Cooperative Structural Transitions Induced by Non-homogeneous Intramolecular Interactions in Compact Globular Proteins

I. Bahar* and R. L. Jernigan‡

*Chemical Engineering Department, Polymer Research Center, Bogazici University, and TUBITAK Advanced Polymeric Materials Research Center, Bebek 80815, Istanbul, Turkey; and ‡Laboratory of Mathematical Biology, Division of Cancer Biology and Diagnosis and Centers, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892 USA

ABSTRACT The role played by non-homogeneous interactions in stabilizing cooperative structural changes in proteins was investigated by exhaustive simulations of all compact conformations compatible with several well-defined globule-like shapes in three dimensions. Conformational free energies corresponding to the association of residues *i* and *j* were computed both for the unperturbed system, all subject to identical intramolecular interactions, and for the perturbed system in which a single pair of residues is probed by changing its interactions with an attractive or repulsive interaction. The high packing density leads to strong coupling between residues so that specific interactions between a given pair of residues are accompanied by considerable enthalpy changes. Relatively weak, about 1–2 kcal/mol, attractive interactions can exert a dramatic effect on the free energy distribution. Usually, central positions in the sequence most affect the conformational characteristics. Some of these interaction pairs appear to be capable of effecting major conformation transitions because of the high level of cooperativity in the dense state. Effects of repulsive interactions, however, do not depend so strongly on residue pair and cause more localized structural changes. This approach can suggest more, or less, sensitive loci for amino acid substitution.

INTRODUCTION

Simulations of model proteins were performed as described in the accompanying paper (Bahar and Jernigan, 1994) to investigate the role of homogeneous intramolecular attractions in stabilizing intermediate density equilibrium states of globular proteins (Kuwajima, 1989), which have been experimentally observed under mild denaturating conditions. The analysis suggests the possibility for the formation of those compact but denatured states, provided that the intramolecular potential energy favoring compact forms competes efficiently with the entropy opposing it. The choice between the two density regimes, corresponding to 1) the most compact, and 2) the compact but less tightly packed equilibrium structures, respectively, rests on a delicate balance between the entropic and enthalpic effects. The packing density in the most compact state is comparable with that of the native state of globular proteins and is naturally favored in the presence of stronger attractive interactions between residues. On the other hand, the compact but less tightly packed states are entropically driven due to their larger conformational freedom. Those intermediate equilibrium states in which native-like regular structural domains persist might be identified with the so-called molten globule state. Their stability relative to the completely unfolded state does necessitate attractive interactions between residues, but not strong enough to completely overcome the adverse entropic effect.

Received for publication 15 July 1993 and in final form 17 September 1993.

Address reprint requests to Dr. Robert L. Jernigan, Laboratory of Mathematical Biology, DCBDC, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892. Tel.: 301-496-4783; Fax: 301-402-4724; E-mail: jernigan@lmmb.nci.nih.gov.

0006-3495/94/02/467/15 $2.00

In the present work, we studied the most compact states, among these two high density regimes. This regime is represented by the subset of on-lattice, self-avoiding conformations, subject to the maximum number of intramolecular topological contacts allowable by the particular lattice geometry and confined to well-defined globular shapes in three dimensions. The space of accessible conformations in this subset, as well as the number of topological contacts between residues, increases with chain length, i.e., with the number of residues $N + 1$ comprising the chain, due to the absence of heterogeneous and/or specific interactions in the model. Thus, a unique most favorable compact state, equivalent to the native state, is not distinguishable in this regime, unless specific interactions discriminating between different residue-residue interactions are introduced.

Correlations between topological contacts and their perturbation by specific interactions were explored in this ensemble of highly compact structures. More specifically, the response of the overall chain conformation to attractive or repulsive interactions between different pairs of residues was studied. The objective was to understand the relative influence of different residue pair coupling or aversion on the secondary and possibly tertiary organization of compact chains. Experimental data indicate that the replacement of a single amino acid sometimes has no effect and other times significantly alters the structural characteristics of proteins. Both solvent exposure and internal organization of the amino acids exhibit considerable changes for different mutant forms. The purpose here was to search for specific locations along the primary sequence of polypeptides, which are potential candidates for developing cooperative structural changes. The question was to find which residues, if any, and which type of interactions are most effective in generating stabilized correlated structures. In general, residues near the chain termini would be expected

to exert a weaker effect on the equilibrium distribution of conformations, while interactions involving inner residues might be more effective in perturbing conformational energies. This feature follows in a simple manner from chain connectivity. Yet, the effectiveness of a given residue pair interaction to induce cooperative structural changes might be correlated with the particular three-dimensional shape of the globular protein. The particular shape might imply certain loci along the chain primary sequence which are most effective in prescribing tertiary structures. These issues are clearly of fundamental importance in designing or modifying proteins with well-defined structural and functional characteristics. The so-called inverse protein folding problem, i.e., determination of sequences that will fold to specific compact structures, has been the focus of several recent studies (Pabo, 1983; Ponder and Richards, 1987; Yue and Dill, 1992).

Topological contacts of various orders are operative in the ensemble of conformations belonging to the high density regime under study. In conformity with previous work (Ishinabe and Chikahisa, 1986; Chan and Dill, 1989a, b; Chan and Dill, 1990a, b; Bahar and Jernigan, 1994), a topological contact refers to the occupancy of two adjacent sites in the lattice by a pair of non-bonded residues, assuming the residues along the chain to be connected to each other by virtual bonds. The contact order refers to the number of intervening bonds in the sequence between a pair of residues in contact. Contacts of lower order stabilize helical structures in general (Zimm and Bragg, 1959), whereas regular structures such as β-sheets involve association of more distant residues. The relative importance of topological contacts of various orders in prescribing the observed global structure of proteins was studied. A thermodynamic approach was undertaken for evaluating the effectiveness of a given interaction energy $\Delta\epsilon_{ij}$, attractive or repulsive, between non-bonded residues i and j in perturbing the conformational space. Accordingly, the fluctuations in the free energy of all other topological contacts, which are induced by the presumed interaction $\Delta\epsilon_{ij}$, were quantitatively analyzed for the ensemble of all accessible conformations.

MODEL AND METHODS

Lattice simulations of model proteins

As in the accompanying work (Bahar and Jernigan, 1994), complete enumeration of short polypeptide chains within a defined dense space was adopted as the computational method. The conformational space is limited by 1) the three-dimensional shape of the molecule and 2) the requirement of maximum number of topological contacts, which reduce to a large extent the accessible number Ω of conformations, as explained below. Three-dimensional (simple cubic) lattice simulations were performed and compared with the two-dimensional (square) simulations of the accompanying paper (Bahar and Jernigan, 1994) to elucidate any effects arising from the dimensionality of the lattice.

On-lattice simulations of self-avoiding walks (SAWs) have proven to be quite informative in previous studies of highly compact polymeric structures (Wall and Whittington, 1969; Barber and Ninham, 1970; McKenzie, 1973; Schmalz et al., 1984; Ishinabe and Chikahisa, 1986). Entire allocation of space without violation of the volume exclusion requirement and complete sampling of the conformational subspace of compact states are the two main advantages of this approach. The importance of enumerating such compact walks for the theory of glass transition of polymer melts as well as polymer collapse in solution, has been discussed previously (Flory, 1982; Schmalz et al., 1984).

Inasmuch as a compact, globular shape, devoid of any free volume to allow for the accommodation of solvent molecules in the interior, is considered to be the general property of most soluble proteins, the extensive use of compact SAW chain model in previous studies of protein structure is justifiable. Two different approaches are used in lattice simulations of compact proteins: Monte Carlo generation and complete enumeration. The random search mechanism of the former implies a finite probability of never obtaining the native structure. Yet, this method implemented with a Metropolis algorithm has been advantageously used to elucidate the mechanism and kinetics of protein folding/unfolding processes (Taketomi et al., 1975; Go and Taketomi, 1978; Go and Taketomi, 1979a, b; Shakhnovich et al., 1991). The second method, employed here, is limited, on the other hand, to relatively short polypeptide chains but presents the advantage of permitting an exhaustive analysis of all accessible conformations. It has recently been shown by Chan and Dill (1989a, b; 1990a, b) to be a powerful tool for understanding several basic structural features of globular proteins, such as the important role of steric constraints and consequently packing efficiency requirements in the formation of secondary structures in compact globular proteins. We also mention another efficient use of lattice geometries, mainly simulating real globular proteins, by placement of residues at regular sites, based on coordinates from crystallographic data (Ueda et al., 1978; Covell and Jernigan, 1990; Skolnick and Kolinski, 1990; Kolinski et al., 1991).

The model has been presented in the second section of the accompanying paper, and the reader is referred to that section for an outline of the basic assumptions and definitions used throughout both papers. An additional element introduced in the present work is the use of the simple cubic lattice in the calculations instead of the square-planar lattice. The advantages and applicability of two-dimensional square lattices in providing a low resolution model of compact proteins has been discussed by Chan and Dill (1989a). The information extracted from those simulations was also confirmed with subsequent three-dimensional simulations (Chan and Dill, 1990a, b), which compare favorably with observations (Kabsch and Sander, 1983) of protein crystal structures. However, it is evident that a three-dimensional lattice would lead to a better physical representation of real chains. A major disadvantage of the two-dimensional square lattice is its low coordination number $z = 4$. In fact, detailed analysis of the x-ray crystal structure coordinates of several proteins indicates the applicability of a coordination number of $z = 6$–8 in general (Miyazawa and Jernigan, 1985). Also, confinement of a given structure to two dimensions naturally leads to an overestimation of the constraints arising from volume exclusion. The higher conformational degrees of freedom in cubic lattices are more realistic in this respect. Coordinates of individual residues in proteins were shown to fit to a good approximation with all three types of cubic lattice: simple, body-centered, and face-centered, the latter being slightly superior because it provides closely matching angles to those between virtual bonds (Covell and Jernigan, 1990). The errors incurred by using a lattice model to represent secondary structures have recently been discussed by Chan and Dill (1990b). Their analysis demonstrates that all the cubic lattice models represent helix topology to a good approximation.

General approach

The model chains are assumed to be subject to intramolecular attractive or repulsive interactions, in addition to the steric effect inherently precluding self-intersections. Residues along the chain are indexed from 1 to $N + 1$ following the primary sequence. The conformational potential $E\{\mathbf{r}\}$ of a given chain $\{\mathbf{r}\}$ is assumed to result from the additive contributions of binary interaction energies ϵ_{ij} between residues $[i, j]$ in topological contact, i.e., non-bonded residues occupying adjacent lattice sites, as presented in Eq. 3 of the accompanying paper. The ternary interaction parameter ϵ_{ijk} was set equal to 0. The total binary interaction potential ϵ_{mn} was conveniently expressed as $\epsilon + \Delta\epsilon_{mn}$, where ϵ is the homogeneous in-

teraction parameter and $\Delta\epsilon_{mn}$ is the perturbation term arising from specific interaction between the pair of residues $[m, n]$. This overall conformational potential $E\{\mathbf{r}\}$ is used to assign a Boltzmann statistical weight

$$W\{\mathbf{r}\} = \exp[-E\{\mathbf{r}\}/RT] \tag{1}$$

and a probability

$$P\{\mathbf{r}\} = W\{\mathbf{r}\} \Big/ \sum_{\{\mathbf{r}\}} W\{\mathbf{r}\} \tag{2}$$

to each conformation $\{\mathbf{r}\}$. The probability of occurrence of a chain subject to a given topological contact $[i, j]$ reads

$$P(i,j) = \sum_{\{\mathbf{r}|(i,j)\}} W\{\mathbf{r}|[i,j]\} \Big/ \sum_{\{\mathbf{r}\}} W\{\mathbf{r}\} \tag{3}$$

where $W\{\mathbf{r}|(i, j)\}$ denotes the statistical weight of a chain conformation $\{\mathbf{r}\}$ subject to the contact $[i, j]$.

In the case of systems subject to homogeneous interactions between residues such that $\epsilon_{ij} = \epsilon$ for all $[i, j]$, the probability of occurrence of a chain with a given topological contact becomes identical to

$$P_0(i,j) = \sum_{\{\mathbf{r}|(i,j)\}} \Omega\{\mathbf{r}|[i,j]\} \Big/ \sum_{\{\mathbf{r}\}} \Omega\{\mathbf{r}\} \tag{4}$$

Here $\Omega\{\mathbf{r}|[i, j]\}$ refers to the number of chains with conformation $\{\mathbf{r}\}$ exhibiting topological contact between residue pair $[i, j]$. The summation in the denominator is equal to the total number Ω of generated SAWs, and the subscript 0 appended to $P_0(i, j)$ indicates the particular case of homogeneous interactions, referred to as the unperturbed state. Depending on the particular type of intramolecular interactions, homogeneous or heterogeneous, two expressions

$$\Delta G_0(i,j) = -RT \ln P_0(i,j) \tag{5}$$

and

$$\Delta G(i,j) = -RT \ln P(i,j) \tag{6}$$

may be written for the free energy of contacts $[i, j]$ in compact structures. The former represents the entropic contribution, inasmuch as it accounts for the reduction in the number of accessible conformations arising from the particular choice of topological contact. The second incorporates both entropic and enthalpic effects. The difference between those two values yields the change in free energy of contacts or the enthalpic effect brought about by specific interactions between residues.

In general, specific interactions between residues m and n induce perturbations in the free energies associated with all other contacts $[i, j]$, due to the fact that residues are strongly-coupled to each other in the highly constrained system of compact structures. The fluctuations $[\Delta G(i, j) - \Delta G_0(i, j)]_{m,n}$ in the free energy of all contact pairs $[i, j]$, where $1 \le i \le N - 2$ and $i + 3 \le j \le N + 1$, may be adopted as a quantitative measure of the strength of specific interactions between residues m and n in perturbing the conformational space. The root-mean-square (RMS) fluctuation in contact free energies, $C(m, n)$, resulting from the specific interaction between residues $[m, n]$ is evaluated from

$$C(m,n) \equiv \left[C_0^{-1} \sum_{i,j} [\Delta G(i,j) - \Delta G_0(i,j)]^2_{m,n} \right]^{1/2} \tag{7}$$

where the summation is performed over all types of accessible contacts $[i, j]$, C_0 is the normalization factor which is equal to the number of all accessible contacts, and the subscript m, n indicates the pair of residues responsible for the perturbation of the system being subject to nonhomogeneous interaction.

RESULTS FROM CALCULATIONS AND DISCUSSION

General structural characteristics

The subset of conformations with the maximum number of topological contacts may result in various shapes, depending on chain length, as illustrated in Fig. 1 of the accompanying paper. The chain length is characterized by the number of residues $N + 1$. On a square lattice, when this number is equal to s^2, where s is an integer ≥ 1, only a single shape, the square, is attainable for compact chains with the maximum number of topological contacts. Also, as discussed by Chan and Dill (1989a), the rectangular shape with s and $s + 1$ sites on the edges is the only possible compact shape for the case $N + 1 = s(s + 1)$. These two categories of numbers of residues have been referred to as "magic numbers," as they lead to unique compact shapes. Likewise, numbers leading to cubes and near-perfect cubes (square-based parallelepipeds with the base edge differing from the height by one unit) are classified as magic numbers in simple cubic lattices (Chan and Dill, 1990a). In the present work, compact structures compatible with different shapes, near perfect cubes and/or asymmetric shapes, are also analyzed. The shape is prescribed by a simulation box divided into lattice cells, each accommodating a given residue or a closely associated structural unit. Confinement of chains to a finite space (or simulation box) reduces to a large extent the number of accessible conformations. As a result, exhaustive enumeration of all distinct conformations compatible with the presumed shape is possible for chain lengths which, otherwise, could not be explored within reasonable computation time. Here, the set of distinct conformations refers to SAWs which are not superimposable by rigid body rotation or mirror image reflection. The heads and tails of the generated chains are assumed to be distinguishable, inasmuch as specific interactions assigned at precise locations are of interest.

The smallest possible system leading to a compact cubic structure on simple cubic lattice is a chain of eight units, each unit occupying the corner of a single cell. Three distinct conformations shown in Fig. 1 are accessible to this short chain. These three conformations may be viewed as a set of building blocks that appear frequently in longer chains, inasmuch as they comply with the basic requirement of efficient packing in restricted geometry. The dashed lines indicate the types of topological contacts accessible in those conformations. Conformation (*a*) presents two contacts of order 3, one of them simultaneously participating in a contact of order 5 (residues 1, 4, and 6). Contacts of order 7 appear in conformations (*b*) and (*c*), both coupled with contacts of lower order. Those simple structures give us indications of correlations existing between different pairs of contacts. For example, half of the contacts between residues i and $i + 3$ are found to be accompanied by contacts between residues i and $i + k$, where $k = 5$ or 7, indicating the tendency for correlated contacts engendered whenever residues i and $i + 3$ come into close proximity. Conformations (*a*) and (*b*) form the repeat units of helical structures of type (i) and (ii) defined by Chan and Dill (1990a). Residues 3–8 in conformation (*c*) may be identified by units participating in an antiparallel β-sheet at a turn, which is another regular secondary structural element. Two other types of helices have been defined for chains simulated on a simple cubic lattices, both occupying more than one lattice cell. The first is the

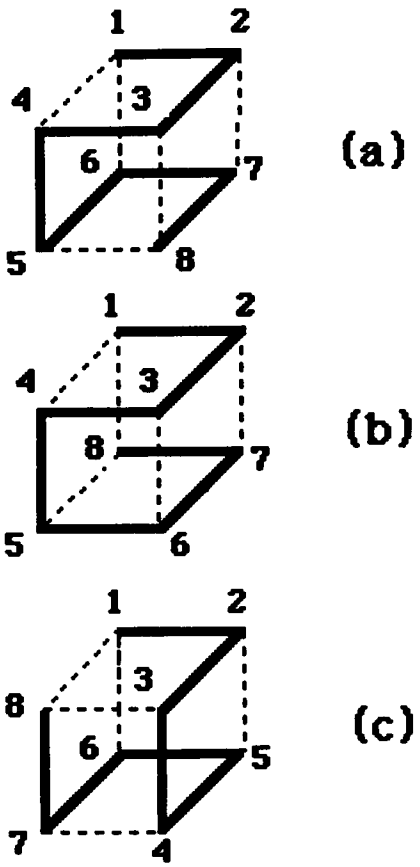

FIGURE 1 Schematic representation of all of the three distinct conformations (*a*) to (*c*) of compact chains of eight residues on a simple cubic lattice. Dashed lines indicate the pairs of residues with topological contacts.

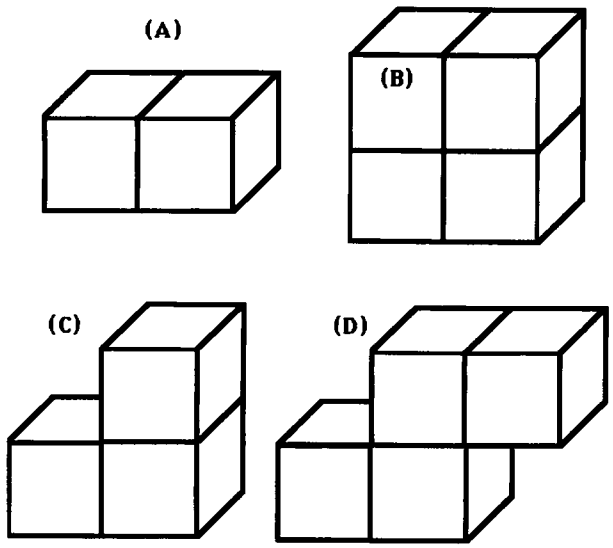

FIGURE 2 Schematic representation of the four shapes (*A*) to (*D*) of model proteins. Residues or closely associated units are assumed to occupy all corners and nodes of the cubic sites. Structural characteristics of the four cases are given in Table 1.

TABLE 1 Structural characteristics of generated model chains

Shape	$N + 1$	$N_{c,\max}$	Ω	$\overline{N_c^{(3)}}$	$\overline{N_c^{(4)}}$
(A)	12	9	73	4.90	0.55
(B)	18	16	2110	8.28	2.60
(C)	16	13	1632	5.00	2.26
(D)	20	17	8143	5.05	4.07

planar zigzag structure, referred to as helix type (iii), already encountered on the square planar lattice. Residue pairs $[i, i + 3]$ and $[i + 2, i + 5]$ are simultaneously associated in this type of helix. The second involves additional contacts between residues $i + 4$ and $i + 7$. It is not confined to a single plane but is propagated instead along the diagonal direction of lattice cells. From the high proportion of contacts of lowest order and their frequent coupling to higher order contacts, one might anticipate that their role could be of considerable importance in achieving particular structural organizations. On the other hand, the fact that those contacts are highly localized along the chain suggests that their effects on the tertiary structure might be limited. The efficacy of short-range contacts in promoting structural changes is explored in the present work.

Fig. 2 displays the four types of simple compact shapes (*A*) to (*D*) investigated here. Residues are placed at each node, so on form (*A*) there are exactly 12 residues, on (*B*) 18, on (*C*) 16, and on (*D*) 20. The former two belong to the set of magic numbers mentioned above. Most of the calculations were performed for shape (*B*), which represents the most compact structure of a chain composed of 18 units. The third and fourth shapes have been considered to search for effects arising from the asymmetry of the structure. Table 1 summarizes the structural characteristics of the four cases, such as the numbers $N + 1$ of residues or units in the generated chains, the total numbers $\Omega\{\mathbf{r}\}$ of accessible distinct conformations, and the numbers of topological contacts $N_{c,\max}$ per chain. Here the subscript *max* indicates that these are the maximum attainable numbers of contacts for the investigated chain lengths. The fifth and sixth columns indicate the average number of three-body and four-body contacts, $\overline{N_c^{(3)}}$ and $\overline{N_c^{(4)}}$, respectively, occurring in chains under those compact structures, evaluated from the total ensemble of generated SAWs. It is noted that the conformations with the shape (*B*) exhibit the largest number of multiple-body topological contacts, which indicates the enhanced tendency for intrachain interactions in globular structures with lower surface/volume ratio.

Comparison with two-dimensional model chains

The two pie charts in Fig. 3 display the distribution of contacts of various order $3 \leq k \leq N$ for a two-dimensional and a three-dimensional compact case of (*a*) 16-residue chains on a square lattice and (*b*) 18-residue chains on a simple cubic lattice. Those two chains are of comparable length and lead to unique shapes, square and near-perfect cube (*B*), belonging both to the set of magic numbers, in

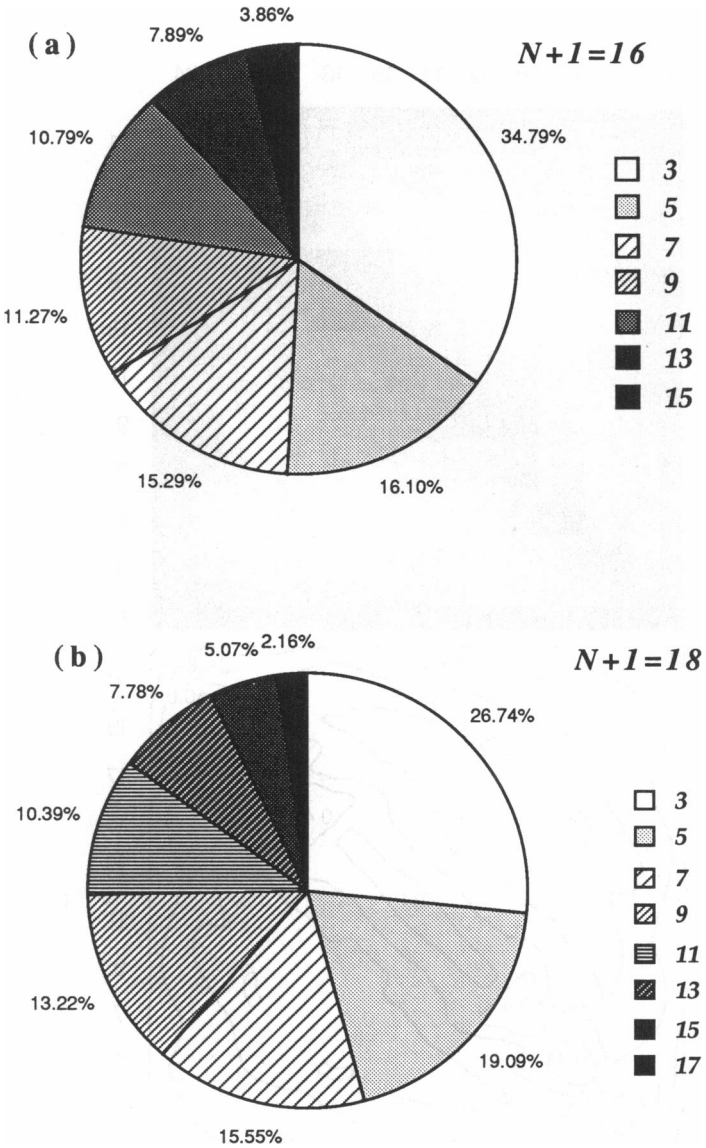

FIGURE 3 Distribution of topological contacts of various orders k, $3 \leq k \leq N$, in the most compact structures of (a) model chains of $N + 1 = 16$ residues on two-dimensional square lattice (4×4), and (b) model chain of 18 residues on simple cubic lattice of shape (B), indicating that two-dimensional simulations emphasize contacts of lower order.

corresponding lattice geometries. We note that the amount of lowest order ($k = 3$) contacts is considerably larger in part (a). Likewise, the proportion of contacts between most distant residues such as those separated by more than 14 virtual bonds is lower in two-dimensional model chains (<4%) compared with three-dimensional chains (>6%). Thus, confinement of a given model chain to two dimensions intrinsically enhances interactions between nearby residues, at the expense of those occurring between more distant residues. One might argue that the lower frequency of contacts between distant residues in part (a) arises from the smaller size of the chain, constraining the cyclization or folding of the chain so as to restrict contacts between more distant residues, and that the proportion of lowest order contacts would decrease with increasing chain length. Results in the accompanying paper demonstrate in fact that the contribution of lowest order contacts is depressed with increasing chain length. However, the pie chart displayed for the 25-residue chain in Fig. 13 of the accompanying paper indicates that contacts between nearest neighbors along the primary sequence of the chain are still overestimated compared with three-dimensional simulations, although a substantially larger variety of contact orders ($5 \leq k \leq 13$) competing with $k = 3$ is now accessible.

A more detailed examination of the probabilities of contacts of various orders may be achieved with the aid of contact maps. Accordingly, the subset of chains subject to a given topological contact $[i, j]$, $\Omega\{\mathbf{r} \mid [i, j]\}$, among the whole set $\Omega\{\mathbf{r}\}$ of compact structures are considered in order to deduce the free energy $\Delta G_0(i, j)$ associated with a given contact using Eqs. 4 and 5; these display the changes in free energy $\Delta G_0(i, j)$ for the most compact conformations of 25-residue chains on a square lattice and 18-residue chains on a simple cubic lattice [form (B)], respectively. The contact maps (*top*) and the corresponding isoenergetic contours (*bottom*) are shown in each figure. The lower triangular portions of the maps/contours refer to the indicated residues with indices $1 \leq i \leq N - 2$ and

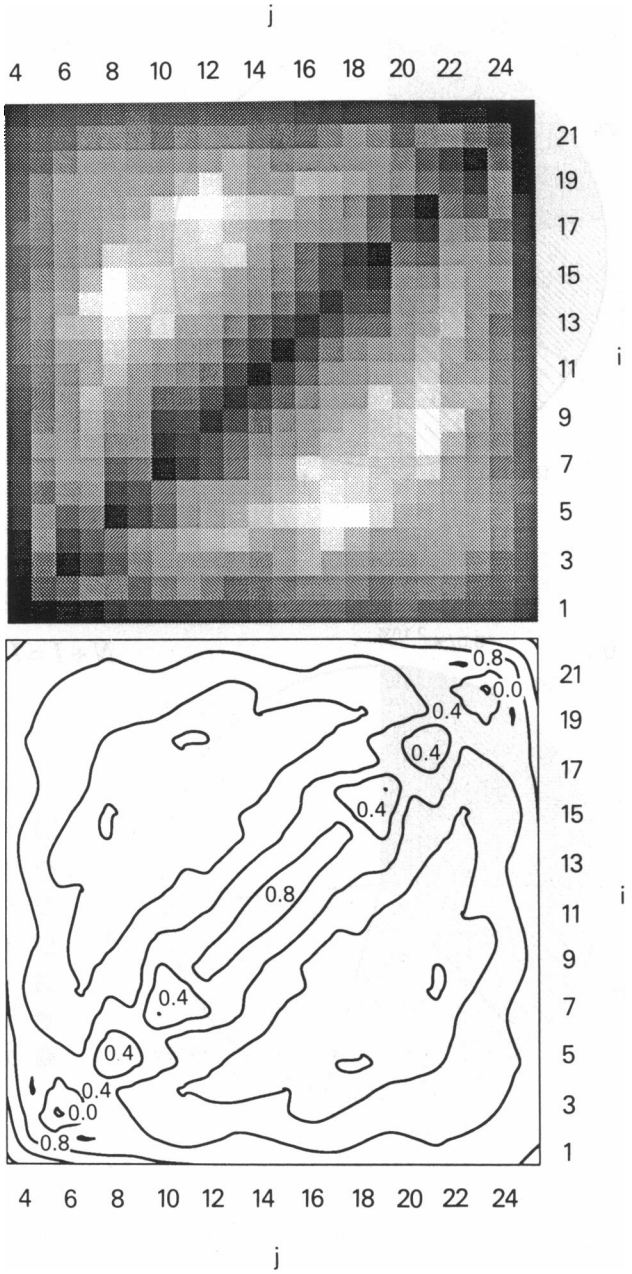

FIGURE 4 Map and contour plots of entropic free energies $\Delta G_0(i, j)$ associated with the pair of residues $[i, j]$ in compact structures of 25 residue chains simulated on square lattice. The lower triangular portions on the right in the map and contour plots are to be considered here and in the following figures. The upper triangular portion is the mirror image. Points along the diagonal with slope 1 represent contacts of lowest order between residues i and $j = i + 3$. Energy levels of the contours are expressed in dimensionless units $\Delta G_0(i, j)/RT$ with increments of 0.4. Highest energy contacts occur between the two residue pairs [5, 18] and [8, 21].

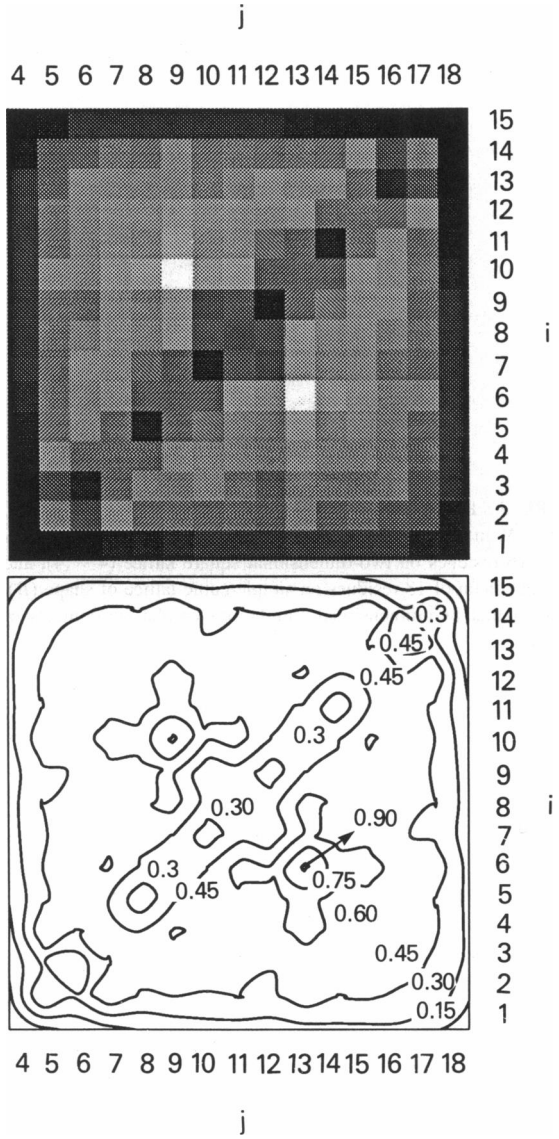

FIGURE 5 Map and contour plots of entropic free energies $\Delta G_0(i, j)$ associated with the pair of residues $[i, j]$ in compact structures of 18-residue chains of shape (B) simulated on a simple cubic lattice. Energy levels are expressed with increments of $\Delta G_0(i, j)/RT = 0.15$. The energy surface is substantially smoother than that obtained in two dimensions (Fig. 4) and also indicates an enhanced preference for contacts involving terminal residues.

$i + 3 \leq j \leq N + 1$, which constitute the set of interacting pairs. The upper triangular portions are the mirror images of the lower parts. Points along the diagonal (with slope 1) represent contacts of lowest order, i.e., those occurring between residues i and $i + 3$. Contacts between more distant residues are situated off the diagonal, their distance from the diagonal increasing with their separation in the primary sequence of the protein. Thus, in the present representation, helices involving residues i and $i + 3$ are represented by points along the diagonal, parallel β-sheets result in an array of points parallel to the main diagonal, and antiparallel β-sheets appear as an array of points with slope -1. The simple cubic lattice geometry permits topological contacts between residues separated by an odd number (≥ 3) of virtual bonds only. Points corresponding to residues separated by an even number of bonds have been assigned mean energy values on the basis of the neighboring points on the maps for visual clarity. Lower energy regions are darker in the contact maps. The energies of the contours

are expressed in terms of reduced units $\Delta G(i, j)/RT$, ranging from 0 to 2.0 with increment 0.4 in Fig. 4, and from 0 to 0.9 with increment 0.15 in Fig. 5.

From the comparison of the maps/contours in Figs. 4 and 5, the following features emerge. First, there is an enhanced tendency for contacts between more distant residues in three-dimensional model chains, compared with two dimensions, as indicated by the darker frame in Fig. 5. In particular, contacts involving terminal residues are now rendered almost as probable as lowest order contacts. We note that the apparent high fraction of lower order contacts in the pie charts in Fig. 3 is not related to the lower energy of those contacts but simply results from the larger number of candidate pairs ($N - 2$ of them) participating in contacts of order 3. Comparison of the free energies for individual pairs in the maps reveals that contacts between most distant residues are not the least probable, but on the contrary, almost as favorable as lowest order contacts, and preferred over most intermediate range contacts. A second important difference between the results obtained for the two cases is the relative heights of the free energy surfaces. In two dimensions, the least probable contacts, situated at positions [5, 18] and [8, 21], have free energy of 2.1 *RT* with respect to the zero level of contact [1, 4], whereas in three dimensions, the highest energy level that occurs at a single point [6, 13] is only 0.9 *RT* above the same minimum energy level. Thus, confinement of chains in two dimensions exaggerates the variations in the frequencies of different topological contacts in favor of closest neighbor interactions. In spite of those distinctions, we note that several qualitative features are preserved in simulations in both dimensions, such as the weaker probability of intermediate range contacts, the alternating frequency of lowest order contacts along the diagonals indicating an enhanced correlation between residues i and $i + 2$ where $i = 1, 3, 5$, etc., rather than between $i = 2, 4, 6$, etc., an effect discussed in the accompanying paper.

Perturbations induced by specific interactions

The contact free energies $\Delta G_0(i, j)$ may be alternatively shown in the form of a three-dimensional surface, which allows for a clearer perspective on regions of minimum or maximum energies. Fig. 6 shows the free energy surface corresponding to the 18-residue chain of shape (*B*). This is another rendition of the data of Fig. 5, which are useful for the comparison with the free energy surfaces corresponding to non-homogeneous interactions. The energy surface is rather flat, with borders curved down, as also implied by the darker edges of the contact map in Fig. 5. The pair of residues [6, 13] that exhibits the largest free energy $\Delta G_0(i, j)$ is the least probable contact from the entropic point of view. In this respect, one might expect to observe a large perturbation in the equilibrium distribution of topological contacts upon introduction of specific attractive interactions between residues 6 and 13 in this chain.

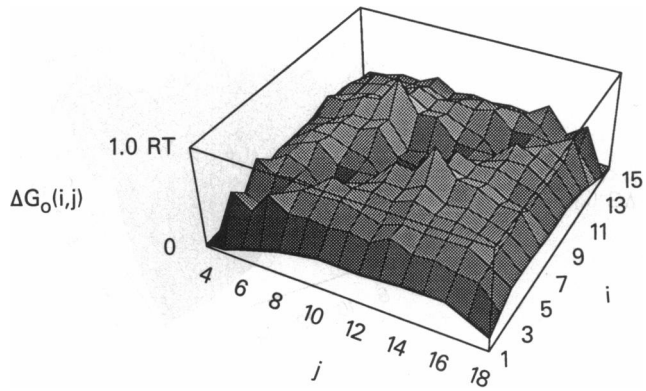

FIGURE 6 Entropic free energy surface $\Delta G_0(i, j)$ for the model chains of shape (*B*) in the unperturbed state, i.e., in the presence of homogeneous interactions between residues. This is another perspective of the map and contour plots presented in Fig. 5, which can be used in comparison with Figures 7 (*a*) and 8 (*a*) to visualize the relative contributions of the entropic and enthalpic factors to the free energy of the perturbed system.

Computations performed by adopting a perturbation $\Delta\epsilon_{mn} = -2.0$ kcal/mol in the binary interaction parameter corresponding to residues $[m, n] = [6, 13]$ and $\Delta\epsilon_{ij} = 0$ for all $[i, j] \neq [m, n]$ lead to the free energy surface displayed in Fig. 7 (*a*) for $T = 300$ K. It is useful to examine this surface along with its corresponding contact map in Fig. 7 (*b*). This is a striking change relative to the unperturbed state (Fig. 6), indicating how strong conformational correlations between residues can be in compact structures. The reduction in the free energy of the pair [6, 13] is accompanied by a substantial decrease in that of residues [7, 12], and also [5, 14] to a certain extent, leading to the low energy valley in the free energy surface. This configuration is indicative of two arrays of residues running in opposite directions, i.e., forming an antiparallel β-sheet. It is further observed that even contacts between residues near chain termini, such as [1, 18] and [2, 17] pairs, are now rendered highly favorable. This suggests the occasional extension of the generated regular structure, the antiparallel β-sheet, to several adjoining residues along the chain in some of the accessible conformations. In contrast, a tendency for the formation of a parallel β-sheet involving the three or four residues at both ends is discernible from an examination of the map in Fig. 7 (*b*), as is a significant decrease in the frequency of contacts between nearest neighbors.

The above example concerns a pair of residues located at relatively distant positions in the primary structure of the protein. We can also examine a pair of residues closer to each other, such as [8, 11]. The contact map in Fig. 5 shows that this is neither a high energy pair nor one of the most probable among the lowest order contacts along the diagonal. Its location in the center of the chain would certainly favor its effectiveness in inducing structural reorganizations, but the latter might be expected to be relatively limited to short ranges along the chain. Fig. 8 (*a*) illustrates the free energy surface resulting from an interaction

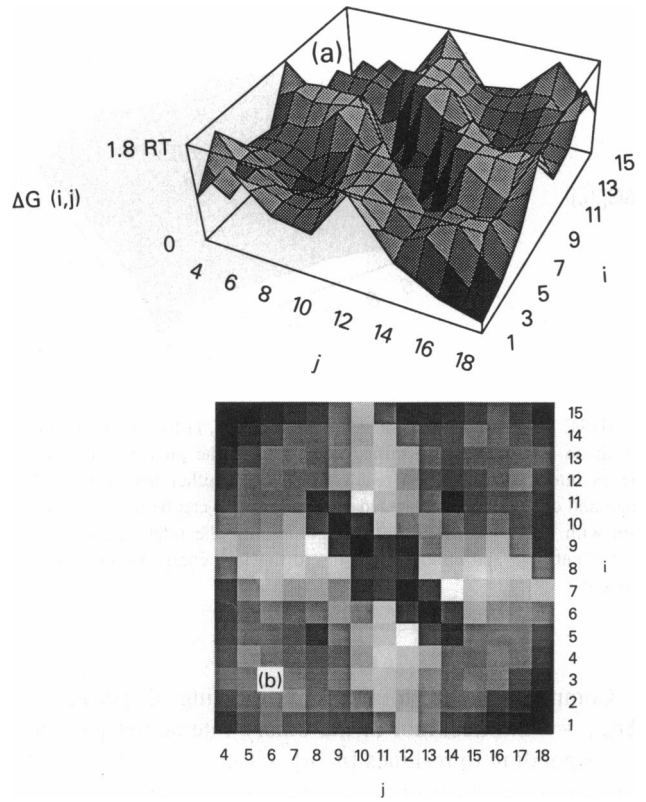

FIGURE 7 (*a*) Free energy surface $\Delta G(i, j)$ for the shape (*B*) in the perturbed state. Residue pair [6, 13] is subjected to an attractive potential of $\Delta\epsilon_{mn} = -2$ kcal/mol at 300 K. A minimum energy valley covering residue pairs [7, 12], [6, 13], and [5, 14] appears, indicative of an antiparallel β-structure. (*b*) Contact energy map corresponding to (*a*). The enhanced probability of occurrence of the new regular structural motifs may be verified from comparison with Fig. 5.

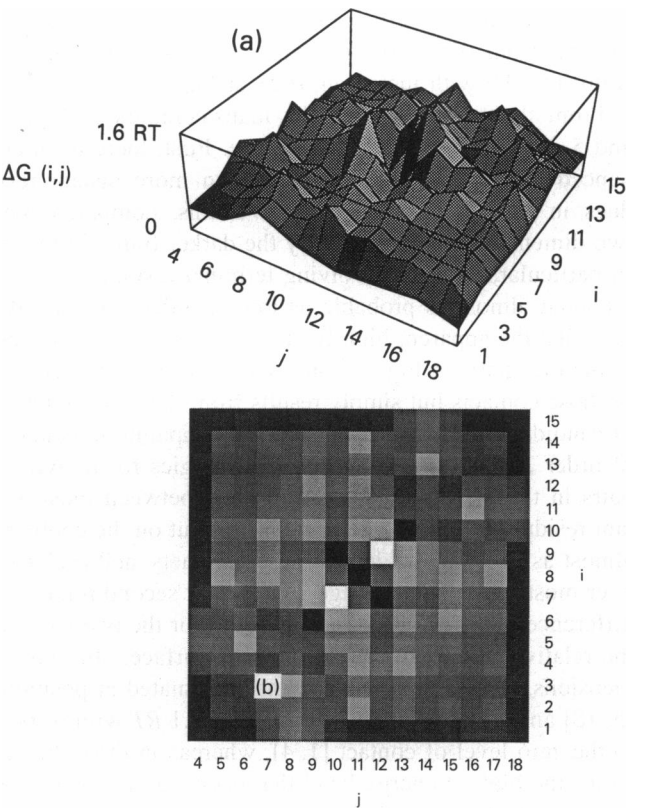

FIGURE 8 (*a*) Free energy surface $\Delta G(i, j)$ for the model chains of shape (*B*) in the presence of a potential of $\Delta\epsilon_{mn} = -2$ kcal/mol at 300 K between residues [8, 11] perturbing the homogeneous system. A series of peaks and wells appearing on the diagonal indicate the enhanced probability for the formation of helical structures. (*b*) Contact energy map corresponding to (*a*) to be compared with that of the unperturbed system shown in Fig. 5.

of −2.0 kcal/mol between residues 8 and 11. The corresponding contact map is given in Fig. 8 (*b*). The new feature which emerges is an array of peaks and wells, still along the diagonal. In particular, the two adjacent contacts [7, 10] and [10, 12], which were originally strongly favored, are practically prohibited, while contacts [6, 9] and [10, 13] are enhanced, in contrast to the relative probabilities of the same residues in the unperturbed state. Basically the helix interactions have shifted their positions, inasmuch as coupled contacts between $[i, i + 3]$ and $[i \pm 2, i \pm 5]$ are indicative of helical structures of various types, as discussed above. Thus, it is interesting to note that the present simple model confirms the widely recognized role of intrinsic propensities, i.e., that of establishing or stabilizing helical regions in globular proteins. However, it should be noted that correlated contacts leading to helical structures might not be so easily achieved if the latter were not supplemented by the constraints imposed by efficient packing in highly compact structures. In Fig. 8 (*b*) some enhancement of antiparallel alignment is also discernible from the lowering of the energy of the pair [7, 12] in response to the attractive interaction of the pair [8, 11].

A similar analysis performed with various shapes indicates the tendency of low order interactions to generate helical structures, while attractions between more distant residues ($k = 5, 7$) induce ridges and valleys in the free energy surfaces, similar to Fig. 7 (*b*). Also, calculations repeated with weaker interactions such as $\Delta\epsilon_{mn} = -1$ kcal/mol between the same residues yield free energy surfaces that are qualitatively very similar to those of stronger interactions in Figs. 7 (*b*) and 8 (*b*). Specific interactions of a given strength and order have an even more pronounced effect on chain topology in shorter chains such as shape (*A*). The residue pairs [4, 9] and [5, 8] are found to be most effective, in this case, in enhancing the respective regular structures identified here with antiparallel sheets and helices.

We considered attractive interactions between a pair of residues separated by five virtual bonds belonging to 16-residue chains with shape (*C*). In contrast to the previous examples, the newly selected residues for perturbation [4, 9] are not situated at symmetric positions with respect to the center of the chain. However, this particular pair exerts a dramatic effect on the contact free energy surface, as illustrated in Figs. 9 and 10. Fig. 9 (*a*) and (*b*) display the free energy distribution $\Delta G_0(i, j)$ of the unperturbed system. A rather flat surface with the maximum located at [4, 13] is seen. The change in free energy $[\Delta G(i, j) - \Delta G_0(i, j)]_{m,n}$ at $T = 300$ K induced by a potential of $\Delta\epsilon_{mn} =$

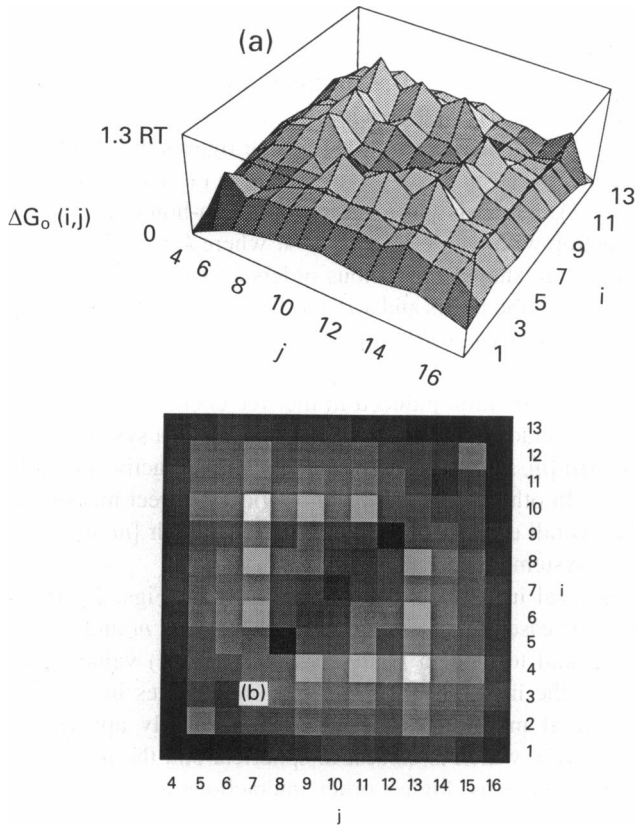

FIGURE 9 (*a*) Free energy surface $\Delta G_0(i, j)$ for the structure (*C*) in the unperturbed state. The surface represents the entropic contribution to free energies and is compared with the enthalpic contribution induced by an interaction of −2 kcal/mol between residues [4, 9] displayed in Fig. 10 (*a*). (*b*) Contact energy map corresponding to (*a*). Minimum energy contacts occur between residues [5, 8] and [9, 12], while the pair [4, 13] is the least probable from the entropic point of view.

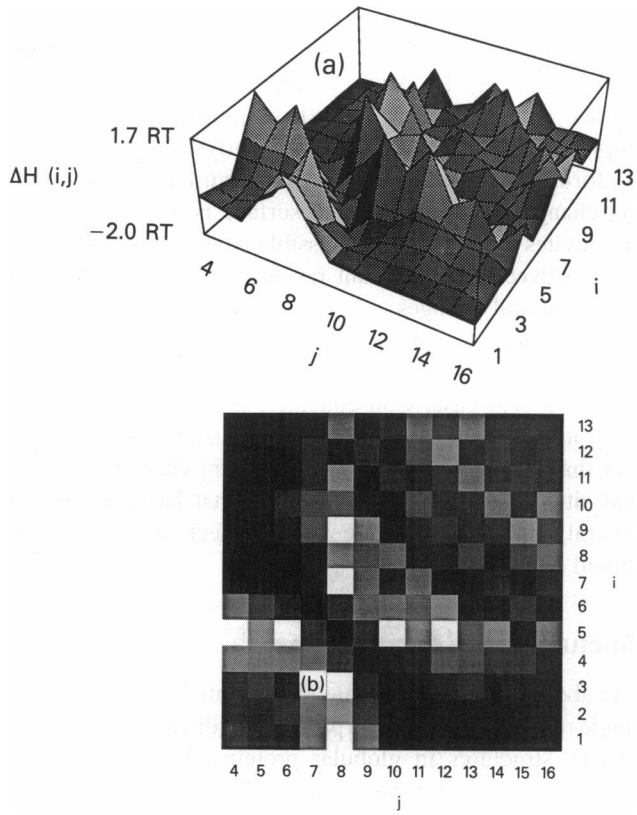

FIGURE 10 (*a*) Enthalpic contribution $\Delta H(i, j)$ to the free energy distribution $\Delta G(i, j)$ of residue pairs [*i, j*] in structure (*C*) in the presence of an attractive potential of 2.0 kcal/mol between residues 4 and 9. Formation of an antiparallel β-sheet associated with residue pairs [3, 10] and [5, 8] is favored. (*b*) Contact map corresponding to $\Delta H(i, j)$ shown in (*a*).

−2 kcal/mol between residues 4 and 9 is shown in Fig. 10 (*a*) and (*b*). This difference represents the enthalpic contribution to free energies, $\Delta H(i, j)$, arising from the non-homogeneous interactions. The heights of the surfaces displayed in Figures 9 (*a*) and 10 (*a*) reveal that the accessible range of enthalpy values $-2.0\ RT < \Delta H(i, j) < 1.7\ RT$, is substantially larger than that of unperturbed (purely entropic) free energy values $0 < \Delta G_0(i, j) < 1.3\ RT$. As a result, the free energy surface of the perturbed system (data not shown) is dominated by the enthalpic contribution and exhibits the same structural characteristics as Fig. 10 (*a*). It was observed that the attraction between residues 4 and 9 favors the coupling of residue pairs [5, 8] and [3, 10], enhancing an antiparallel β-sheet formation again. Examination of the energy surface reveals the possible occurrence of regular structures involving other residues as well at various locations along the chain. It is interesting to note that the attractive interaction between a given pair of residues results in an unfavorable enthalpic effect on the interaction of several other residue pairs. In fact, the residue pairs [1, 8], [3, 8], [5, 10], and [5, 12] are found to experience an enthalpy increase of approximately 1.7 *RT* because of the attractive interaction between residues 4 and 9.

As a final example, we consider shape (*D*), which presents the largest conformational space (8143 distinct SAWs), and the pair of residues [5, 16] which are rather distant from each other. Fig. 11 shows the unperturbed free energy surface $\Delta G_0(i, j)$ corresponding to this case. It is similar to Fig. 6 but is more articulated in general. The

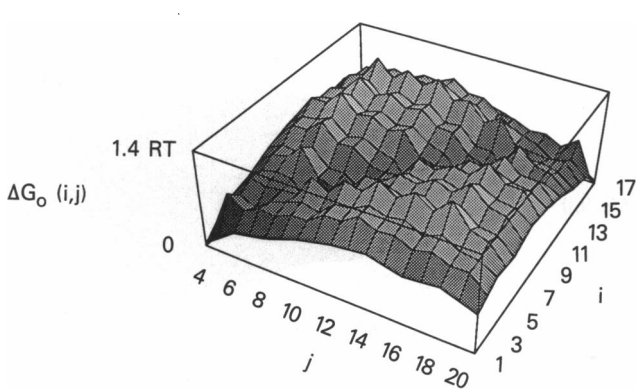

FIGURE 11 Free energy surface $\Delta G_0(i, j)$ for the shape (*D*) in the unperturbed state. The surface represents the entropy contribution to free energies, as does Fig. 6 for shape (*C*). Its perturbed counterpart is shown in Fig. 12.

corresponding free energy surface in the presence of an attractive interaction of 2 kcal/mol between residues [5, 16] is shown in Fig. 12. Although topological contact between residue pair [5, 16] is relatively improbable from the entropic point of view, and consequently a small subset of conformations are affected by the perturbation, the resulting change in the free energy surface is quite remarkable and invites attention to the possible contribution of interactions between rather distant residues in inducing cooperative structural changes.

The surface-to-volume ratios higher than those of real proteins imply that there are fewer internal interactions than in a real case. Consequently, the effect of changing one interaction pair here may have a somewhat larger effect than in the real case. An equivalent change in the actual situation might require a somewhat larger change of several interactions to achieve the effects as large as observed here.

Fluctuations in free energies

The above examples indicate the potential importance of a single residue's character in prescribing distinct secondary or tertiary structures in globular proteins. However, several questions remain to be answered. Which type of interactions, attractive or repulsive, and which pairs among a group of residues, are expected to induce the largest structural changes? Is there a direct correlation between the a priori probability $P_0(m, n)$ of occurrence of a given contact and the extent of disturbance created by $\Delta\epsilon_{mn}$? In other words, is it possible to classify the contacts which are not favorable from the entropic point of view as the least efficient in promoting cooperative structural organization, irrespective of their order? Or, on the contrary, would an attractive interaction between such residues induce the largest perturbation in the free energy surface? A systematic analysis of the perturbation in contact free energies was performed with the aim of answering these questions.

As described above, the effectiveness of a given residue pair $[m, n]$ for driving large-scale cooperative structural changes was estimated by computing the free energy changes $[\Delta G(i, j) - \Delta G_0(i, j)]_{m, n}$ produced for all $[i, j]$ and combining them as in Eq. 7. The resulting RMS fluctuations in free energy $C(m, n)$ are shown in Figs. 13–16 for the four shapes (A) to (D) listed in Table 1. In these figures the abscissa is the index $1 \leq m \leq N - 2$ of residue m ($1 \leq m \leq n - 3 \leq N + 1$), which is participating in a non-homogeneous interaction with residue $n = m + k$, where $k = 3, 5$, etc. The curves are shown for various orders k of interaction, as indicated by the labels and arrows. The ordinate is the reduced fluctuation, $C(m, n)/RT$, induced by an attractive potential of $\Delta\epsilon_{mn} = -2$ kcal/mol at 300 K. $C(m, n)$ may be viewed as the RMS perturbation induced in the free energy of interaction of any randomly selected pair of residues in a system where the pair $[m, n]$ is subjected to the specific attractive potential $\Delta\epsilon_{mn}$. In other words, $C(m, n)$ provides a direct measure of the overall effect of a specific interaction pair $[m, n]$ on the total system.

Several interesting features stand out in Figs. 13–16. In all four cases, interactions between residues m and $m + 3$ are found to lead to relatively large $C(m, n)$ values, indicating the importance of intrinsic propensities in inducing structural motifs. This feature is particularly apparent in shape (B), which is closest to spherical and the most symmetric, whereas, higher order interactions are observed to compete more efficiently in the case of more asymmetric shapes (C) and (D). Residues that are relatively distant ($k \geq 9$) from each other along the chain primary sequence

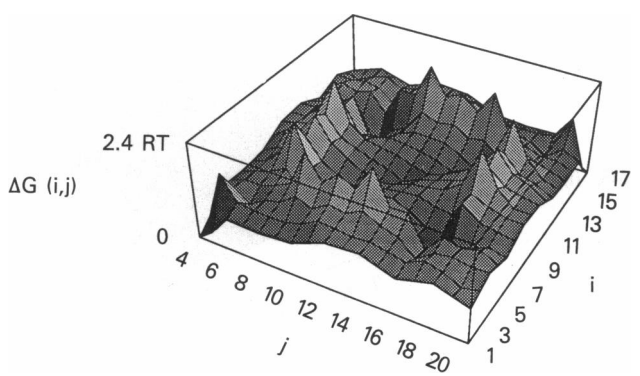

FIGURE 12 Free energy surface for shape (D) in the presence of an attractive interaction of −2.0 kcal/mol between residues [5, 16], an interaction of order $k = 11$, to be compared with the unperturbed surface in Fig. 11.

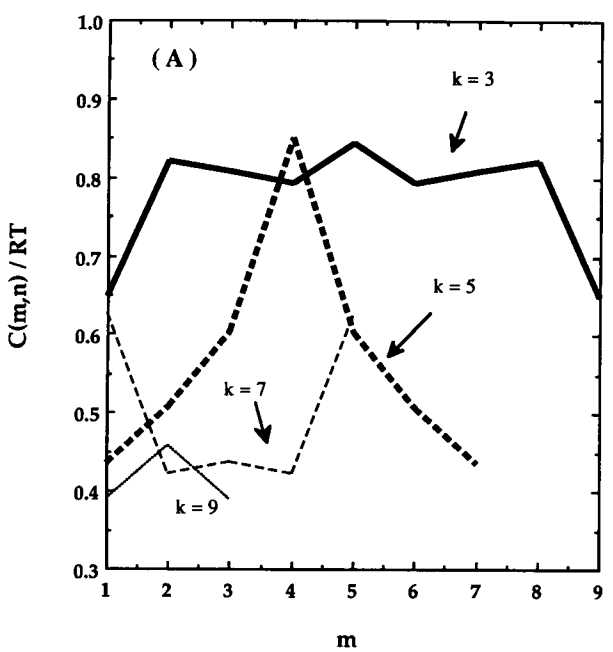

FIGURE 13 RMS fluctuation in free energies, $C(m, n)$, induced by an attractive interaction of −2.0 kcal/mol between residues m and $n = m + k$ for the model chains of shape (A). Curves are computed using Eq. 7 for contacts of order $k = 3, 5, 7$ and 9, as indicated. Lowest order contacts affect the free energies most effectively, whereas only the higher order perturbation induced by the interaction between residues 4 and 9 is observed to be as significant.

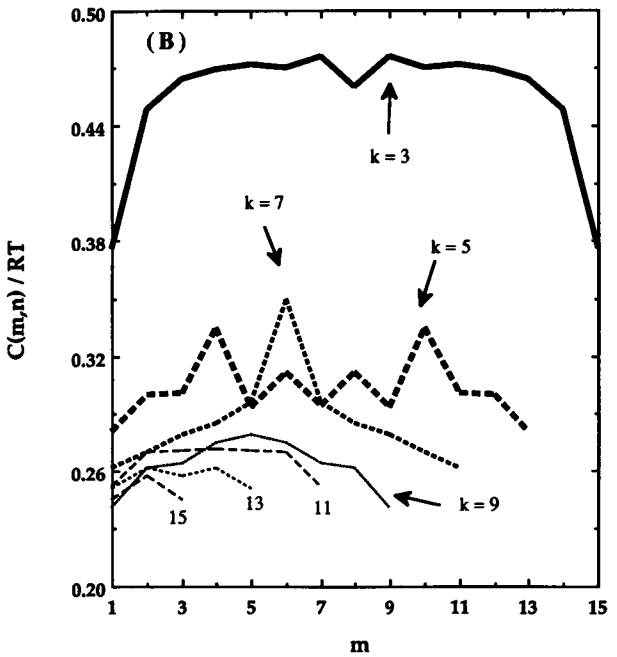

FIGURE 14 RMS fluctuations $C(m, n)$ in free energies due to attractive interaction of −2.0 kcal/mol between residues m and $n = m + k$ for the shape (B). Curves are shown in dimensionless units for various contact orders $3 \leq k \leq 15$, as indicated. Attractions in which one of the terminal residues participates have a weaker effect on the free energy distribution. The effect of lowest order contacts is strongest, mainly due to their high a priori probabilities.

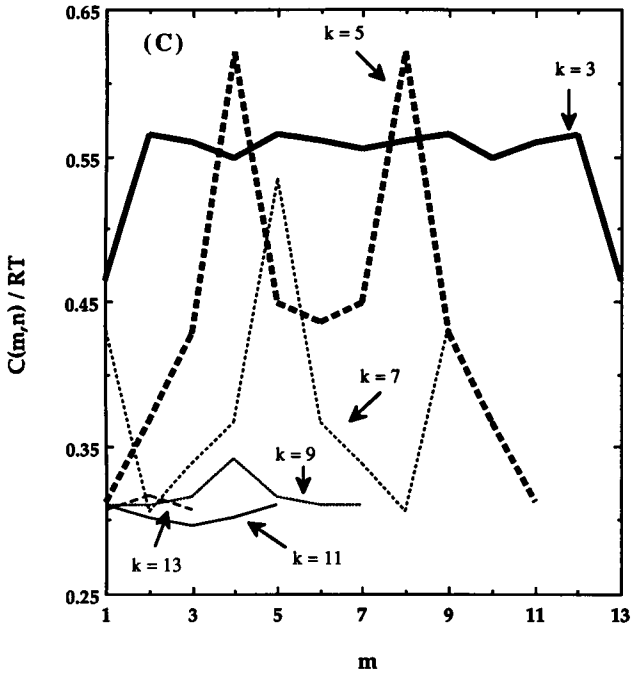

FIGURE 15 $C(m, n)/RT$ values corresponding to an attractive interaction of −2.0 kcal/mol between residues m and $n = m + k$ for the structure (C). Curves are shown for $3 \leq k \leq 13$, as indicated. Residue pairs [4, 9] and [8, 13] are observed to induce the largest fluctuations in free energies.

impart a weaker perturbation in the free energy distribution. Likewise, interactions in which either one of the terminal residues participates always have a relatively weaker effect on the distribution of free energies than do those with mid-sequence positions. This feature may be viewed as a natural consequence of chain connectivity; specific interactions involving terminal residues are presumably accommodated with relatively localized configurational changes, leaving a considerable region of the free energy surface unaffected.

The magnitude of $C(m, n)$ is largest in the shortest chain and decreases with chain length. This is understandable, since $C(m, n)$ results from averaging over all possible pairs, and in longer chains the fraction of pairs $[i, j]$ which are not directly affected by the specific interaction $\Delta\epsilon_{mn}$ is larger. Although some specific conformations lead to large-scale reorganization of the chain, the highest absolute perturbation occurs in the close neighborhood of $[m, n]$. Inasmuch as the fluctuations are evaluated from the ensemble of all accessible conformations, the results should naturally reflect the behavior of the majority of conformations. This feature may be verified from the free energy surfaces presented above.

Several peaks appear in Figs. 13–16, although the magnitude of the specific interaction is identical in all cases. For instance, the attractive interaction between residues [4, 9] and its symmetric counterpart [9, 14] are found to lead to the maximum perturbation in the free energy distributions in the chain (C) of 16 residues. Likewise the pair [4, 9] in shape (A) affects the conformational space quite strongly. It is to be noted that $[\Delta G(i, j) - \Delta G_0(i, j)]_{m, n}$ and consequently $C(m, n)$ values are affected by two factors: 1) the probability that the residue pair $[m, n]$ comes into close proximity (i.e., at a distance d_0, the size of lattice edge in the present model); and 2) the specific location of the interacting pair $[m, n]$ in the particular globular shape. The former is the entropic effect. It is prerequisite for the possibility of a perturbation, i.e., two residues which are not allowed to come into close proximity due to specific chain geometry (or shape) will not affect the free energy distribution, regardless of the magnitude of a specific interaction. Alternately, residue pairs that have a higher chance of being located close to each other, such as residues m and $m + 3$ in general, will lead to larger apparent fluctuations in free energy, inasmuch as they affect a larger fraction of conformations in the whole ensemble. This feature explains the relatively high $C(m, n)$ values obtained for $n - m = 3$ in all cases, as well as the gradual depression of the curves to lower values with increasing k values. The appearance of peaks for a given k, on the other hand, is related to the specific location of the interacting pair in the specific shape. The fact that those peaks emerge in spite of the lower entropic probability of the participating residues is quite significant. If normalized with respect to their a priori probabilities of occurrence $P_0(m, n)$, the effect of those particular pairs might be even more pronounced. As an example, we might consider the two residue pairs [6, 13] and [8, 11] in case (B), which are observed to yield the respective values of 0.35 and 0.46 for

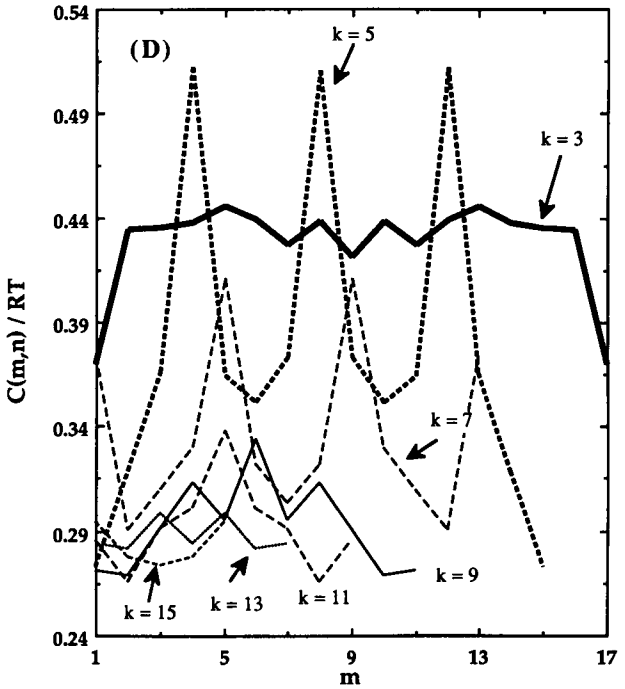

FIGURE 16 $C(m, n)/RT$ values corresponding to an attractive interaction of −2.0 kcal/mol between residues m and $n = m + k$ for the structure (D). Curves are shown for $3 \leq k \leq 15$, as indicated. Three cases of interactions of order 5 exhibit the strongest perturbations.

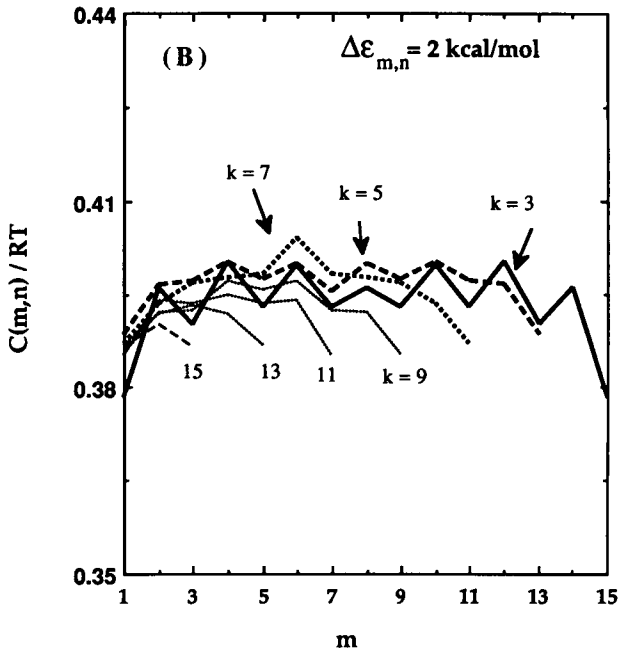

FIGURE 17 $C(m, n)/RT$ values induced by a repulsive interaction of 2.0 kcal/mol between residues m and $n = m + k$ for the structure (B) for $3 \leq k \leq 13$, as indicated. Approximately the same perturbation values are found for all contact orders, regardless of the residue location along the chain, indicating the non-specificity of repulsive interactions.

$C(m, n)/RT$. If those values are normalized with respect to their a priori probabilities 0.009 and 0.016, the effect of residue pair [6, 13] exceeds that of [8, 11] by about 35%. These results rest on the conjecture that the investigated residue pairs are in topological contact. This type of analysis has been performed by Chan and Dill (1989a, 1990a) for compact chains of 27 residues. For generality, we consider here the $C(m, n)$ values that incorporate both entropic and enthalpic contributions.

Repulsive interactions

We studied the effectiveness of repulsive interactions in perturbing the free energy surfaces. Fig. 17 shows the fluctuations $C(m, n)/RT$ resulting from repulsive interactions of $\Delta\epsilon_{mn} = +2$ kcal/mol between residues $[m, n]$ for chain (B). In contrast to the effect of attractive interactions which are quite specific with respect to the location of the embraced residues $[m, n]$, repulsive interactions are observed to lead to approximately the same amount of perturbation, $0.38 \leq C(m, n)/RT \leq 0.41$, regardless of the choice of $[m, n]$. A similar behavior is observed in all of the other shapes. This result, which might seem unexpected at first, can be explained: upon inclusion of attractive interactions between a given pair of residues $[m, n]$ the probability of a small subset of chains is enhanced; this is the subset of conformations in which topological contact between m and n takes place, i.e., in which those residues occupy adjacent sites. The observed behavior in the presence of attractive interactions directly reflects that of this subset, which is now assigned a relatively large statistical weight. In the case of repulsive interactions, the same subset is assigned a low probability of occurrence. Its contribution to the observed behavior is negligible, inasmuch as this subset already presents a small group among all accessible conformations. Thus, the much larger subset of conformations without the perturbed topological contact $[m, n]$ dominates the observed behavior. Exclusion of the chains with a specific contact do not perturb to a large extent the statistical behavior of the ensemble, leading to free energy fluctuations of about the same size, regardless of the location of the residues repelling each other.

This insensitivity of fluctuations to the position of residues repelling each other suggests that repulsive interactions do not induce a specific cooperative structural rearrangement but have a rather localized effect, perturbing only limited regions of the free energy surface. Free energy surfaces computed by adopting $\Delta\epsilon_{mn} = 2$ kcal/mol for various $[m, n]$ confirm this prediction. Two extreme cases may be considered. In one case, the a priori probability of the interacting pair is low. In this case, the perturbation induced by a positive $\Delta\epsilon_{mn}$ is rather limited and diffuse, since the probability of a highly improbable contact is further depressed. The other extreme case is the repulsion between a pair of residues whose contact is highly favorable from the entropic point of view. For instance, we might consider the pair [7, 10] in chain (B), which should be expected to exert the strongest effect on the free energy distribution. Fig. 18 shows that the perturbation induced by a repulsion between residues [7, 10] in shape (B)

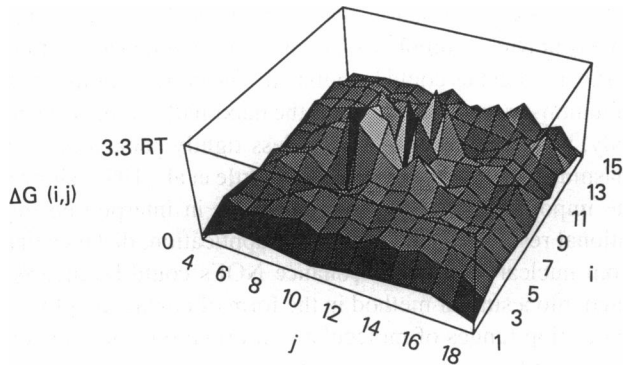

FIGURE 18 Free energy surface for shape (B) in the presence of repulsive interactions of +2.0 kcal/mol between residues [7,10]. Comparison with the unperturbed surface in Fig. 6 and the surfaces of Figs. 7 (a) and 8 (a) resulting from attractive interactions indicates that the effect of repulsive interactions is restricted to a much more localized region of free energy surface.

is confined to a restricted region of the free energy surface, leaving the other regions almost unaffected. These considerations suggest that *intramolecular attractive interactions, rather than repulsive interactions, are more effective in establishing or stabilizing well-defined structural patterns in compact proteins.*

This result leads to the consideration of the important role of the type of interaction, attractive or repulsive, in establishing the preference for a given regular structure by a given chain. Attractive interactions are more specific; they enhance specific structural patterns easily, and their effect strongly depends on the location of residues along the primary sequence. Repulsive interactions, on the other hand, are less selective in favoring specific regular structures, and their effect does not depend on the primary position of residues.

Effectiveness of individual residues

The effectiveness of a given residue m to induce cooperative structural changes in a well-defined shape may also be estimated by considering all of the RMS fluctuations in contact energies, $C(m, n)$, obtained for a given m. A weighted average over all possible pairings for a given m is performed by

$$C(m) = \sum_n C(m, n)P_0(m, n) \qquad (8)$$

where $C(m, n)$ is evaluated from Eq. 7 and $P_0(m, n)$ is the equilibrium probability of contact $[m, n]$ in the unperturbed state given by Eq. 4. Fig. 19 displays $C(m)$ values obtained as a function of position m of the perturbed residue for the chains (A) to (D), as indicated, using $\Delta\epsilon_{mn} = -2$ kcal/mol at 300 K. The points represent the results of simulations (joined by lines to guide the eye). The ordinate values of the curve for the shape (B) is lowered by 0.05 along the vertical axis to achieve visual separation.

In all of the four distinct shapes, terminal residues are found to be the least efficient in perturbing topological free

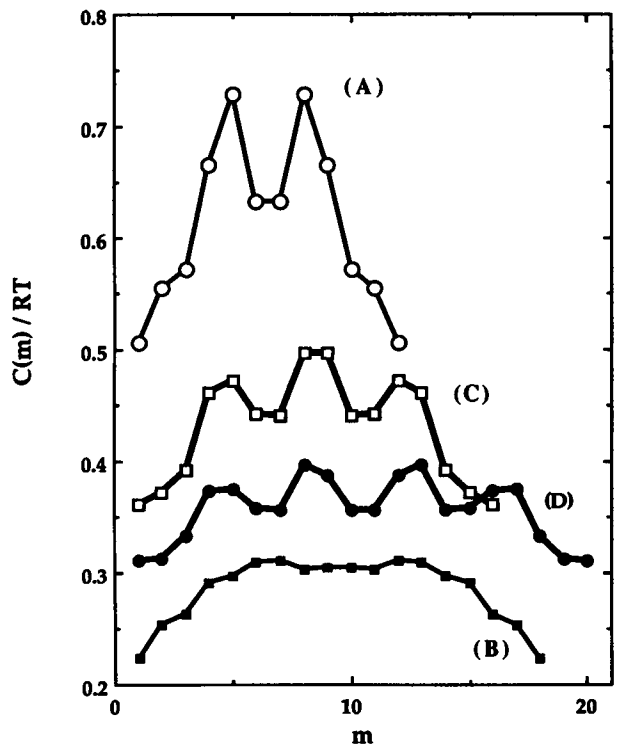

FIGURE 19 Fluctuations $C(m)/RT$ in free energies induced by a given unit m participating in binary attractive interaction (−2 kcal/mol) with any other non-bonded unit in the chain evaluated from Eq. 8. $C(m)$ provides a measure of the strength of perturbation exerted by specific points in the different shapes (A) to (D) (see Fig. 2). Inner units are found to affect more strongly the free energy distribution in all cases as compared with terminal units. Curves for shapes (A), (C), and (D) exhibit a distinctive dependence on m, whereas that for the more symmetric shape (B) is substantially smoother.

energies. However, it is interesting to see that the extent of perturbation, as characterized by $C(m)$, does not necessarily increase smoothly as one proceeds toward inner residues. Certain inner positions along the chain produce relatively weak free energy perturbations. This feature is even more pronounced in shorter chains and asymmetric shapes. Thus, *the potency of a given residue to induce structural changes appears to be closely associated with the shape of the protein to which it belongs.* A smoother dependence of $C(m)$ on m is observable in case (B), but the curve for the asymmetric chain (D) exhibits an elaborate specific structure in spite of the larger number of units in the chain.

The contact maps or free energy surfaces displayed in Figs. 5 (or 6), 9, and 11 for the respective model chains of shapes (B), (C), and (D) in the unperturbed state, as well as those obtained (Bahar and Jernigan, 1994) for two-dimensional compact chains, generally indicate a relatively high frequency of occurrence of contacts involving terminal residues. Yet, the enhancement of the equilibrium probabilities $P_0(m, n)$ of these particular contacts (driven by entropic effects) is not sufficient alone to enable chain ends to effectively perturb free energy distributions. Thus, irrespective of the weighting terms $P_0(m, n)$ in Eq. 8, $C(m)$ seems to preserve the characteristics of $C(m, n)$, mainly weaker

fluctuations in contact free energies induced by specific interactions involving units closer to chain ends. This is another way of stating that enthalpic effects (brought about by an attractive interaction of 1–2 kcal/mol between a given pair of residues) are predominantly important in prescribing the free energy distribution, rather than the entropic effects which lead to relatively flat free energy surfaces.

CONCLUSIONS

Enumeration of all conformations is feasible for relatively short chains whenever each residue site compatible with a given shape can be occupied once and only once (Covell and Jernigan, 1990). A comparison of results for two and three dimensions indicates much similarity but with enhanced frequencies formed in three dimensions for contacts with terminal portions of the chain. A wider variety of shapes than considered here can be generated (Jernigan, 1992).

The above analysis indicates that certain regions along the primary sequence present a higher potential for 1) perturbing conformations and 2) establishing regular structural motifs when subjected to attractive interactions with other residues. These critical positions depend on the particular shape of the globular protein. Inner residues are generally shown to be more efficient in inducing structural changes compared with terminal residues, but the perturbation induced by a given residue does not exhibit a simple dependence on its location along the primary sequence. In real proteins, instead of a homogeneous base of interactions upon which a change is imposed, there is actually a highly heterogeneous background of interactions.

An interesting aspect of intramolecular interactions in compact structures with prescribed shapes is the different character of attractive and repulsive interactions. The effect of attractive interactions strongly depends on the position of the interacting pair, as well as on the particular shape of the protein. The perturbing effect of repulsive interactions is of approximately the same strength, irrespective of the loci of the residues involved, and the induced topological changes are rather localized, in contrast to the highly cooperative response of chain units to attractive interactions.

This study suggests that certain residues in a given protein shape may contribute more effectively to the stabilization of a given structure than other residues, even if they are not necessarily participating in strong specific interactions such as hydrogen bonding or electrostatics. Their capability for imparting substantial structural changes originates from the severe constraints of high packing density within a well-defined restrictive shape. It would be interesting to apply this type of statistical analysis to proteins of known shape and identify regions of utmost sensitivity for inducing large scale cooperative structural rearrangements. Such an approach might prove useful in providing approximate guidelines for protein engineering/design purposes.

Other types of perturbations could be studied. The present method might be applied to investigate the effects of biases corresponding to secondary structure segments, a much larger scale perturbation than considered here. For longer chains in which complete enumeration is not feasible, Monte Carlo approaches could be pursued. The perturbations due to mutations are explored here for the maximally compact states only, but might be extended to less tightly packed models, inasmuch as recent modeling by Shortle et al. (1992) showed the importance of the denatured states in interpreting mutational results. In another type of application, distance data from nuclear magnetic resonance NOEs could be incorporated into a similar method in the form of contact pair biases to develop ranges of molecular conformations that are consistent with the experimental data.

We acknowledge the National Cancer Institute for allocation of computing time and staff support at the Biomedical Supercomputing Center at the Frederick Cancer Research and Development Center.

REFERENCES

Bahar, I., and R. L. Jernigan. 1994. Stabilization of intermediate density states in globular proteins by homogeneous intramolecular attractive interactions. *Biophys. J.* In press.

Barber, M. N., and B. W. Ninham. 1970. Random and Restricted Walks: Theory and Applications. Gordon & Breach, New York. 176 pp.

Chan, H. S., and K. A. Dill. 1989a. Compact polymers. *Macromolecules* 22:4559–4573.

Chan, H. S., and K. A. Dill. 1989b. Intra-chain loops in polymers. Effects of excluded volume. *J. Chem. Phys.* 90:492–509.

Chan, H. S., and K. A. Dill. 1990a. The effects of internal constraints on the configurations of chain molecules. *J. Chem. Phys.* 92:3118–3135.

Chan, H. S., and K. A. Dill. 1990b. Origins of structures in globular proteins. *Proc. Natl. Acad. Sci. USA.* 87:6388–6392.

Covell, D. G., and R. L. Jernigan. 1990. Conformations of folded proteins in restricted spaces. *Biochemistry.* 29:3287–3294.

Flory, P. J. 1982. Treatment of disordered and ordered systems of polymer chains by lattice methods. *Proc. Natl. Acad. Sci. USA.* 79:4510–4514.

Go, N., and H. Taketomi. 1978. Respective roles of short- and long-range interactions in protein folding. *Proc. Natl. Acad. Sci. USA.* 75: 559–563.

Go, N., and H. Taketomi. 1979a. Studies on protein folding, unfolding and fluctuations by computer simulation. III. Effect of short-range interactions. *Int. J. Pept. Protein Res.* 13:235–252.

Go, N., and H. Taketomi. 1979b. Studies on protein folding, unfolding and fluctuations by computer simulations. IV. Hydrophobic interactions. *Int. J. Pept. Protein Res.* 13:447–461.

Ishinabe, T., and Y. Chikahisa. 1986. Exact enumeration of self-avoiding walks with different nearest-neighbor contacts. *J. Chem. Phys.* 85: 1009–1017.

Jernigan, R. L. 1992. Generating general shapes and conformations with regular lattices for compact proteins. *In* Structure and Function, Volume 2: Proteins. R. H. Sarma and M. H. Sarma, editors. Adenine Press, New York. 169–182.

Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* 22:2577–2637.

Kolinski, A., M. Milik, and J. Skolnick. 1991. Static and dynamic properties of a new lattice model of polypeptide chains. *J. Chem. Phys.* 94: 3978–3985.

Kuwajima, K. 1989. The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structures. *Proteins* 6:87–103.

McKenzie, D. S. 1973. The end-to-end distribution of self-avoiding walks. *J. Phys. A.* 6:338–352.

Miyazawa, S., and R. L. Jernigan. 1985. Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules.* 18:534–552.

Pabo, C. 1983. Designing proteins and peptides. *Nature* 301:200–201.

Ponder, J. W., and F. M. Richards. 1987. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J. Mol. Biol.* 193:775–791.

Schmalz, T. G., G. E. Hike, and D. J. Klein. 1984. Compact self-avoiding circuits on two-dimensional lattices. *J. Phys. A.* 17:445–453.

Shakhnovich, E., G. Farztdinov, A. M. Gutin, and M. Karplus. 1991. Protein folding bottlenecks: a lattice Monte Carlo simulation. *Phys. Rev. Lett.* 67:1665–1668.

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.

Skolnick, J., and A. Kolinski. 1990. Simulation of the folding of a globular protein. *Science (Washington DC).* 250:1121–1125.

Taketomi, H., Y. Ueda, and N. Go. 1975. Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions. *Int. J. Pept. Protein Res.* 7:445–459.

Ueda, Y., H. Taketomi, and N. Go. 1978. Studies on protein folding, unfolding and fluctuations by computer simulations. II. A three-dimensional lattice model of lysozyme. *Biopolymers.* 17:1531–1548.

Wall, F. T., and S. G. Whittington. 1969. The density function of end-to-end lengths of self-avoiding random walks on a lattice. *J. Phys. Chem.* 73: 3953–3959.

Yue, K., and K. A. Dill. 1992. Inverse protein folding problem: designing polymer sequences. *Proc. Natl. Acad. Sci. USA.* 89:4163–4167.

Zimm, B. H., and J. K. Bragg. 1959. Theory of phase transition between helix and random coil in polypeptide chains. *J. Chem. Phys.* 31: 526–535.