequences have σ-values larger than that of the $h_N = 0$ sequences; e.g., an $h_N = 1$ sequence that allows an HH contact between positions 2 and 17 has $\sigma = 37.3$.

performing all 18 possible H → P or P → H substitutions on it to identify the mutation that leads to the largest possible decrease in $\langle R_N^2 \rangle$ (minimum $\Delta\langle R_N^2 \rangle$, steepest descent) and the mutation that leads to the largest increase in $\langle R_N^2 \rangle$ (maximum $\Delta\langle R_N^2 \rangle$, steepest ascent). The resulting scatter plots of σ with these minimum and maximum $\Delta\langle R_N^2 \rangle$ values for all sequences are shown, respectively, in Fig. 7, *a* and *b*.

The two scatter plots are dramatically different. Because of how the mutations are chosen above, it is not surprising that $\Delta\langle R_N^2 \rangle < 0$ for almost all steepest-descent mutations and $\Delta\langle R_N^2 \rangle > 0$ for almost all steepest-ascent mutations. What is striking, however, is that a correlation between σ and $\Delta\langle R_N^2 \rangle$ exists for the steepest-descent mutations in Fig. 7 *a* but no meaningful correlation is observed for the steepest-ascent mutations in Fig. 7 *b*. This asymmetry between Fig. 7, *a* and *b* is similar to that in Fig. 6 between the two classes of mutations moving in opposite directions with respect to the prototype sequence. A reason for the similar trends may be that in Fig. 6, $\langle R_N^2 \rangle$ decreases for most mutations toward the prototype sequence whereas $\langle R_N^2 \rangle$ increases for most mutations away from the prototype sequence, even though the steepest-descent and steepest-ascent mutations in Fig. 7 were not constructed with respect to any prototype sequence.

For the steepest-descent mutations in Fig. 7 *a*, large decreases in $\langle R_N^2 \rangle$ are possible only for sequences with large fluctuations in R_N^2. Fig. 7 *a* shows clearly that the magnitude of the mutational change $\Delta\langle R_N^2 \rangle$ is limited by σ of the original sequence. The trend may be summarized, very roughly, by the inequalities $\Delta\langle R_N^2 \rangle > -4.8\sigma$ for $\sigma \lesssim 3$ ($\Delta\langle R_N^2 \rangle \gtrsim -13$) and $\Delta\langle R_N^2 \rangle > -1.2\sigma - 10$ for $\sigma \gtrsim 3$ ($\Delta\langle R_N^2 \rangle \lesssim -13$). (Note that $\Delta\langle R_N^2 \rangle$ and σ have the same [length]^2 unit.) However, Fig. 7 *a* also shows that having a large σ, per se, is not sufficient to guarantee a sequence's ability to achieve a large mutational decrease in $\langle R_N^2 \rangle$. For $\sigma \lesssim 15$, $\Delta\langle R_N^2 \rangle$ for different sequences may take virtually any value within a range from $\Delta\langle R_N^2 \rangle \approx 0$ to the approximate lower bounds delineated above. For $\sigma \gtrsim 15$, the largest $\Delta\langle R_N^2 \rangle$ values become negative, roughly satisfying the inequality $\Delta\langle R_N^2 \rangle < -1.2\sigma + 17.6$. This trend indicates that for each of these sequences with larger fluctuations in R_N^2, at least one mutant can bring about an appreciable decrease in $\langle R_N^2 \rangle$, although the magnitude of that decrease may still be substantially smaller than $-\Delta\langle R_N^2 \rangle \approx 1.2\sigma + 10$.

In stark contrast, the steepest-ascent mutations in Fig. 7 *b* do not show any of the above-described or other correlative features between σ and $\Delta\langle R_N^2 \rangle$. This lack of correlation means that a sequence's fluctuation in R_N^2 alone is not predictive of its evolvability to another sequence with a larger $\langle R_N^2 \rangle$. Fig. 7, *c* and *d*, show further that the changes in ground-state degeneracy g for the steepest-descent and steepest-ascent mutations are also very different. Steepest-descent mutations tend to decrease ground-state degeneracy, and in this respect making the sequence more similar to natural globular proteins. For the data in Fig. 7 *c*, the median and average values of Δg are, respectively, –28 and -2.6×10^4. This trend means that selecting for a mutant with smaller $\langle R_N^2 \rangle$ would likely yield, as a byproduct, a mutant that also has fewer ground-state conformations. Such coevolution of various properties of the same sequence may be viewed as a single-genotype analog of the hitchhiking effect (58). Steepest-ascent mutations, however, tend to lead to large increases in ground-state degeneracy. For the data in Fig. 7 *d*, the median and average values of Δg are, respectively, +282 and $+2.2 \times 10^5$. It will be instructive to explore how the lack of correlation between fluctuation and evolvability among the steepest-ascent mutations in Fig. 7 *b* might be related to the more rugged conformational landscapes (27–29) entailed by the $\Delta g > 0$ increases in Fig. 7 *d*.

To ascertain the robustness of the trend we observed, we have also investigated the relationship between σ and $\Delta\langle R_N^2 \rangle$ for all single-point mutations at $\varepsilon = -2$, –3, and –4 (detailed results not shown). In addition, we have also studied single-point mutations with steepest descents (maximum decreases) and steepest ascents (maximum increases) in fluctuation σ (Fig. S4). In all of these other studies, a level of correlation between σ and $\Delta\langle R_N^2 \rangle$ similar to that in Fig. 7 *a* was observed for mutations with steepest descents in either $\Delta\langle R_N^2 \rangle$ or σ. However, as in Fig. 7 *b*, no correlation was observed for corresponding mutations with steepest ascents.

We have also generalized the interpretation of Eq. 5 to consider the relationship between $\Delta\langle R_N^2 \rangle$ and the standard deviation of R_N^2 of the mutated sequence (denoted as σ(final)) instead of σ for the original sequence. For the results in Fig. 6, *b*, *c*, *e*, and *f*, changing the variable σ to σ(final) amounts to swapping Fig. 6 *b* with *c*, swapping Fig. 6 *e* with *f*, and changing the sign of $\Delta\langle R_N^2 \rangle$. It is clear from the existing results in Fig. 6 that after these changes the same data would indicate a correlation of σ(final) with $\Delta\langle R_N^2 \rangle$ for mutations away from the prototype sequence (especially those leading to an increase in both σ and $\langle R_N^2 \rangle$) but not for mutations toward the prototype sequence. For the steepest-descent and steepest-ascent results in Fig. 7 for all sequences, although the scatter plots for σ(final) (Fig. S5) are not exact mirror images of those for σ, they nonetheless show a degree of correlation of σ(final) with steepest-ascent mutations toward larger $\langle R_N^2 \rangle$ (Fig. S5 *b*) but not with steepest-descent mutations toward smaller $\langle R_N^2 \rangle$ (Fig. S5 *a*). This behavior thus follows a trend similar to that for σ(final) deduced above for the $(R_N^2)_0 = 1$ and $(R_N^2)_0 = 9$ nets in Fig. 6. Therefore, our results suggest in general that for a pair of sequences that differ by one single-point substitutive mutation, the magnitude of the difference $\Delta\langle R_N^2 \rangle$ between the sequences of the pair tends to correlate with the larger but not with the smaller of the two σ-values for the two sequences.

Taken together, our results indicate consistently that conformational fluctuation of a sequence is correlated with evolvability of that sequence toward a mutant with decreased conformational fluctuation (smaller σ, more order); but the

extent of conformational fluctuation by itself is not predictive of a sequence's evolvability toward a mutant with increased conformational fluctuation (larger σ, less order). Therefore, it appears that the fluctuation-response idea of Sato et al. (37) is applicable, with caveats, for protein mutations toward more ordered conformational states; but the idea may not be so applicable for protein mutations toward more disordered conformational states.

CONCLUSIONS

Using a simple exact model of the mapping between protein sequence and structure, we have now characterized several statistical mechanical aspects of the relationship between evolvability and single-genotype phenotypic fluctuation, which was modeled as conformational fluctuation of a single model protein sequence. Biological functions of proteins are often related to conformational fluctuations (59). For single-domain cooperatively folding proteins (51), native conformational fluctuation is often small (60), perhaps to guard against harmful aggregation. For larger proteins, however, conformational flexibility is often critical for function (59). Here, we have verified that a simple formula proposed by Sato et al. to relate fluctuation and response (37) can indeed provide a semiquantitative rationalization for the correlation between evolvability and fluctuation among the mutations that move sequences toward the prototype sequence of a superfunnel (9) (Fig. 6). At the same time, the present explicit-chain modeling has also revealed subtle, unanticipated aspects of the fluctuation-evolvability relationship. Our results suggest that significant single-genotype phenotypic fluctuation is, in general, a likely requirement for a sequence's evolvability to other sequences with less phenotypic fluctuations. However, not every sequence with significant single-genotype phenotypic fluctuation is highly evolvable. Although single-genotype phenotypic fluctuation is an indicator of evolvability toward more conformational order, single-genotype phenotypic fluctuation per se is not predictive of evolvability toward less conformational order. These asymmetric behaviors deserve more future theoretical attention; and it would be extremely interesting to inquire experimentally whether a similar asymmetry exists in the evolution of real polypeptides.

SUPPORTING MATERIAL

Five figures are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(10)00326-7.

We thank Tobias Sikosek for a critical reading of an earlier version of this article.

This work was supported by Canadian Institutes of Health Research grant No. MOP-84281 (to H.S.C., who holds a Canada Research Chair in Proteomics, Bioinformatics, and Functional Genomics). D.V. was supported through a "PPP Travel Grant Canada" from the German Academic Exchange Service (DAAD).

REFERENCES

1. Maynard Smith, J. 1970. Natural selection and the concept of a protein space. *Nature*. 225:563–564.
2. Lau, K. F., and K. A. Dill. 1990. Theory for protein mutability and biogenesis. *Proc. Natl. Acad. Sci. USA*. 87:638–642.
3. Lipman, D. J., and W. J. Wilbur. 1991. Modeling neutral and selective evolution of protein folding. *Proc. R. Soc. Lond. B. Biol. Sci*. 245:7–11.
4. Shortle, D., H. S. Chan, and K. A. Dill. 1992. Modeling the effects of mutations on the denatured states of proteins. *Protein Sci*. 1:201–215.
5. Chan, H. S., and K. A. Dill. 1996. Comparing folding codes for proteins and polymers. *Proteins*. 24:335–344.
6. Li, H., R. Helling, …, N. Wingreen. 1996. Emergence of preferred structures in a simple model of protein folding. *Science*. 273:666–669.
7. Abkevich, V. I., A. M. Gutin, and E. I. Shakhnovich. 1996. How the first biopolymers could have evolved. *Proc. Natl. Acad. Sci. USA*. 93:839–844.
8. Govindarajan, S., and R. A. Goldstein. 1997. Evolution of model proteins on a foldability landscape. *Proteins*. 29:461–466.
9. Bornberg-Bauer, E., and H. S. Chan. 1999. Modeling evolutionary landscapes: mutational stability, topology, and superfunnels in sequence space. *Proc. Natl. Acad. Sci. USA*. 96:10689–10694.
10. Blackburne, B. P., and J. D. Hirst. 2001. Evolution of functional model proteins. *J. Chem. Phys*. 115:1935–1942.
11. Cui, Y., W. H. Wong, …, H. S. Chan. 2002. Recombinatoric exploration of novel folded structures: a heteropolymer-based model of protein evolutionary landscapes. *Proc. Natl. Acad. Sci. USA*. 99:809–814.
12. Xia, Y., and M. Levitt. 2002. Roles of mutation and recombination in the evolution of protein thermodynamics. *Proc. Natl. Acad. Sci. USA*. 99:10382–10387.
13. Xia, Y., and M. Levitt. 2004. Funnel-like organization in sequence space determines the distributions of protein stability and folding rate preferred by evolution. *Proteins*. 55:107–114.
14. Bloom, J. D., J. J. Silberg, …, F. H. Arnold. 2005. Thermodynamic prediction of protein neutrality. *Proc. Natl. Acad. Sci. USA*. 102: 606–611.
15. Wroe, R., E. Bornberg-Bauer, and H. S. Chan. 2005. Comparing folding codes in simple heteropolymer models of protein evolutionary landscape: robustness of the superfunnel paradigm. *Biophys. J*. 88:118–131.
16. Zeldovich, K. B., P. Chen, …, E. I. Shakhnovich. 2007. A first-principles model of early evolution: Emergence of gene families, species, and preferred protein folds. *PLOS Comput. Biol*. 3:1224–1238.
17. Chan, H. S., and E. Bornberg-Bauer. 2002. Perspectives on protein evolution from simple exact models. *Appl. Bioinformatics*. 1:121–144.
18. Xia, Y., and M. Levitt. 2004. Simulating protein evolution in sequence and structure space. *Curr. Opin. Struct. Biol*. 14:202–207.
19. Ancel, L. W., and W. Fontana. 2000. Plasticity, evolvability, and modularity in RNA. *J. Exp. Zoo. (Mol. Dev. Evol.)*. 288:242–283.
20. James, L. C., and D. S. Tawfik. 2003. Conformational diversity and protein evolution—a 60-year-old hypothesis revisited. *Trends Biochem. Sci*. 28:361–368.
21. Aharoni, A., L. Gaidukov, …, D. S. Tawfik. 2005. The 'evolvability' of promiscuous protein functions. *Nat. Genet*. 37:73–76.
22. Amitai, G., R. D. Gupta, and D. S. Tawfik. 2007. Latent evolutionary potentials under the neutral mutational drift of an enzyme. *HFSP J*. 1:67–78.
23. Wroe, R., H. S. Chan, and E. Bornberg-Bauer. 2007. A structural model of latent evolutionary potentials underlying neutral networks in proteins. *HFSP J*. 1:79–87.
24. Bloom, J. D., P. A. Romero, …, F. H. Arnold. 2007. Neutral genetic drift can alter promiscuous protein functions, potentially aiding functional evolution. *Biol. Direct*. 2:17.
25. Depristo, M. A. 2007. The subtle benefits of being promiscuous: adaptive evolution potentiated by enzyme promiscuity. *HFSP J*. 1:94–98.

26. Tokuriki, N., and D. S. Tawfik. 2009. Protein dynamism and evolvability. *Science*. 324:203–207.
27. Bryngelson, J. D., J. N. Onuchic, …, P. G. Wolynes. 1995. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins*. 21:167–195.
28. Dill, K. A., S. Bromberg, …, H. S. Chan. 1995. Principles of protein folding—a perspective from simple exact models. *Protein Sci*. 4: 561–602.
29. Dill, K. A., and H. S. Chan. 1997. From Levinthal to pathways to funnels. *Nat. Struct. Biol*. 4:10–19.
30. Badasyan, A., Z. Liu, and H. S. Chan. 2009. Interplaying roles of native topology and chain length in marginally cooperative and noncooperative folding of small protein fragments. *Int. J. Quantum Chem*. 109:3482–3499.
31. Queitsch, C., T. A. Sangster, and S. Lindquist. 2002. Hsp90 as a capacitor of phenotypic variation. *Nature*. 417:618–624.
32. Ito, Y., H. Toyota, …, T. Yomo. 2009. How selection affects phenotypic fluctuation. *Mol. Syst. Biol*. 5:264.
33. Tompa, P. 2002. Intrinsically unstructured proteins. *Trends Biochem. Sci*. 27:527–533.
34. Mittag, T., and J. D. Forman-Kay. 2007. Atomic-level characterization of disordered protein ensembles. *Curr. Opin. Struct. Biol*. 17:3–14.
35. Boehr, D. D., R. Nussinov, and P. E. Wright. 2009. The role of dynamic conformational ensembles in biomolecular recognition. *Nature Chem. Biol*. 5:789–796, Correction 5:954.
36. Kirschner, M., and J. Gerhart. 1998. Evolvability. *Proc. Natl. Acad. Sci. USA*. 95:8420–8427.
37. Sato, K., Y. Ito, …, K. Kaneko. 2003. On the relation between fluctuation and response in biological systems. *Proc. Natl. Acad. Sci. USA*. 100:14086–14090.
38. Yomo, T., Y. Ito, …, K. Kaneko. 2005. Phenotypic fluctuation rendered by a single genotype and evolutionary rate. *Physica A*. 350:1–5.
39. Keefe, A. D., and J. W. Szostak. 2001. Functional proteins from a random-sequence library. *Nature*. 410:715–718.
40. Hayashi, Y., H. Sakata, …, T. Yomo. 2003. Can an arbitrary sequence evolve towards acquiring a biological function? *J. Mol. Evol*. 56: 162–168.
41. Ito, Y., T. Kawama, …, T. Yomo. 2004. Evolution of an arbitrary sequence in solubility. *J. Mol. Evol*. 58:196–202.
42. Pathria, R. K. 1980. Statistical Mechanics. Pergamon Press, Oxford, UK.
43. Kaneko, K., and C. Furusawa. 2006. An evolutionary relationship between genetic variation and phenotypic fluctuation. *J. Theor. Biol*. 240:78–86.
44. Earl, D. J., and M. W. Deem. 2004. Evolvability is a selectable trait. *Proc. Natl. Acad. Sci. USA*. 101:11531–11536.
45. Chan, H. S. 2000. Modeling protein density of states: additive hydrophobic effects are insufficient for calorimetric two-state cooperativity. *Proteins*. 40:543–571.
46. Schuler, B., and W. A. Eaton. 2008. Protein folding studied by single-molecule FRET. *Curr. Opin. Struct. Biol*. 18:16–26.
47. Yamauchi, A., T. Yomo, …, I. Urabe. 1998. Characterization of soluble artificial proteins with random sequences. *FEBS Lett*. 421:147–151.
48. Chan, H. S. 1999. Folding alphabets. *Nat. Struct. Biol*. 6:994–996.
49. Irbäck, A., and E. Sandelin. 2000. On hydrophobicity correlations in protein chains. *Biophys. J*. 79:2252–2258.
50. Stout, M., J. Bacardit, …, J. Blazewicz. 2006. From HP lattice models to real proteins: coordination number prediction using learning classifier systems. *In* Applications of Evolutionary Computing, Proceedings; Lecture Notes in Computer Science., Vol. 3907. F. Rothlauf, editor. Springer, Berlin/Heidelberg. 208–220.
51. Chan, H. S., S. Shimizu, and H. Kaya. 2004. Cooperativity principles in protein folding. *Methods Enzymol*. 380:350–379.
52. Noivirt-Brik, O., R. Unger, and A. Horovitz. 2009. Analyzing the origin of long-range interactions in proteins using lattice models. *BMC Struct. Biol*. 9:4.
53. Chan, H. S., and Z. Zhang. 2009. Liaison amid disorder: non-native interactions may underpin long-range coupling in proteins. *J. Biol*. 8:27.
54. Noivirt-Brik, O., A. Horovitz, and R. Unger. 2009. Trade-off between positive and negative design of protein stability: from lattice models to real proteins. *PLOS Comput. Biol*. 5:e1000592.
55. Chan, H. S., H. Kaya, and S. Shimizu. 2002. Computational methods for protein folding: scaling a hierarchy of complexities. *In* Current Topics in Computational Molecular Biology. T. Jiang, Y. Xu, and M. Zhang, editors. MIT Press, Cambridge, MA. 403–447.
56. Leopold, P. E., M. Montal, and J. N. Onuchic. 1992. Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. USA*. 89:8721–8725.
57. van Nimwegen, E., J. P. Crutchfield, and M. Huynen. 1999. Neutral evolution of mutational robustness. *Proc. Natl. Acad. Sci. USA*. 96:9716–9720.
58. Maynard Smith, J., and J. Haigh. 1974. The hitchhiking effect of a favorable gene. *Genet. Res. Camb*. 23:23–35.
59. Gunasekaran, K., B. Y. Ma, and R. Nussinov. 2004. Is allostery an intrinsic property of all dynamic proteins? *Proteins*. 57:433–443.
60. Petsko, G. A., and D. Ringe. 2004. Protein Structure and Function. New Science Press, Waltham, MA.