PROTEINS: Structure, Function, and Genetics 49:403–412 (2002)

Designability of Protein Structures: A Lattice-Model Study using the Miyazawa-Jernigan Matrix

Hao Li, Chao Tang,* and Ned S. Wingreen
NEC Research Institute, Princeton, New Jersey

ABSTRACT **We study the designability of all compact 3 × 3 × 3 and 6 × 6 lattice-protein structures using the Miyazawa–Jernigan (MJ) matrix. The designability of a structure is the number of sequences that design the structure, i.e., sequences that have that structure as their unique lowest-energy state. Previous studies of hydrophobic-polar (HP) models showed a wide distribution of structure designabilities. Recently, questions were raised concerning the use of a two-letter (HP) code in such studies. Here, we calculate designabilities using all 20 amino acids, with empirically determined interaction potentials (MJ matrix) and compare with HP model results. We find good qualitative agreement between the two models. In particular, highly designable structures in the HP model are also highly designable in the MJ model—and vice versa—with the associated sequences having enhanced thermodynamic stability. Proteins 2002;49:403–412.**

Key words: protein folding; designability; lattice models; statistical potential; alphabet; Miyazawa–Jernigan matrix

INTRODUCTION

The sequences and structures of natural proteins form special classes among all possible sequences and structures. A natural protein sequence has, as its native state, a unique global minimum of free energy that is well separated in energy from other misfolded states[1]—a property not typically shared by random sequences of amino acids. Protein structures in general possess striking geometric regularities,[2,3] characterized by preferred secondary structures and motifs[4] and often by tertiary symmetries. It has been noted that a large number of proteins are accounted for by a small number of folds[5,6] or superfolds.[7] Several authors have proposed possible physical mechanisms behind nature's selection of protein folds. Finkelstein and coworkers argued that certain motifs are easier to stabilize and thus more common either because they have lower structural (e.g., bending) energies or because they have unusual energy spectra over random sequences.[8–10] Yue and Dill observed in a lattice hydrophobic-polar (HP) model that protein-like folds are associated with sequences that have a minimal number of degenerate lowest-energy states.[11] Govindarajan and Goldstein suggested that the evolutionary pressure on protein structures is to fold fast. They studied the "foldability" of structures in a lattice model and found that the foldability, optimized over sequences, varies from structure to structure.[12,13] They further argued that structures with larger optimal foldability should tolerate more sequences and be more robust to mutations.

More recently, this issue has been investigated from the perspective of "designability."[14–17] The designability of a structure is defined as the number of sequences that can design the structure, that is, sequences that possess the structure as their unique lowest-energy state. Li et al. studied the designability of all compact structures in HP lattice models of sizes 3 × 3 × 3 and 6 × 6.[14] They found that structures differ drastically in their designabilities and that a small number of structures emerge with designabilities much larger than the average. The sequences associated with these highly designable structures are also thermodynamically more stable[14,15] and fold much faster than typical sequences.[16] Further, these structures possess regular secondary structures and motifs and, in some cases, global symmetries.[18] Studies of designability for a larger lattice model (4 × 3 × 3)[19] and for an off-lattice model[20] yielded similar overall results.

However, most studies of designability have been based on HP-type models. It is a legitimate concern to ask how the designability of structures depends on interaction potentials and on the alphabet size (the number of different kinds of amino acids in the model).[21–24] In a recent lattice-model study, it was concluded that the designability of a structure depends sensitively on the size of the alphabet in the model—in particular, structures that were highly designable for a two-letter alphabet were found not especially designable with a many-letter alphabet.[23] In this article, we study the designability of all compact structures in two lattice models using all 20 amino acid types, with interactions given by the Miyazawa–Jernigan (MJ) matrix.[25] We compare the results with those of Ref. 14, which were obtained using only two types (H and P) of amino acids. We find that the designability of a structure is *not* sensitive to the alphabet size when the hydrophobic interaction is included in the potential.

H. Li's present address is the Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, CA 94143.

*Correspondence to: Chao Tang, NEC Research Institute, 4 Independence Way, Princeton, NJ 08540. E-mail: tang@research.nj.nec.com

Received 15 March 2002; Accepted 8 July 2002

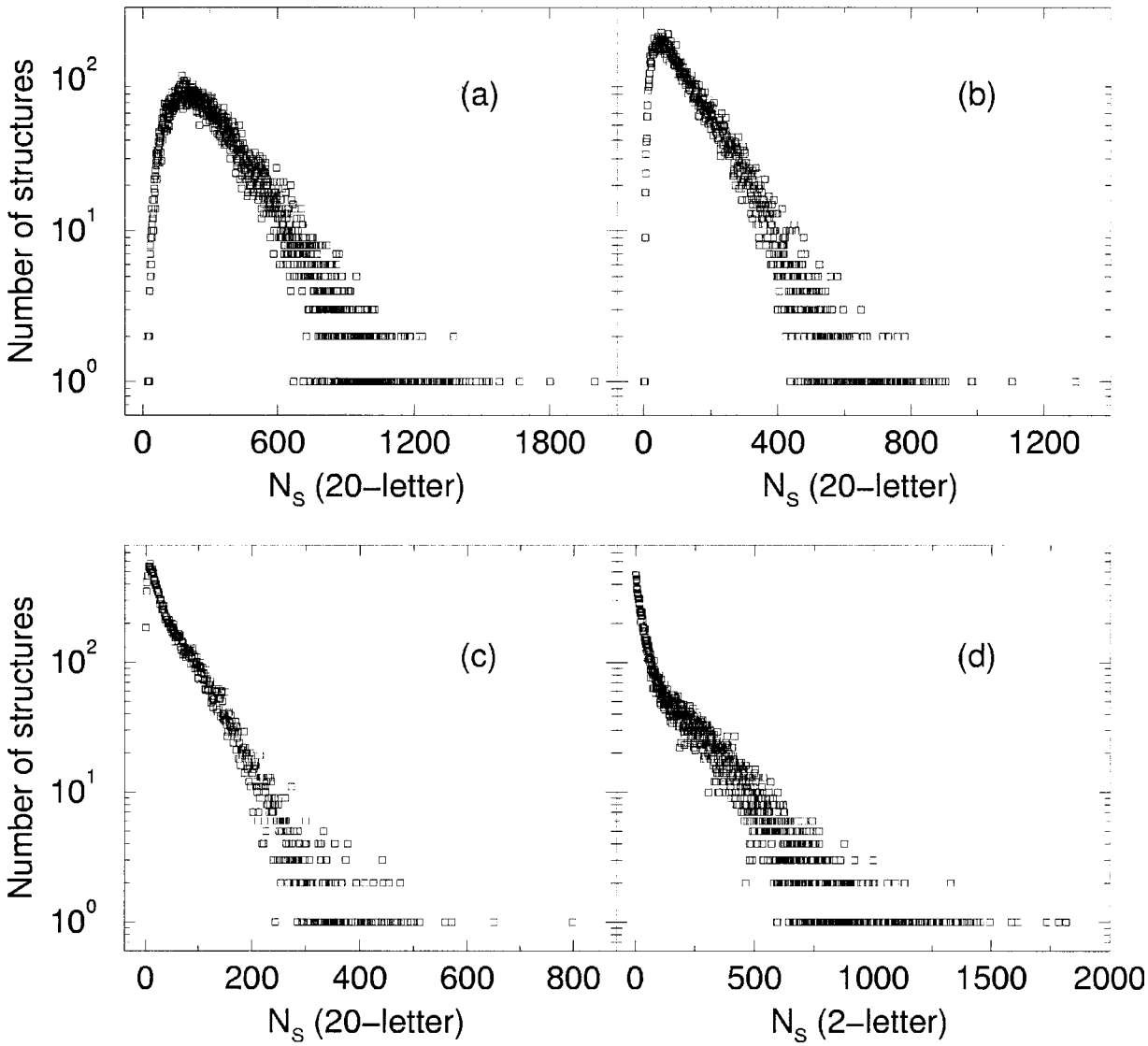

Fig. 1. Histograms of designability N_S for the 6 × 6 system for the MJ matrix with gap cutoff (**a**) $g_c = 0.0$, (**b**) $g_c = 0.4$, (**c**) $g_c = 0.8$, and (**d**) for the HP model. Results for the MJ matrix were obtained by using 9,095,000 random sequences.

MODELS AND METHODS

We study a lattice-protein model in both 2D and 3D. In the 2D case, the model is a self-avoiding chain of $N = 36$ residues on a square lattice. We consider only the maximally compact structures, i.e., structures contained in a 6×6 square. We also study a 2D system with a chain length $N = 30$ (6×5). In the 3D case, the chain has a length of $N = 27$ and folds into a maximally compact configuration of $3 \times 3 \times 3$. The study of 3D structures is computationally limited to short chains ($N \approx 30$) with corresponding surface-to-core ratios of approximately 2:1, much larger than that of typical natural proteins. 2D models can achieve a more realistic 1:1 surface-to-core ratio with manageable chain lengths, at the risk of introducing other unphysical effects due to the dimensional reduction. Thus, it is more convincing to draw conclusions based on a combined study of 2D and 3D models.

In the model, a sequence of length N is specified by the residue type μ_i, ($i = 1, 2, \ldots, N$) along the chain, where μ is one of the 20 natural amino acids. A structure is specified by the position $\mathbf{r}_i$, ($i = 1, 2, \ldots, N$) of each residue along the chain. The energy for a sequence folded into a structure is taken to be the sum of the contact energies, that is

$$E = \sum_{i<j} e_{\mu_i\mu_j} \Delta(r_i - r_j)\,, \tag{1}$$

where $e_{\mu_i\mu_j}$ is the contact energy between residue types μ_i and μ_j, and $\Delta(\mathbf{r}_i - \mathbf{r}_j) = 1$ if $\mathbf{r}_i$ and $\mathbf{r}_j$ are adjoining lattice sites with i and j not adjacent along the chain, and $\Delta(\mathbf{r}_i - \mathbf{r}_j) = 0$ otherwise. The contact energies $e_{\mu\nu}$ are taken from the MJ matrix.[25] Note that the water solvent is implicit in the Hamiltonian (1). The energy $e_{\mu\nu}$ of a contact between residue μ and ν is the relative energy with respect to

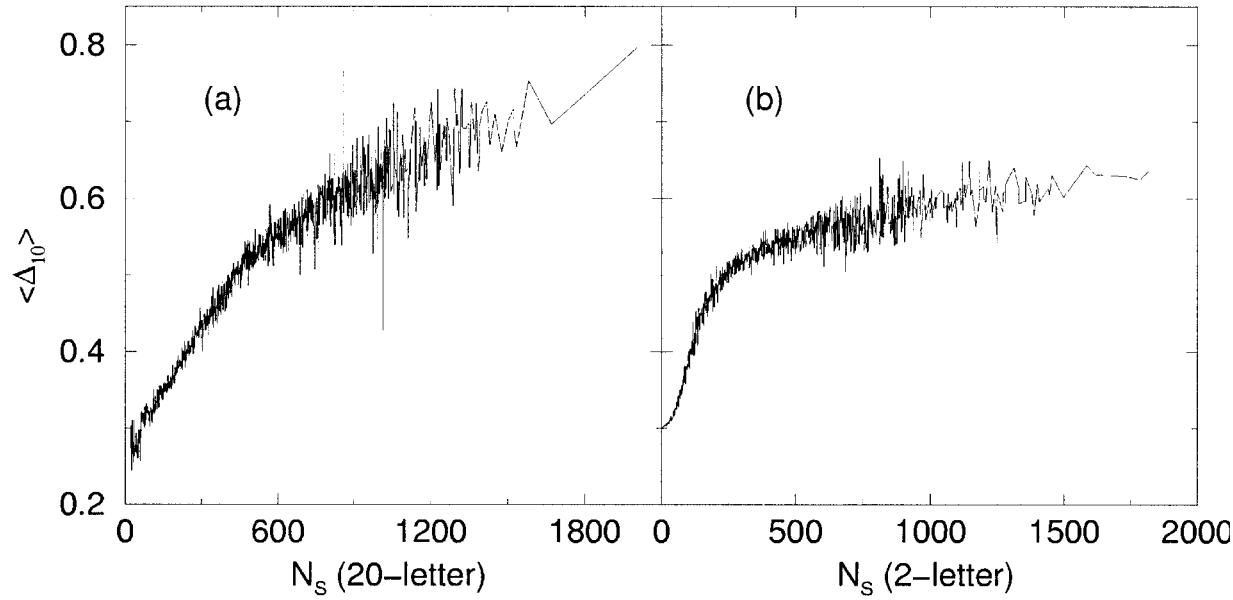

Fig. 2. Average energy gap $\langle \Delta_{10} \rangle$ vs designability N_S for the 6 × 6 system. (**a**) For the MJ matrix. (**b**) For the HP model.

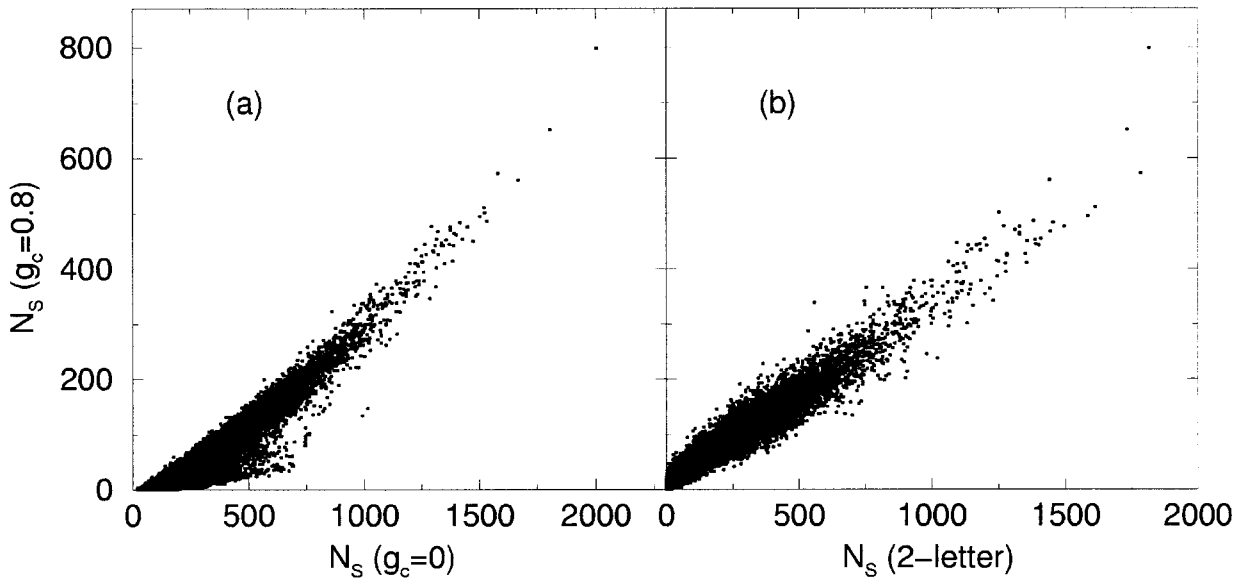

Fig. 3. (**a**) Designability N_S with gap cutoff $g_c = 0.8$ vs N_S with $g_c = 0$ for the MJ model for each 6 × 6 structure. (**b**) N_S with $g_c = 0.8$ for the MJ model vs N_S for the HP model for each 6 × 6 structure. Note that most structures are close to the origin.

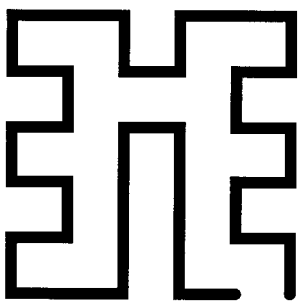

Fig. 4. Most designable structure in the 6 × 6 system.

separated μ and ν in water. So, the residues on the surface of a structure are considered to be in contact with water.

There are several different MJ matrices in Ref. 25. We use matrix e_{ij} (the upper half and diagonal of Table V in Ref. 25). This matrix contains all the contributions to the interaction energy including, in particular, the hydrophobic or solvation contribution. The hydrophobic contribution, although nonspecific, is residue dependent and is the dominant contribution to the MJ matrix e_{ij}.[26] For other MJ matrices in Ref. 25, the hydrophobic contribution has been, to various degrees, removed. Thus, they are not appropriate for folding studies like this one (cf. discussion and conclusion sections). We also used updated versions of the same MJ matrix.[27,28] The results are similar.

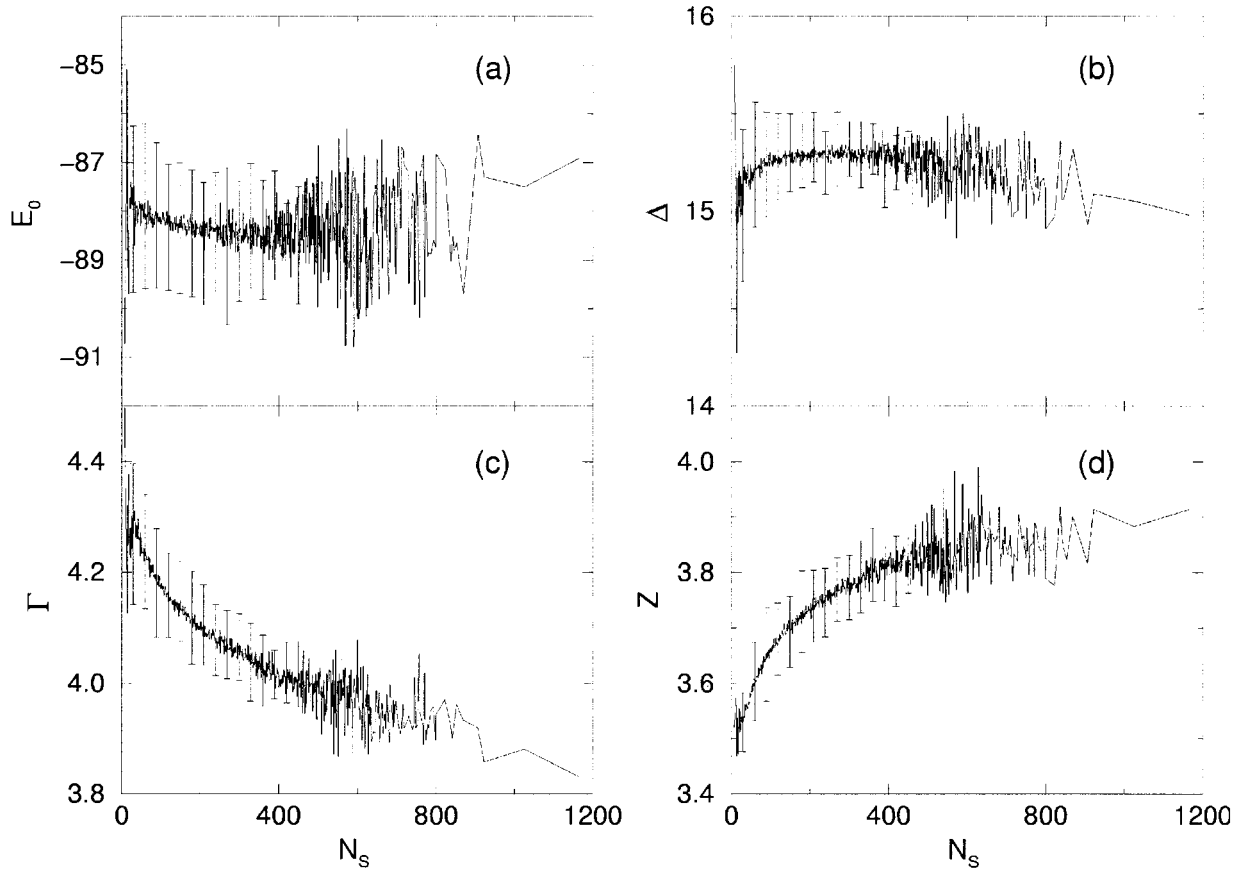

Fig. 5. (**a**) Ground-state energy E_0, (**b**) depth of ground state Δ, (**c**) width of compact spectrum Γ, and (**d**) $Z = \Delta/\Gamma$ vs N_S for the 6 × 6 system with gap cutoff $g_c = 0$. The solid lines are averages for given N_S and the error bars indicate the variances. Data were obtained from 5,100,000 random sequences.

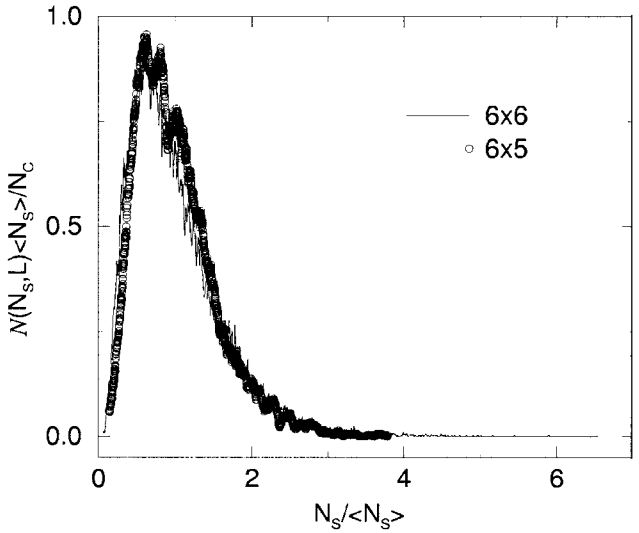

Fig. 6. Scaled histogram vs scaled designability N_S for 6 × 6 and 6 × 5 systems. For a chain of length L, we denote by $\mathcal{N}(N_S,L)$ the number of structures with designability N_S. The average designability is $\langle N_S \rangle$ and N_C is the total number of structures.

In each case, we enumerate all maximally compact self-avoiding structures. We then randomly select a large number of sequences. For each of these sequences, we evaluate its energy on all the structures using Eq. 1. If the sequence has a unique lowest-energy state, or ground state (the criterion of being unique will be defined below), we say the sequence can design the structure and the following quantities are recorded: the ground-state structure, the ground-state energy E_0, the second lowest energy E_1, the depth of the ground state

$$\Delta = \langle E \rangle' - E_0 \,, \tag{2}$$

and the variance of the energy spectrum

$$\Gamma^2 = \langle E^2 \rangle' - \langle E \rangle'^2 \,, \tag{3}$$

where $\langle \cdot \rangle'$ denotes averaging over all compact structures other than the ground state. The quantities Δ, $Z = \Delta/\Gamma$,[29,30] and $\Delta_{10} = E_1 - E_0$[31,32] have been widely used to characterize how protein-like a sequence is because of their correlations with the folding rate.[16,33,34] The ground state of a sequence is said to be unique if for the sequence there are no other structures with energy lower than $E_0 + g_c$, where the gap cutoff g_c is a parameter. We used $g_c = 0$, 0.4, and 0.8 (in the unit of RT at room temperature) in our calculations. After the calculation is completed with all randomly selected sequences, we measure the designability of a structure, N_S, by the number of sequences that design the structure.

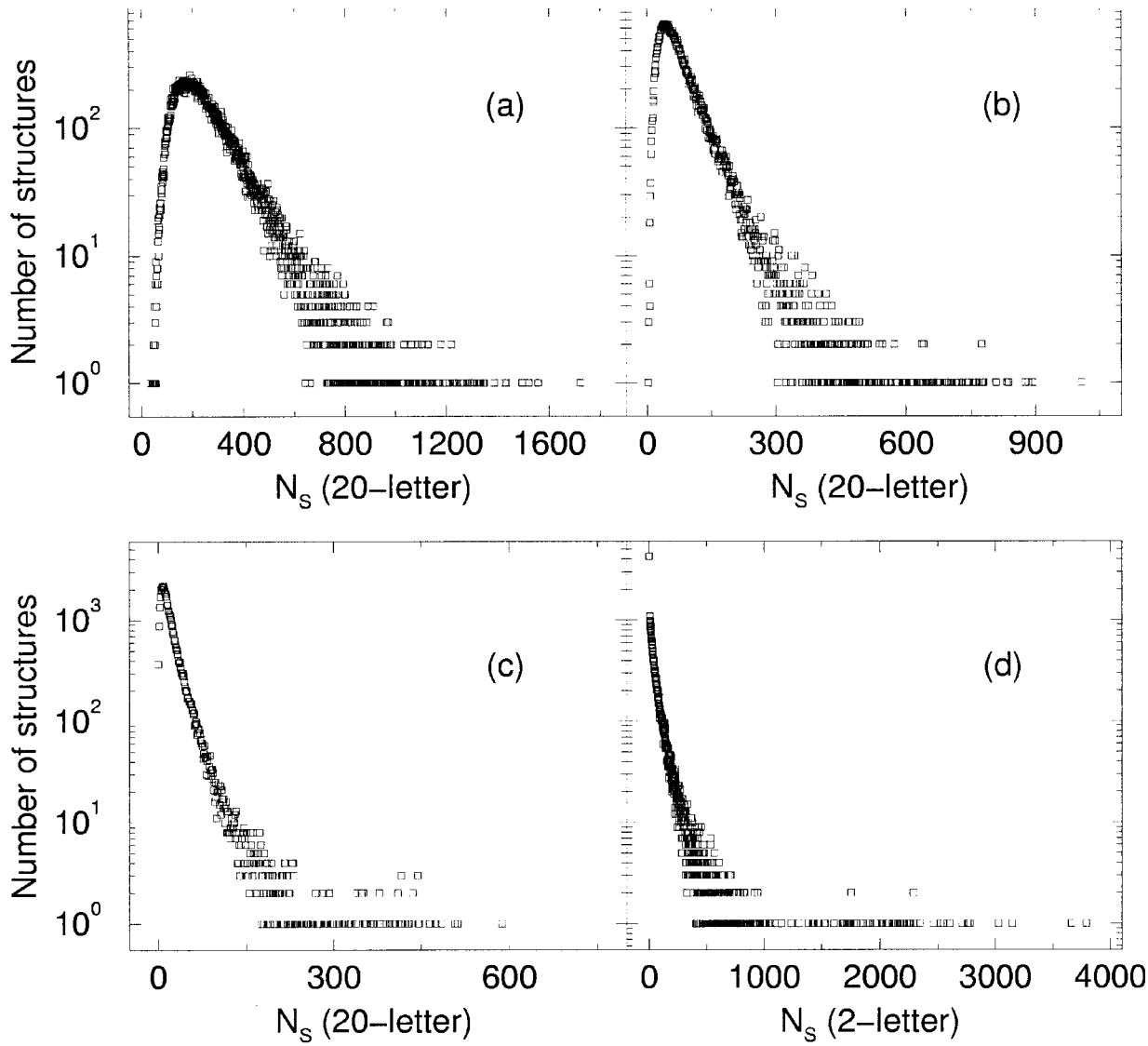

Fig. 7. Histogram of designability N_S for the $3 \times 3 \times 3$ system for the MJ matrix with gap cutoff (**a**) $g_c = 0.0$, (**b**) $g_c = 0.4$, (**c**) $g_c = 0.8$, and (**d**) for the HP model. Results for MJ matrix were obtained by using 13,550,000 random sequences. Results for HP model were obtained by enumerating all 27 sequences.

We compare our results with those of Ref. 14, which were obtained using an HP model. The parameters used in Ref. 14 for Eq. 1 are: $e_{HH} = -2.3$, $e_{HP} = -1$, and $e_{PP} = 0$, which were derived from and can be viewed as the two-letter simplification of the MJ matrix e_{ij}.[14,26]

RESULTS

First, we present results for the 2D 6×6 system. There are 28,728 maximally compact structures unrelated by symmetries of rotation, reflection, or reverse labeling. In the calculation with the MJ matrix, we used up to 9,095,000 randomly selected sequences of 20 amino acids. We found that 96.74, 42.46, and 17.79% of sequences had a unique ground state when the gap cutoff g_c was set to 0, 0.4, and 0.8, respectively. In Figure 1(a)–(c), we plot the histogram of the designability N_S, i.e., the number of structures with a given N_S versus N_S. As in the case of the HP model [shown in Fig. 1(d)], the distribution of N_S has a long tail, that is, there are some structures with much higher than average designability.[14,23,35] Further, for large gap cutoff g_c [Fig.1(c)] the curve resembles that of the HP model. One measure of the thermodynamic stability of a ground state is the energy gap Δ_{10} between the ground state and the next lowest energy state. To display the correlation between thermodynamic stability and designability, we average Δ_{10} over all sequences that design a structure, and then average over all structures with a given N_S. This doubly averaged energy gap is plotted against designability N_S in Figure 2. In both models (MJ and HP), there is a strong positive correlation between the average gap and designability N_S.

For the designability N_S to be a useful characterization of structures, it should be robust with respect to some variation in model parameters. We found a good correla-

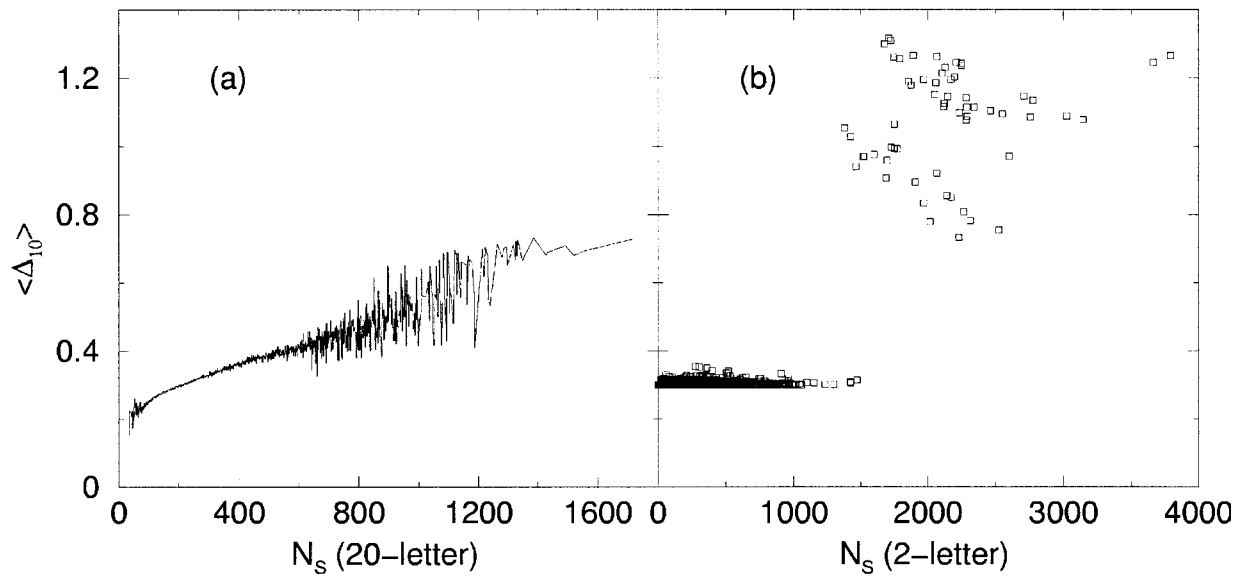

Fig. 8. The average gap $\langle\Delta_{10}\rangle$ vs designability N_S for the $3 \times 3 \times 3$ system. (**a**) For the MJ matrix. (**b**) For the HP model.

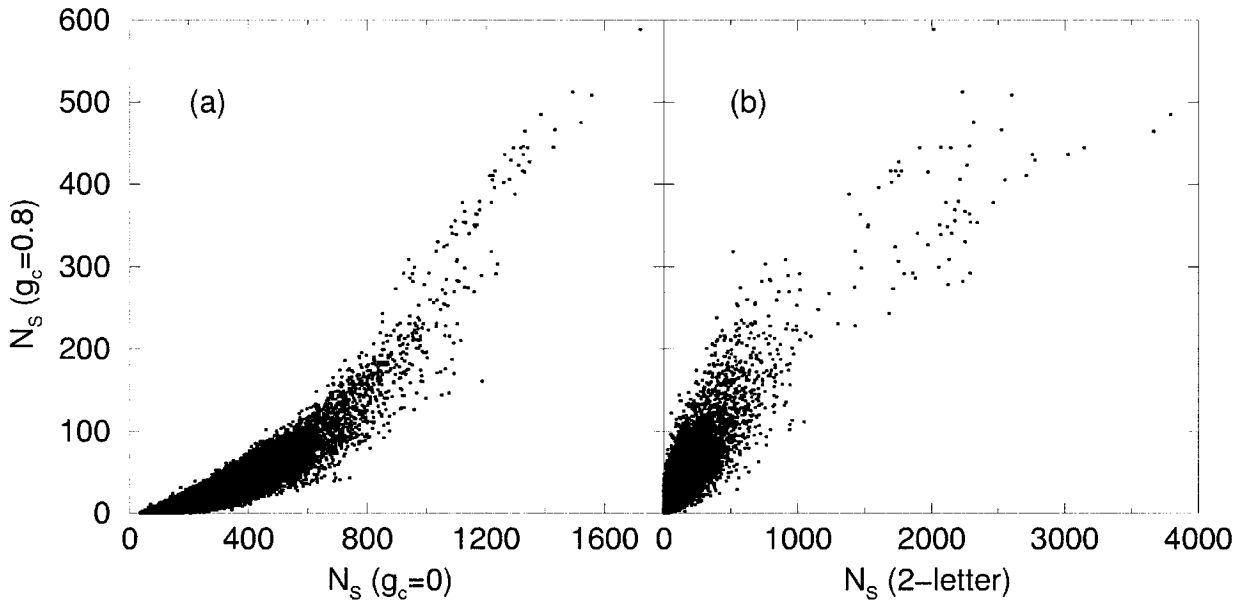

Fig. 9. (**a**) Designability N_S with gap cutoff $g_c = 0.8$ vs N_S with $g_c = 0$ for the MJ matrix for each $3 \times 3 \times 3$ structure. (**b**) N_S with $g_c = 0.8$ for the MJ matrix vs N_S for the HP model for each $3 \times 3 \times 3$ structure.

tion between the N_Ss of a given structure obtained with various gap cutoffs g_c [Fig. 3(a)] and obtained with the HP model [Fig. 3(b)]. In particular, highly designable structures in the HP model are also highly designable in the MJ matrix model and vice versa. The top structure is the same for both models (Fig. 4).

Do the sequences that design highly designable structures have unusual ground-state energies E_0, or ground-state depths Δ (Eq. 2), or spectral widths Γ (Eq. 3)? In Figure 5 we plot the average over sequences of E_0, Δ, Γ, and $Z = \Delta/\Gamma$ versus N_S. It is clear from the figure that there are no significant correlations between N_S and average E_0 or Δ. Thus, a highly designable structure does not have a lower E_0 or a larger Δ. On the other hand, N_S correlates inversely with the average width of the spectrum Γ and therefore correlates positively with the Z score. However, the scatter of data for structures of given N_S is so large that small Γ does not necessarily imply large N_S. A small spectral width Γ is a necessary but not a sufficient condition for a structure to be highly designable.

To see how the distribution of sequences among the structures changes with the length of the chain, we also studied a 6×5 system. In this case, there are 6802 maximally compact structures unrelated by symmetries. We used 5,200,000 randomly selected sequences. The percentage of sequences that had a unique ground state was 96.86, 43.96, and 19.11%, for $g_c = 0$, 0.4, and 0.8, respectively. These percentages are slightly larger than

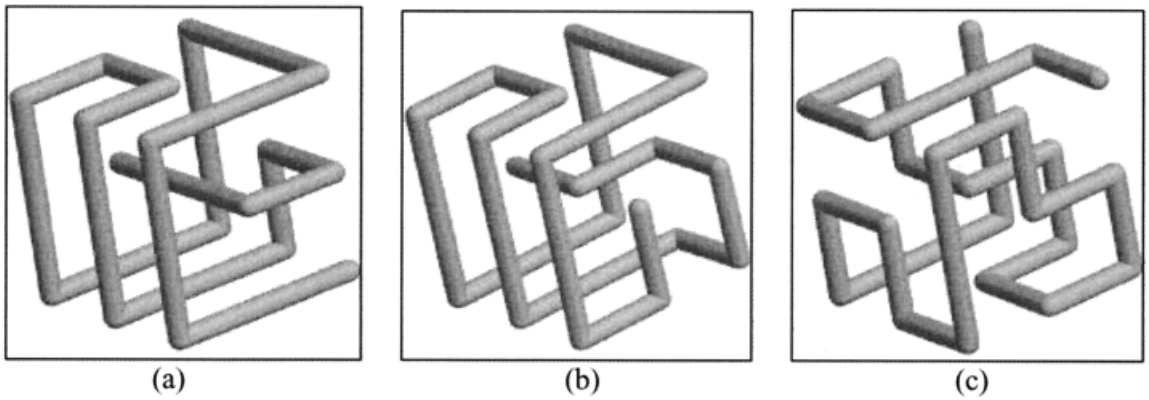

Fig. 10. (**a**) Top structure for the $3 \times 3 \times 3$ system with the MJ matrix. (**b**) Top structure for the HP model. (**c**) Structure with low N_S in both the MJ and HP models. Poorly designable structures typically show less geometric regularity than highly designable structures.

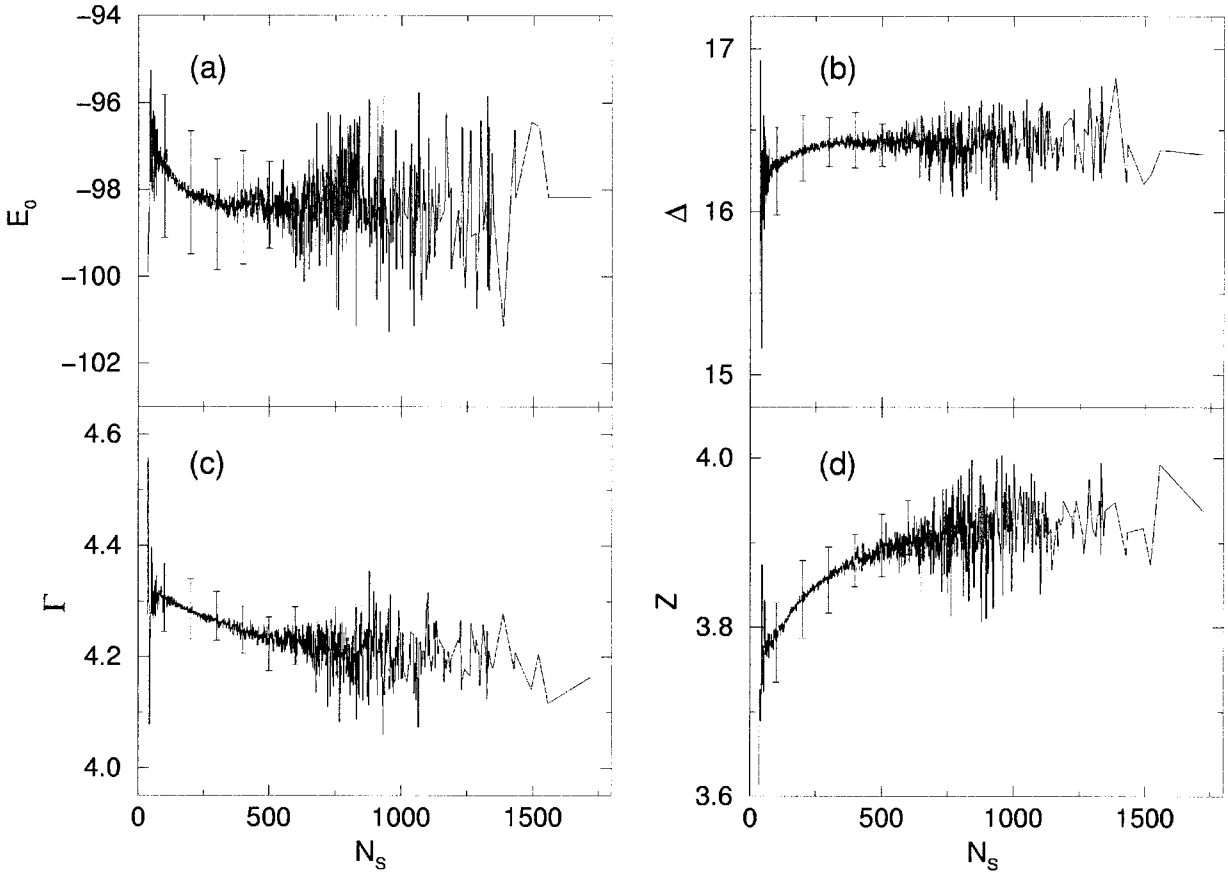

Fig. 11. (**a**) Ground-state energy E_0, (**b**) ground-state depth Δ, (**c**) width of compact spectrum Γ, and (**d**) $Z = \Delta/\Gamma$ vs N_S for the $3 \times 3 \times 3$ system. Solid lines are averages for given N_S. Error bars indicate variances.

those in the 6×6 system, indicating a slight decrease in the probability of a unique ground state with increasing chain length. In HP models, the histograms of designabilities for different system sizes were found to be identical after rescaling. To test for this property in the MJ model, we let $\mathcal{N}(N_S,L)$ be the number of structures with designability N_S in the system of chain length L. The dependence of $\mathcal{N}(N_S,L)$ on L may be "scaled out," and $\mathcal{N}(N_S,L)$ may be reduced to a "universal" form. We make the scaling *ansatz* (guess)

$$\mathcal{N}(N_S, L) = \frac{N_c}{\langle N_S \rangle} f\left(\frac{N_S}{\langle N_S \rangle}\right), \tag{4}$$

where N_c is the total number of structures and $\langle N_S \rangle$ the average designability for chain length L. If Eq. 4 holds, then the universal function $f(x)$ should be independent of L. In Figure 6, we plot $f = \mathcal{N}\langle N_S \rangle / N_c$ versus $x = N_S/\langle N_S \rangle$ for systems of 6×6 and 6×5. The two curves match well, supporting the scaling *ansatz* (4).

We now turn our attention to the 3D $3 \times 3 \times 3$ system. There are 51,704 compact structures unrelated by symmetries. A total of 13,550,000 randomly selected sequences of 20 amino acids were used in the calculation. With the gap cutoff $g_c = 0$, 0.4, and 0.8, the percentage of the sequences that had a unique ground state was, respectively, 96.67, 30.20, and 8.26%. In the HP model this percentage is 4.75%.[14] Histograms of the designability N_S, along with the histogram for the HP model, are plotted in Figure 7. Similar to the 2D case, there is a long tail to the distribution and the histogram for $g_c = 0.8$ resembles that of the HP model. In Figure 8, we show the average gap $\langle \Delta_{10} \rangle$ versus N_S. Again, the sequences that design structures with larger N_S have larger gaps, on average.

Note that there seems to be a qualitative difference in Figures 8(a) and (b): $\langle \Delta_{10} \rangle$ varies continuously with N_S for the 20-letter code but has a jump in the HP model. This jump in the HP model reflects two different kinds of rearrangements between the ground state and the next lowest energy structure. One kind of rearrangement changes the position of an H monomer from relatively buried to relatively exposed, so the number of H water bonds is increased. This kind of rearrangement has an energy > 1. The second kind of rearrangement breaks an H—H bond and a P—P bond to form two H—P bonds, which has an energy cost of only $2e_{HP} - e_{HH} - e_{PP} = 0.3$. The jump in Figure 8(b) indicates that rearrangements of the first kind are required for structures with large N_S but not for structures with small N_S. The difference in energy between the two kinds of rearrangements is smeared out when the 20-letter code is used.

The designability of structures is rather robust characterization—we observe good correlations between N_Ss obtained with different g_cs and between the MJ and HP models (Fig. 9). The most designable structure in the MJ model [shown in Fig. 10(a)] is not the same as in the HP model [shown in Fig. 10(c)], although they share some common geometric features, e.g., many antiparallel long lines. In Figure 11 we plot the quantities E_0, Δ, Γ, and $Z = \Delta/\Gamma$ for the 3D $3 \times 3 \times 3$ system versus N_S. Similar to the 2D case, there is little dependence of E_0 and Δ on N_S. On average, there is an inverse correlation between Γ and N_S and therefore a positive correlation between Z and N_S. The scatter of the data is large.

Finally, we consider the set of sequences that design the top $3 \times 3 \times 3$ structure in the MJ model. Of 13,550,000 randomly selected sequences, 1721 of them design the top structure, namely, they have the top structure as their unique ground state. In Figure 12, we plot the average hydrophobicity of the residue as a function of the chain index i, averaged over all the 1721 sequences that design the top structure. It is clear that there is a strong correlation between the average hydrophobicity of the residues and the exposure to water of the site—the more buried the site, the more hydrophobic the residue, on average. In Figure 13 we plot several quantities versus the ground-state energy E_0 for the sequences that design the top structure. We see that there is no correlation between the gap Δ_{10} and the ground-state E_0, whereas E_0 is inversely correlated with both Δ and Γ. However, no obvious correlation is seen between E_0 and $Z = \Delta/\Gamma$, as if the effect of a lower E_0 is just to uniformly pull down the energy spectrum, enlarging Δ and Γ by the same factor. Similar statistical behaviors are found for all sequences.

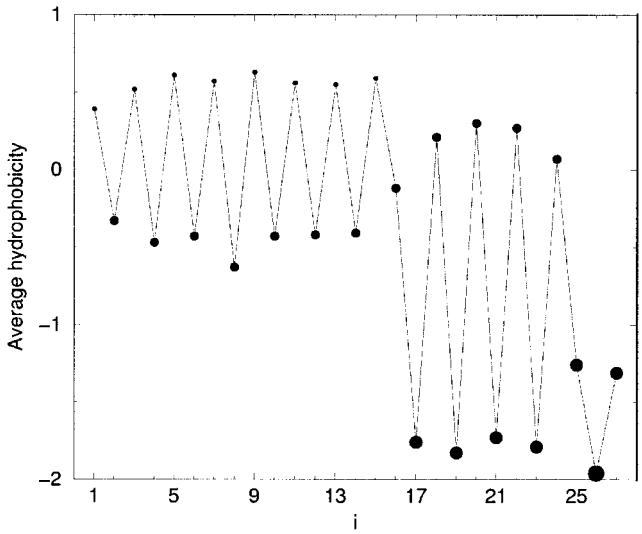

Fig. 12. Average hydrophobicity over the sequences that design the top $3 \times 3 \times 3$ structure in the MJ model. The size of each black dot, from small to large, represents the position of the residue: corner, edge, face, and center. The hydrophobicity scale of the 20 amino acids is taken from: Creighton TE. Proteins. Freeman: New York; 1993. Note that the hydrophobic scale ranges from negative (hydrophobic) to positive (hydrophilic).

DISCUSSION AND CONCLUSION

There has been much discussion of the minimum alphabet size for protein folding.[22,36–41] The answer undoubtedly depends on the questions addressed. The above results show no sensitive dependence of designability on the alphabet size when the potential includes the hydrophobic interaction. Recently, Buchler and Goldstein studied the designability for structures on a 5×5 lattice using various alphabet sizes for the sequence.[23,24] They obtained poor or no correlation between the designability N_S obtained with our HP parameters and with an MJ matrix. The reason for this discrepancy is that they used a different MJ matrix than the one we used to derive our HP parameters. Note that there are several matrices in Miyazawa and Jernigan's original articles.[25,27] The one we used for this study, and for deriving our HP parameters, is the matrix e_{ij}, which is the upper half of Table V in Ref. 25 or the upper half of Table 3 in Ref. 27. This is the matrix containing all interactions including the hydrophobic interaction. We analyzed this matrix via eigenvalue decomposition[26] and found that the matrix can be well approximated by the following form:

$$e_{ij} \approx \tilde{e}_{ij} = h_i + h_j + c(i,j)\,. \tag{5}$$

The additive term $h_i + h_j$ originates from the hydrophobic interaction and it dominates the potential (5).[26] The "two-body" term, $c(i,j)$, is small compared to the additive term and represents the tendency of similar amino acids to

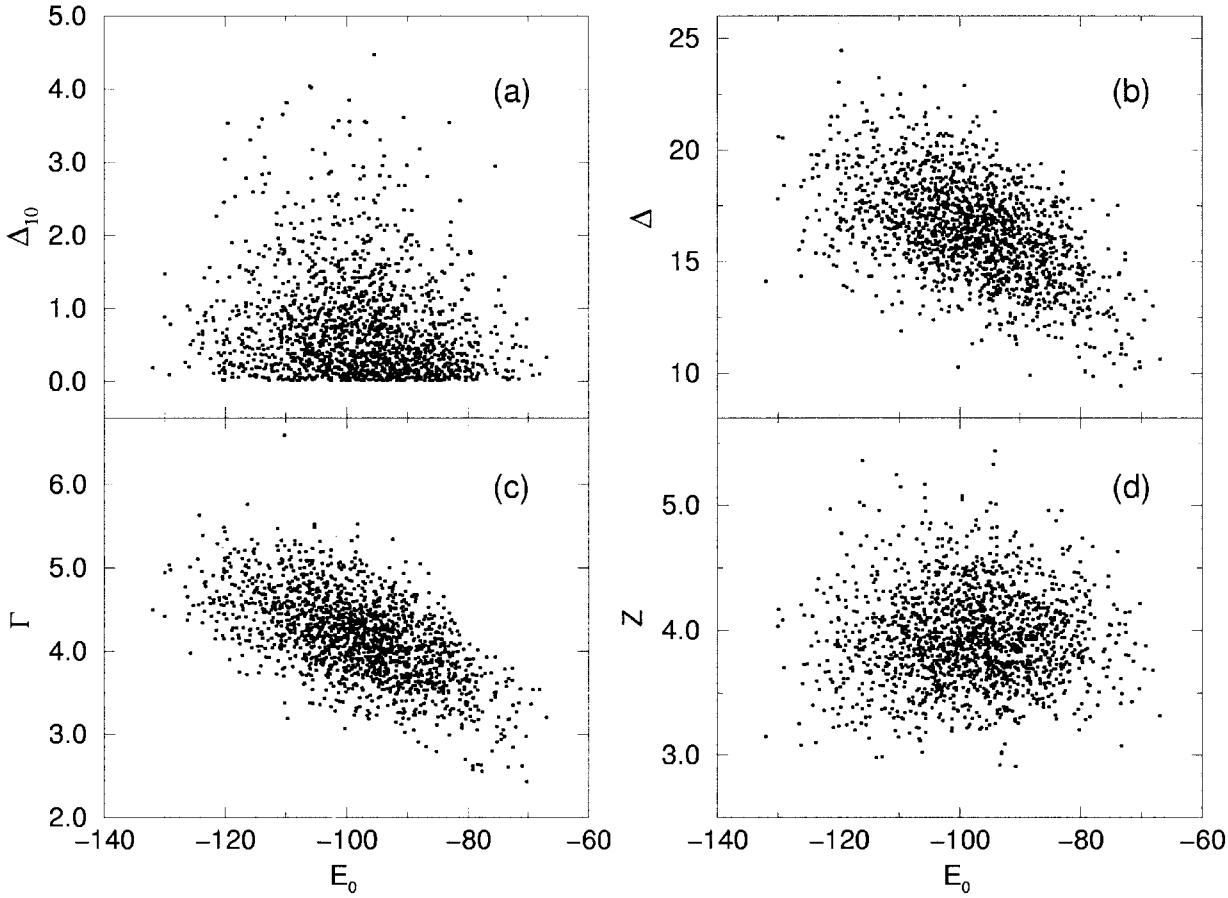

Fig. 13. (**a**) Energy gap Δ_{10}, (**b**) depth of ground state Δ, (**c**) width of compact spectrum Γ, and (**d**) $Z = \Delta/\Gamma$ vs the ground-state energy E_0 for the sequences that design the top structure in the MJ model for the $3 \times 3 \times 3$ system.

segregate.[26] Note that the hydrophobic interaction $h_i + hj$ not only imposes an overall drive toward compactness but, more importantly, it is also the main determinant of structural specificity.[15,26] The choice of $e_{\mathrm{HH}} = -2.3$, $e_{\mathrm{HP}} = -1$, and $e_{\mathrm{PP}} = 0$ in our HP study can be viewed as the result of a hydrophobic part $h_{\mathrm{H}} = -1$ plus a small two-body part $c(\mathrm{H,H}) = -0.3$, with $h_{\mathrm{P}} = 0$ and $c(\mathrm{H,P}) = c(\mathrm{P,P}) = 0$. The current study shows that there is no qualitative difference between our two-letter HP model and the full MJ matrix as far as designability is concerned. Thus, the designability of structures has no significant dependence on the alphabet size, as long as the potential is dominated by the hydrophobic or solvation force.[24] However, the outcome can be different for qualitatively different amino acid interaction potentials. For example, in the MJ matrix that Buchler and Goldstein used in their calculation (Table VI of Ref. 25) the hydrophobic force has been removed. It is a different potential, dominated by the pairing term $c(i,j)$ of Eq. 5. Its set of highly designable structures is different from that of the full MJ matrix and similar to that obtained for a *random* pairing potential.[24] Several authors have investigated the effect of the two-body pairing term in Eq. 5 on designability.[35,42–45] In particular, Shahrezaei and Ejtehadi showed in the case of a two-letter code that the set of highly designable structures is robust with respect to potential parameters and is largely determined by the structures' geometry.[25] It would be revealing to study how the designability of structures changes as the potential is changed from solvation-like to random-pairing-like.[46] It is not yet clear what role the alphabet size plays in the case of a random-pairing potential.[24]

REFERENCES

1. Anfinsen C. Principles that govern the folding of protein chains. Science 1973;181:223–230.
2. Richardson JS. Handedness of crossover connections in β-sheets. Proc Natl Acad Sci USA 1976;73:2619–2623.
3. Richardson JS. The anatomy and taxonomy of protein structures. Adv Protein Chem 1981;34:167–339.
4. Levitt M, Chothia C. Structural patterns in globular proteins. Nature 1976;261:552–558.
5. Chothia C. One thousand families for the molecular biologist. Nature 1992;357:543–544.
6. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of protein database for the investigation of sequences and structures. J Mol Biol 1995;247:536–540.
7. Orengo CA, Jones DT, Thornton JM. Protein superfamilies and domain superfolds. Nature 1994;372:631–634.
8. Finkelstein AV, Ptitsyn OB. Why do globular proteins fit the limited set of folding patterns? Prog Biophys Mol Biol 1987;50:171–190.
9. Finkelstein AV, Gutin AM, Badretdinov AY. Why are the same protein folds used to perform different functions? FEBS Lett 1993;325:23–28.
10. Finkelstein AV, Badretdinov AY, Gutin AM. Why do protein

architectures have Boltzmann-like statistics? Proteins 1995;23: 142–150.
11. Yue K, Dill KA. Forces of tertiary structural organization in globular proteins. Proc Natl Acad Sci USA 1995;92:146–150.
12. Govindarajan S, Goldstein RA. Searching for foldable protein structures using optimized energy functions. Biopolymers 1995;36: 43–51.
13. Govindarajan S, Goldstein RA. Why are some protein structures so common? Proc Natl Acad Sci USA 1996;93:3341–3345.
14. Li H, Helling R, Tang C, Wingreen N. Emergence of preferred structures in a simple model of protein folding. Science 1996;273: 666–669.
15. Li H, Tang C, Wingreen NS. Are protein folds atypical? Proc Natl Acad Sci USA 1998;95:4987–4990.
16. Mélin R, Li H, Wingreen NS, Tang C. Designability, thermodynamic stability, and dynamics in protein folding: a lattice model study. J Chem Phys 1999;110:1252–1262.
17. Tang C. Simple models of the protein folding problem. Physica A 2000;288:31–48.
18. Wang T, Miller J, Wingreen NS, Tang C, Dill KA. Symmetry and designability for lattice protein models. J Chem Phys 2000;113: 8329–8336.
19. Cejtin H, Edler J, Gottlieb A, Helling R, Li H, Philbin J, Wingreen N, Tang C. Fast tree search for enumeration of a lattice model of protein folding. J Chem Phys 2002;116:352–359.
20. Miller J, Zeng C, Wingreen NS, Tang C. Emergence of highly-designable protein-backbone conformations in an off-lattice model. Proteins 2002;47:506–512.
21. Borman S. Protein folding model focuses on "designability." Chem Eng News 1996;74:36.
22. Shakhnovich EI. Protein design: a perspective from simple tractable models. Fold Design 1998;3:R45–R58.
23. Buchler NEG, Goldstein RA. Effect of alphabet size and foldability requirements on protein structure designability. Proteins 1999;34: 113–124.
24. Buchler NEG, Goldstein RA. Surveying determinants of protein structure designability across different energy models and amino-acid alphabets: A consensus. J Chem Phys 2000;112:2533–2547.
25. Miyazawa S, Jernigan RL. Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. Macromolecules 1985;18:534–552.
26. Li H, Tang C, Wingreen NS. Nature of driving force for protein folding: a result from analyzing the statistical potential. Phys Rev Lett 1997;79:765–768.
27. Miyazawa S, Jernigan RL. Residue–residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. J Mol Biol 1996;256: 623–644.
28. Miyazawa S, Jernigan RL. Self-consistent estimation of inter-residue protein contact energies based on an equilibrium mixture approximation of residues. Proteins 1999;34:49–68.
29. Bowie JU, Lüthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. Science 1991;253:164–170.
30. Goldstein RA, Luthey-Schulten ZA, Wolynes PG. Optimal protein-folding codes from spin-glass theory. Proc Natl Acad Sci USA 1992;89:4918–4922.
31. Shakhnovich EI, Gutin AM. Enumeration of all compact conformations of copolymers with random sequence of links. J Chem Phys 1990;93:5967–5971.
32. Shakhnovich EI, Gutin AM. Implications of thermodynamics of protein folding for evolution of primary sequences. Nature 1990; 346:773–775.
33. Sali A, Shakhnovich E, Karplus M. How does a protein fold. Nature 1994;369:248–251.
34. Klimov DK, Thirumalai D. Factors governing the foldability of proteins. Proteins 1996;26:411–441.
35. Shahrezaei V, Ejtehadi MR. Geometry selects highly designable structures. J Chem Phys 2000;113:6437–6442.
36. Lau KF, Dill KA. A lattice statistical mechanics model of the conformational and sequence spaces of proteins. Macromolecules 1989;22:3986–3997.
37. Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH. Protein design by binary patterning of polar and nonpolar amino acids. Science 1993;262:1680–1685.
38. Riddle DS, Santiago JV, Bray-Hall ST, Doshi N, Grantcharova VP, Yi Q, Baker D. Functional rapidly folding proteins from simplified amino acid sequences. Nature Struct Biol 1997;4:805–809.
39. Wolynes PG. As simple as can be? Nature Struct Biol 1997;4:871–874.
40. Wang J, Wang W. A computational approach to simplifying the protein folding alphabet. Nature Struct Biol 1999;6:1033–1038.
41. Chan HS. Folding alphabets. Nature Struct Biol 1999;6:994–996.
42. Skorobogatiy M, Guo H, Zuckermann MJ. A deterministic approach to protein design problem. Macromolecules 1997;30:3403–3410.
43. Ejtehadi MR, Hamedani N, Seyed-Allaei H, Shahrezaei V, Yahyanejad M. Stability of preferable structures for a hydrophobic-polar model of protein folding. Phys Rev E 1998;57:3298—3301.
44. Ejtehadi MR, Hamedani N, Seyed-Allaei H, Shahrezaei V, Yahyanejad M. Highly designable protein structures and inter-monomer interactions. J Phys A 1998;31:6141–6155.
45. Ejtehadi MR, Hamedani N, Shahrezaei V. Geometrically reduced number of protein ground state candidates. Phys Rev Lett 1999;82: 4723–4726.
46. Kussell EL, Shakhnovich EI. Analytic approach to the protein design problem. Phys Rev Lett 1999;83:4437–4440.