Technical University of Denmark

DTU

Mapping of mutation-sensitive sites in proteinlike chains

Skorobogatiy, M.; Tiana, Guido

Link back to DTU Orbit

DTU Library
Technical Information Center of Denmark

Mapping of mutation-sensitive sites in proteinlike chains

M. Skorobogatiy[1] and G. Tiana[2]

[1]*Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139*
[2]*Department of Physics, Technical University of Denmark, Building 307, 2100 Lyngby, Denmark*

In this work we have studied, with the help of a simple on-lattice model, the distribution pattern of sites sensitive to point mutations (''hot'' sites) in proteinlike chains. It has been found that this pattern depends on the regularity of the matrix that rules the interaction between different kinds of residues. If the interaction matrix is dominated by the hydrophobic effect (a Miyazawa-Jernigan–like matrix), this distribution is very simple: All the hot sites can be found at the positions with the maximum number of closest nearest neighbors (bulk). If random or nonlinear corrections are added to such an interaction matrix the distribution pattern changes. The rising of collective effects allows the hot sites to be found in places with a smaller number of nearest neighbors (surface) while the general trend of the hot sites to fall into a bulk part of a conformation still holds. [S1063-651X(98)13008-8]

PACS number(s): 87.15.Da, 61.43.−j, 64.60.Cn, 64.60.Kw

I. INTRODUCTION

In this paper we study how the choice of a particular Hamiltonian is responsible for the distribution pattern of sites sensitive to point mutations in a heteropolymeric chain. As shown in [1] using a very simple model [2], in every optimized [3] sequence there are sites at which point mutations are likely to cause misfolding of the native state (we call them ''hot'' sites), while there are other sites at which point mutations have no relevant thermodynamic effect (''cold'' sites), and we call intermediate sites ''warm.'' As known from both experimental [4] and theoretical studies [1] usually proteins are made of few hot sites, while the majority of the other sites are cold. Because of the thermodynamic importance of the hot sites it is of general interest to investigate the principles guiding their positioning along the protein chain.

The model used in this work describes the polymer as a chain of beads in a cubic lattice, interacting through the Hamiltonian

$$H = \frac{1}{2} \sum_{i,j}^{L} B_{ij} \Delta(r_i - r_j),\tag{1}$$

where B_{ij} is the contact energy between the two residues situated on the ith and jth positions, L is the length of the chain, and $\Delta(r_i - r_j)$ has the value 1 if the ith and jth monomers are nearest neighbors and zero if not.

In the literature, different choices of the matrix B_{ij} have been used. The results of Refs. [1,2] have been found with a matrix [5] whose elements are distributed according to a Gaussian, with the average $\bar{B} = 0$ and standard deviation $\sigma = 0.3$ (in units of $kT = 0.6$ kcal/mol). With these matrix elements, for every target structure it is possible to select sequences whose native state is unique, stable, and kinetically accessible [6]. It has been found that a reliable condition for a site to be cold or hot is intimately connected to the change in the native state energy caused by a point mutation: The bigger this difference, the more probable for this site to be a

hot site. Hot sites of such sequences are mainly in the bulk sites of the native conformation, but can be on the surface as well, while some bulk sites can be rather insensitive to mutations [Fig. 1(a)]. Repeating the same calculations with a random-generated, Gaussian-distributed set of interaction energies, we have observed similar distribution patterns for hot and cold sites [Fig. 1(b)].

Another choice of the interaction matrix can be made to take into account explicitly the hydrophobic effect encountered in real proteins. The simplest way is to choose the matrix B_{ij} to be composed of only three different elements, namely, B_{HH}, B_{PP}, and $B_{HP} = B_{PH}$, where $B_{HH} < B_{PP}$
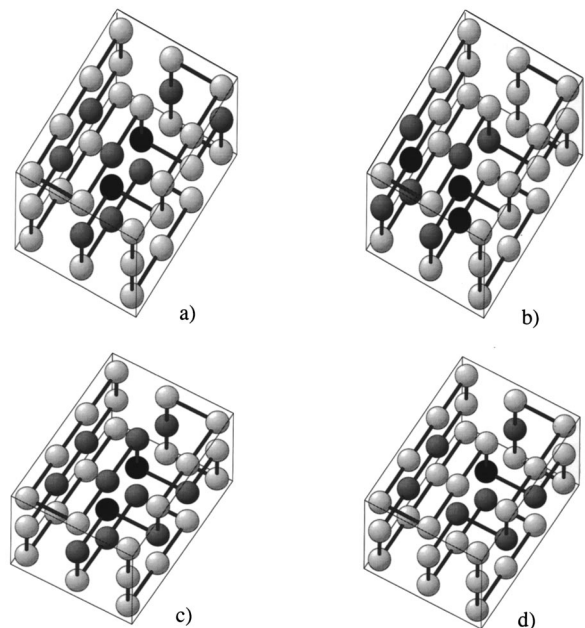

FIG. 1. (a) Map of the ''hot'' (black), ''warm'' (dashed), and ''cold'' (white) sites for the random Gaussian matrix. (b) Map of the mutation sites for the randomly generated Gaussian matrix. (c) Map of the mutation sites for the Li-Tang-Wingreen (LTW) parametrization of the MJ matrix with $\beta = 0$. (d) Map of the mutation sites for the MJ interaction matrix (HP-like model).

$<B_{HP}$. These elements are responsible for the interaction between hydrophobic (H) and polar (P) residues. Unfortunately, it was found [2] that for a given target structure and hydrophobic-polar (HP) interaction matrix it is difficult to construct optimal sequences for which this target structure would be kinetically accessible. As a rule, the optimization of the sequence puts H residues in the bulk sites (the sites with the greatest number of nearest neighbors) and every substitution of an H residue with a P residue causes the chain misfolding (see Figs. 2 and 3 in Ref. [7]).

On the way between the random matrix and the highly regular HP interaction model stands the matrix deduced by Miyazawa and Jernigan in [8]. This matrix contains 210 different elements that can still be grouped into three big blocks, according to their hydrophobicity. In this case it is possible to find sequences for which the native state is both stable and kinetically accessible. It was also found that, as in the situation with only two kinds of residues, hot sites are invariably the sites with the highest number of contacts [bulk sites, Fig. 1(d)].

The question again is what the principles guiding the positioning of hot sites along the protein chain are and how a particular form of the interaction matrix can influence the distribution pattern of hot and cold mutation sites. This paper is organized as follows. In Sec. II we will present a convenient representation of an interaction matrix as a function of the ''mixing'' parameter β. Variation of the mixing parameter will correspond to a change from a highly ordered HP-like interaction matrix at $\beta=0$ to a highly nonlinear matrix at nonzero values of the mixing parameter (Secs. III and IV). A distribution pattern of cold and hot sites will be investigated with the use of these interaction matrices. In Sec. IV we will address the same question by introducing a parameter of ''randomness'' γ, which will allow us to investigate the mutation sites distribution pattern with a highly ordered HP-like interaction matrix at $\gamma=0$ and a random Gaussian interaction matrix at large γ. Conclusions will be drawn in Sec. V.

II. LI-TANG-WINGREEN PARAMETRIZATION

In the work by Li, Tang, and Wingreen [9] a particularly interesting parametrization of a semiexperimental Miyazawa-Jernigan (MJ) matrix [8] was introduced as a consequence of its regularity. In their work it was shown that elements of the MJ interaction matrix can be very nicely fitted as

$$B_{\alpha\gamma}=q_\alpha+q_\gamma+\beta q_\alpha q_\gamma, \qquad (2)$$

where q_α is ascribed to a monomer of the type α and β is a constant. This parametrization involves 20 parameters, instead of 210 parameters of the MJ matrix, each specifying the strength of a particular residue. Using this parametrization we have found that, for the best fit, all q_α are negative and range from -2.3 to 0, while $\beta=-0.42$. Furthermore, it has been pointed out that the origin of the additive term is due to the hydrophobic effect, while the second-order term is responsible for the segregation of dissimilar residues.

Using this parametrization, we can write the Hamiltonian of the chain in the form [10]

$$H=\vec{q}\vec{n}+\frac{\beta}{2}\vec{q}C\vec{q}, \qquad (3)$$

where $\vec{n}$ and $\vec{q}$ have dimensions equal to the number of monomers in the chain. The ith coordinate of $\vec{n}$ is the number of nearest neighbors for the ith monomer and the ith coordinate of $\vec{q}$ specifies the strength of the residue in the ith site. C is the contact matrix for a given conformation. In Ref. [10] it has been shown that the Hamiltonian (3) fits very well the original MJ Hamiltonian and, more importantly, it is very convenient to handle analytically.

One of the problems that can be solved easily using the above Hamiltonian is, given a target structure (C matrix and $\vec{n}$) and the composition of the chain in term of residues, to find the sequence that minimizes the energy [10]. In particular, using that the second-order term is usually $2-3$ times smaller than the first-order term, a first-order approximation solution can be found by minimizing only the solvent exclusion term. A straightforward way is to choose a sequence so that the vectors $\vec{q}$ and $\vec{n}$ are as antiparallel as possible, keeping the constraint of a fixed number of different kinds of monomers. Knowing that all components of $\vec{q}$ are negative while all components of $\vec{n}$ are positive, it is necessary to put the residues with the most negative value of q_i in the sites of the target structure with the largest number of nearest neighbors. The effect is, roughly speaking, to put ''hydrophobic'' residues (i.e., low q_i) inside the structure while keeping ''hydrophilic'' ones (high q_i) on the surface. The second-order term in the Hamiltonian (3) is responsible for a fine-tuning of residue distribution, mostly inside the hydrophobic/hydrophilic regions, and causes a further decrease of the sequence energy.

III. MUTATIONS AND HOT SITES

According to [1], we label each mutation with the difference of the native state energy of the wild-type sequence and of the mutated sequence

$$\Delta E_{loc}=\frac{1}{2}\sum_{i,j}(B_{ij}^\alpha-B_{ij}^0)\Delta(r_i-r_j), \qquad (4)$$

where B_{ij}^0 is the interaction element associated with the wild-type sequence and B_{ij}^α with a mutated sequence. It has been shown that the information about the thermodynamic features of the mutated sequence is mostly contained in the value of ΔE_{loc} [6,11,13].

The energetic effects of mutations in a given site i are studied by introducing the average ΔE_{loc} over all possible mutations in this site. For the sake of mathematical simplicity the identical mutation is also included in the definition of ΔE_{loc} so that

$$\overline{\Delta E_{loc}}(i)=\frac{1}{20}\sum_\alpha^{20}\sum_j(B_{ij}^\alpha-B_{ij}^0)\Delta(r_i-r_j). \qquad (5)$$

For optimized sequences there are few sites (up to 10%) where $\overline{\Delta E_{loc}}$ is large ($\sim\Delta$, where Δ is the gap between the native state and the random conformations energy [2]), while

for the others $\overline{\Delta E_{loc}}$ is much lower. We call the former sites hot and the latter cold. If a mutation occurs in a hot site, there is a high probability that it fills the gap, eliminating the feature of design and causing misfolding of the chain.

For random interaction matrices whose elements have a Gaussian distribution, the behavior of hot sites is the same as indicated in Ref. [1]. Hot sites can be both in the bulk and on the surface of the native configuration [see Fig. 1(b)] with a dominance of hot sites in the bulk.

On the other hand, in a model with only two kinds of residues (H and P), the optimization of the sequence puts H residues in the sites with the highest number of nearest neighbors. From Ref. [7], every substitution of an H residue with a P residue causes the chain misfolding, thus signifying that sites containing the highest number of nearest neighbors tend to be the hot sites.

We are now interested in studying the pattern of hot sites in a model where the interaction matrix contains both features of hydrophobic-hydrophilic separation and randomness or complicated nonlinearity.

IV. CONSTRUCTION OF THE MAP OF HOT AND COLD SITES

To investigate the behavior of $\overline{\Delta E_{loc}}(i)$ we use the parametrized Hamiltonian (3). It is straightforward to show that, for a wild-type sequence characterized by $\vec{q}$,

$$\overline{\Delta E_{loc}}(i) = -\left(n_i + \beta \sum_j^L C_{ij} q_j \right)(q_i - \langle q \rangle_\alpha), \qquad (6)$$

where $\langle q \rangle_\alpha = \frac{1}{20} \Sigma_{\alpha=1}^{20} q_\alpha$ is the average of the values q_α corresponding to the 20 possible monomer substitutes for the wild type. With this assumption we have found $\langle q \rangle_\alpha \approx -1.26$.

We shall consider first the case $\beta = 0$, which is exactly solvable, and then the consequences of the nonlinear term introduced via the nonzero mixing parameter β.

A. $\beta = 0$

In the case $\beta = 0$ the Hamiltonian contains only the solvent exclusion term $H_0 = \Sigma_i q_i n_i$. As we discussed earlier, optimization of the sequence, given a structure, consists in putting the most hydrophobic residues into the sites with the greatest number of contacts, keeping with the constraint of a fixed number of different residues, so that

$$q_i \approx \frac{\sum_{i=1}^L q_i}{\sum_{i=1}^L n_i} n_i. \qquad (7)$$

For short chains and the 20 letter code used in the model with $\beta = 0$ the optimization procedure described above works reasonably well in constructing optimal sequences, while for longer chains it seems that the model is not adequate to ensure a single ground state or a ground state well separated from the other states. Nevertheless, the case $\beta = 0$ is a good starting point to investigate the role of the different terms of the Hamiltonian (3).
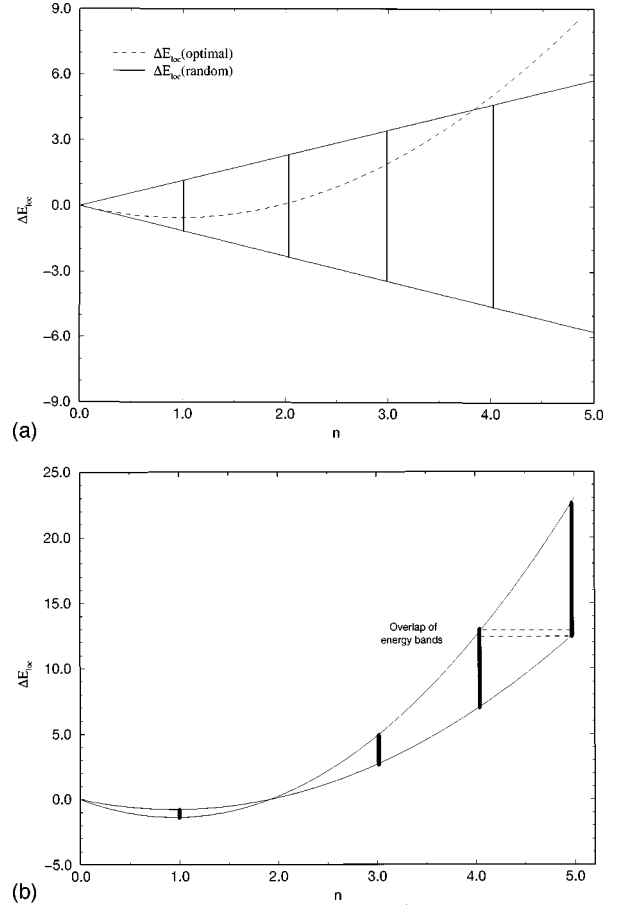

FIG. 2. (a) Energy of mutation as a function of the number of closest nearest neighbors for the LTW parametrized interaction matrix with $\beta = 0$. ΔE_{loc} for the optimized sequence exhibits nonlinear behavior leading to the sharp differentiation of hot and cold sites. ΔE_{loc} for the random sequence can have both positive and negative values at any n leading to the possibility of finding a hot mutation even at $n = 1$. (b) Energy of mutation as a function of the number of closest nearest neighbors for the LTW parametrized interaction matrix with $\beta < 0$. As second nearest neighbors start contributing to the mutation energy the energy line broadens, thus allowing for the sites with equal number of closest nearest neighbors to have different interaction energies. All possible ΔE_{loc} are confined between the two parabolas shown on a graph.

In this case Eq. (6) can be written as

$$\overline{\Delta E_{loc}}(i) = -n_i(q_i - \langle q \rangle_\alpha). \qquad (8)$$

Defining $\langle q \rangle = \Sigma_{i=1}^L q_i / \Sigma_{i=1}^L n_i$ and substituting Eq. (7) in Eq. (8) as an approximation for an optimal sequence, we obtain

$$\overline{\Delta E_{loc}}(i) = -n_i \langle q \rangle \left(n_i - \frac{\langle q \rangle_\alpha}{\langle q \rangle} \right). \qquad (9)$$

As q_i ranges in the interval $[q_{min}, 0]$ and assuming that $q_i \sim n_i$, it is easy to see that $\langle q \rangle \sim q_{min} / n_{max}$. For typical values of $\langle q \rangle \sim -0.6$ and $\langle q \rangle_\alpha \sim -1.26$,

$$\overline{\Delta E_{loc}}(i) \approx 0.6 n_i(n_i - 2.1). \qquad (10)$$

The shape of this function is plotted in Fig. 2. It is interesting to compare the value $\overline{\Delta E_{loc}}(i)$ for the optimized sequence

with the one for a random sequence. We will identify, in the spirit of the random energy model [12], hot sites (as defined in Sec. III) with those sites in which the average impact of mutations is greater than for a random sequence. We define a sequence as random if there is no correlation between the strength q_i at a given site and its number of nearest neighbors n_i, so that

$$\overline{\Delta E_{loc}^{rand}}(i) = -n_i(q - \langle q \rangle_\alpha), \qquad (11)$$

where q ranges between -2.3 and 0. The values that $\overline{\Delta E_{loc}^{rand}}(i)$ can assume are comprised between the two straight lines plotted in Fig. 2(a). While the dependence of $\overline{\Delta E_{loc}}(i)$ for the random sequence is linear, for a selected sequence it is quadratic. In the case of the selected sequence the quadratic behavior of the mutation energy versus the number of closest nearest neighbors induces a sharp distinction between bulk sites ($n \gtrsim 3$) and surface sites. As is clear from Fig. 2(a), all bulk sites are hot, while surface sites are cold. As consequence, hot sites have a certain degree of symmetry in target structures, i.e., none of the sites with the same number of nearest neighbors is privileged to the others. The map of the hot and cold sites for a Hamiltonian $\beta = 0$ and a 36 monomer target structure is presented in Fig. 1(c).

B. $\beta \neq 0$

If β is not zero but still small in absolute value, it is possible to minimize the energy of the sequence in two steps, first minimizing $H_{\beta=0} = \vec{n}\vec{q}$ and finding an initial trial optimal sequence $\vec{q}_{tr}$ and then reminimizing the same $H_{\beta=0}$ with an effective $\vec{n}_{eff} = \vec{n} + (\beta/2)\vec{q}_{tr}C$. As it was shown in Ref. [10], this procedure is quite reliable for small β.

Using the same approximation (7) as in Sec. IV A for $\vec{q}_{tr}$, one finds

$$\overline{\Delta E_{loc}^{\beta}}(i) \approx -n_i \langle q \rangle \left(1 + \beta \langle q \rangle \frac{N_i}{n_i}\right)\left(n_i - \frac{\langle q \rangle_\alpha}{\langle q \rangle}\right), \quad (12)$$

where $N_i = \Sigma_j C_{ij} n_j$ is the number of nearest neighbors of site i and can assume values in the range $\{n_i, \ldots, 4n_i + 1\}$. The expression of $\overline{\Delta E_{loc}^{\beta}}(i)$ then takes into account, through the value of N_i, sites lying further than the nearest neighbors of i. As consequence of this, we can observe [Fig. 2(b)] a broadening of the range of values that $\overline{\Delta E_{loc}^{\beta}}(i)$ can assume for each n_i (we will refer to this broadening as "energy bands" in our further discussion).

From Fig. 2(b) it follows that the effect of the segregation term in Eq. (3) is to differentiate among the sites with the same number of closest nearest neighbors. As $\beta \neq 0$, the second shell of nearest neighbors starts playing its role, thus introducing a cooperative effect in the determination of $\overline{\Delta E_{loc}^{\beta}}(i)$. This splitting of the degeneracies in the $\overline{\Delta E_{loc}^{\beta}}(i)$ at $\beta \neq 0$ can lead to the overlap of energy bands, thus leading to the possibility of encountering a hot site on the surface and cold site in the bulk.

To summarize, from the point of view of single mutations, in the case $\beta = 0$ the spectrum of mutations is composed of two main parts, namely, mutations in bulk sites, with high value of ΔE_{loc}, and mutations in surface sites,

with ΔE_{loc} close to zero or negative. The effect of the coupling term is to broaden the range of possible mutation energies for the mutation sites of the same type (bulk or surface) mixing the energetic levels corresponding to mutations in different kind of sites. So the symmetry of hot sites in the target structure can be broken, thus allowing some hot sites to be found on the surface.

C. Explicit calculations with the MJ parametrized Hamiltonian

To investigate how the energy bands depend upon the strength of the nonlinear contribution in the interaction matrix we have made some explicit calculations, using as the target structure the 36-mer chain displayed in Fig. 1. We have calculated the best values of q_α and β to fit the MJ matrix according to Eq. (2), finding $\beta = -0.42$. Using these values for q_α, we varied β in the range -1.5–1.5, optimizing each time the sequence with a genetic energy minimization technique. The composition has been kept fixed for all values of β and chosen in such a way as to satisfy $q_i \sim n_i$ (condition for optimal composition at $\beta = 0$). For the sake of computational convenience, the interaction matrix elements have been rescaled to have a zero average and standard deviation equal to 1.0. Then we plotted the value of $\overline{\Delta E_{loc}^{\beta}}(i)$ for each lattice site as a function of β. First, we consider the case $\beta < 0$. The different energy bands are shown in Fig. 3(a). For $-0.7 \lesssim \beta < 0$ all the bands lay in four distinct groups. The first two groups, which correspond to the sites with $n_i = 1$ or $n_i = 2$, contribute to cold sites. The other two groups correspond to the sites with $n_i = 3$ or $n_i = 4$ defining the warm and hot sites. This situation is very similar to the case of $\beta = 0$, where bands do not overlap and hot sites are exclusively in the bulk. It is not surprising to find the same distribution pattern for the MJ matrix ($\beta = -0.42$) [Fig. 1(d)] and the HP-like interaction matrix ($\beta = 0$) [Fig. 1(c)]. For $-2.5 \lesssim \beta \lesssim -1.0$ the nonlinear part of the interaction matrix starts playing its role. As it was shown in Sec. IV B, the second-neighbor shell contributes to the value of $\Delta E_{loc}^{\beta}(i)$. At these values of the parameter β we observe that some of the energy bands corresponding to warm ($n = 3$) and cold ($n = 1,2$) sites mix, while the bands corresponding to the hot sites stay well separated from the other bands.

In the case $\beta > 0$ [Fig. 3(b)] the nonlinear effect is much more dramatic. For $\beta \gtrsim 0.2$ all bands start mixing, allowing cold sites to penetrate the bulk while pushing hot sites on the surface. We can rationalize the different pattern of energy bands at $\beta < 0$ and $\beta > 0$ by examining Eq. (6) and the Hamiltonian (3). Noticing that $(\beta/2)\vec{q}C\vec{q}$ is of the order of $(\beta/2)\vec{q}\vec{n}\langle q \rangle$, we can rewrite the Hamiltonian (3) in the form

$$H = \vec{q}\vec{n}\left(1 + \frac{\beta}{2}\langle q \rangle\right) + \beta \eta, \qquad (13)$$

where η contains the collective effects and $\vec{q}\vec{n}[1 + (\beta/2)\langle q \rangle]$ describes a "renormalized" hydrophobic effect. In the case $\beta < 0$, $(\beta/2)\langle q \rangle > 0$ and the interaction matrix is largely dominated by the hydrophobic effect, while the
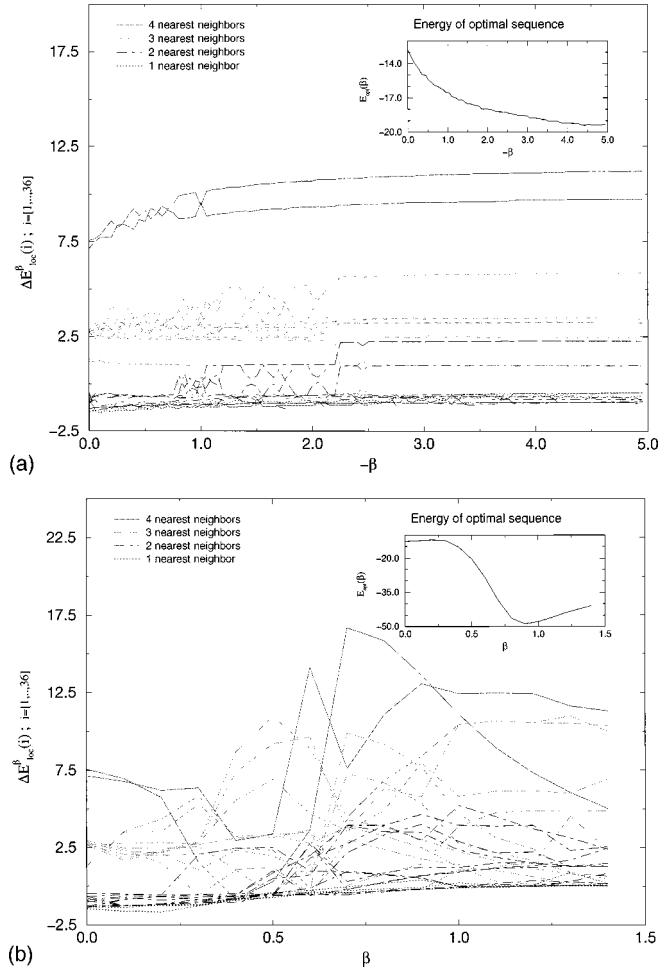
(a)

(b)

FIG. 3. Energy bands for 36 mutation sites. The interaction energy matrix is based on the LTW parametrization with the β parameter introducing a nonlinear segregation energy in the Hamiltonian. Different line types correspond to the sites with different numbers of closest nearest neighbors. For example, solid lines correspond to the mutation energy of sites 16 and 27 that have the most closest nearest neighbors. Sites with 4, 3, 2, and 1 closest nearest neighbors are specified by solid, dashed, solid-dashed, and dotted lines, respectively. (a) corresponds to $\beta < 0$ and (b) corresponds to $\beta > 0$. In the inset the energy of the optimal nonmutated sequence is shown for different values of β.

collective contribution to the Hamiltonian is not strong enough to substantially mix the energy bands.

For the $\beta > 0$ case $(\beta/2)\langle q \rangle < 0$ and the renormalized hydrophobic effect becomes comparable to or smaller than the collective term in the Hamiltonian. This allows all bands to mix substantially.

D. Explicit calculations for the random Hamiltonian

Another interesting question that can be addressed is how our conclusions are modified by the addition of a random term in the $\beta = 0$ Hamiltonian. To study this problem, we have chosen a Hamiltonian in the form

$$H = \vec{n}\vec{q} + \gamma \frac{1}{2} \sum_{i,j}^{L} \epsilon_{ij} \Delta(r_i - r_j), \qquad (14)$$

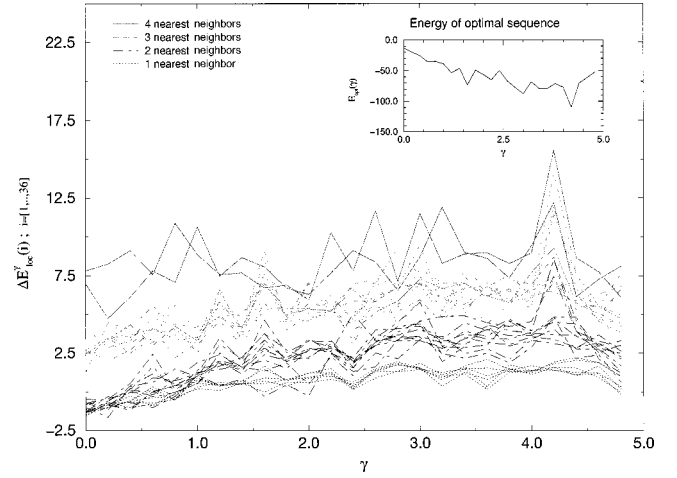where $\vec{n}\vec{q}$ comes from the parametrization (2) of the MJ ma-

FIG. 4. Energy bands for 36 mutation sites. The interaction energy matrix is a mix of the LTW parametrized matrix with $\beta = 0$ and the Gaussian random matrix. The mixing with the random matrix is controlled by the parameter γ. At $\gamma = 0$ the interaction matrix is pure LTW with $\beta = 0$, while at $\gamma \sim 2.0$ the elements of the random matrix become comparable to the elements of the regular matrix. Different line types correspond to the sites with different numbers of closest nearest neighbors. For example, solid lines correspond to the mutation energy of the sites 16 and 27 that have the most closest nearest neighbors. Sites with 4, 3, 2, and 1 closest nearest neighbors are specified by solid, dashed, solid-dashed, and dotted lines, respectively. In the inset the energy of the optimal nonmutated sequence is shown for different values of γ.

trix with $\beta = 0$. The virtue of this Hamiltonian is in separation in a controllable manner the hydrophobic effect due to the $\vec{n}\vec{q}$ and any other nonlinear effects are modeled by the random term. The values of ϵ_{ij} are taken from a Gaussian distribution with mean zero and standard deviation 1.0. In Fig. 4 the energy bands are shown as a function of the mixing parameter γ. The overall pattern is exactly the same as discussed in Sec. IV C for the case in which nonlinear terms in the parametrization are switched on leading to the effect of ''band crossing,'' thus allowing warm sites to appear on the surface and cold sites in the bulk. Even in the case of band crossing, there is still a clear trend for the hot and warm sites to concentrate in the bulk of the structure. This is due to the fact that bulk sites, building the biggest number of contacts, still display the strongest response to point mutations.

The insets of Figs. 3(a), 3(b), and 4 show the energy of the optimal nonmutated sequence for different values of β and γ, respectively. It is possible to observe a sudden decrease in energy of the optimal sequence as β and γ increases, signifying that the addition of nonlinear terms into a Hamiltonian allows for a much better energy minimization.

V. CONCLUSIONS

In this work we have considered how the regularity of the interaction matrix influences the distribution pattern of hot sites. It has been found that if the matrix is polarized (dominant hydrophobic effect, $\beta \sim 0$ or $\gamma \sim 0$) this distribution is very simple. All the hot sites can be found at the positions with the maximum number of closest nearest neighbors (bulk).

With increasing importance of nonlinear terms ($\beta \neq 0$, $\gamma \neq 0$) the distribution pattern changes so that the hot and warm sites can be found in places with a smaller number of nearest neighbors (surface), while the general trend of the hot sites to fall into the bulk part of a conformation still holds.

As pointed out above, this can be rationalized by noticing that if the mixing parameter is different from zero each site starts ''feeling'' not only its nearest neighbors but also the more distant sites. This leads to a collective nature of the interactions giving rise to a modified distribution pattern of hot sites.

In making a connection to the case of real proteins, however, one needs to proceed with caution. In the model considered in this paper we dealt with a $3 \times 3 \times 4$ system where a majority of the monomers could be found on a surface. The surface and bulk in our model were distinguished merely by the difference in the number of closest neighbors due to the local nature of the interaction matrix. In real proteins, however, other factors such as a long-range interaction between charged aminoacids can play an important role and special care might be needed in categorizing the location of a specific monomer.

[1] G. Tiana *et al.*, J. Chem. Phys. **108**, 757 (1998).

[2] E. Shakhnovich, Phys. Rev. Lett. **72**, 3907 (1994).

[3] E. Shakhnovich and A. Gutin, Protein Eng. **6**, 793 (1993).

[4] T. E. Creighton, *Proteins* (Freeman, New York, 1993).

[5] S. Miyazawa and R. L. Jernigan, Macromolecules **18**, 534 (1985), Table V.

[6] R. A. Broglia *et al.* (unpublished).

[7] H. Li, R. Helling, C. Tang, and N. Wingreen, Science **273**, 666 (1996).

[8] Reference [5], Table III.

[9] H. Li, C. Tang, and N. Wingreen, Phys. Rev. Lett. **79**, 765 (1997).

[10] M. Skorobogatiy, H. Guo, and M. Zuckermann, Macromolecules **30**, 3403 (1997).

[11] H. J. Bussemaker, D. Thirumalai, and J. K. Bhattacharjee, Phys. Rev. Lett. **79**, 3530 (1997).

[12] B. Derrida, Phys. Rev. B **24**, 2613 (1981).

[13] V. Abkevich, A. Gutin, and E. Shakhnovich, J. Chem. Phys. **101**, 6052 (1994).