Stabilization of Intermediate Density States in Globular Proteins by Homogeneous Intramolecular Attractive Interactions

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 On-lattice simulations of two-dimensional self-avoiding chains subject to homogeneous intramolecular attractive interactions were performed as a model for studying various density regimes in globular proteins. For short chains of less than 15 units, all conformations were generated and classified by density. The range of intramolecular interactions was found to increase uniformly with density, and the average number of topological contacts is directly proportional to density. The uniform interaction energy increases the probability of high density states but does not necessarily lead to dominance of the highest density state. Typically, several large peaks appear in the probability distribution of packing densities, their location and amplitude being determined by the balance between entropic effects enhancing more expanded conformations and attractive interactions favoring compact forms. Also, the homogeneous interaction energy affects the distribution of most probable interacting points in favor of the longer range interactions over the short range ones, but in addition it introduces some more detailed preferences even among short range interactions. There are some implications about the characteristics of the intermediate density states and also for the likelihood that the native state does not correspond completely to the lowest energy conformation.

INTRODUCTION

Studies of the denatured states of proteins have established the occurrence of partially folded equilibrium intermediate structures (Ptitsyn, 1987; Kuwajima, 1989). These can be stable states, experimentally observed under mild denaturating conditions (Kuwajima et al., 1976; Dolgikh et al., 1981, 1984; Ikeguchi et al., 1986; Ptitsyn, 1987; Kuwajima, 1989). They are more loosely packed compared with the native state and yet exhibit substantial secondary structure. The term "molten globule" was introduced by Oghushi and Wada (1983) to describe these relatively compact but partially denatured states. The presence of compact intermediate states accumulating before native state formation is also supported by kinetic studies and spectroscopic measurements.

In parallel with such experimental work, a sequential folding mechanism based on intermediates of significant population was proposed by Kim and Baldwin (1982), with more detail than the classical "all-or-none" two-state theory of structural transitions in polypeptides. More recently, the transitions among the intermediate, molten globule, and the native states have been theoretically investigated (Finkelstein and Shakhnovich, 1989; Shakhnovich and Finkelstein, 1989). In their view, denaturation to the molten globule state is postulated to be driven by the disruption of the tight packing of the side chains of amino acid residues. The rotational isomerization of side chains is thought to take place only if the intramolecular free volume exceeds a threshold value large enough to entropically compensate the loss in attractive non-bonded energy (Shakhnovich and Finkelstein, 1989). In one estimate, the increase in volume or decrease in density of the molten globule state relative to the native state has been stated (Ptitsyn and Semisetnov, 1991) to be approximately 30%. In another recent statistical mechanical analysis of the stabilities of the three states of globular proteins (unfolded, compact denatured, and native), Alonso et al. (1991) analyzed the role and strength of hydrophobic, electrostatic, and entropic contributions. Reviews of the dominant forces in proteins, either during the process of folding (Dill, 1990) or in denatured states (Dill and Shortle, 1991), have appeared. Conforming with Kauzmann's early viewpoint (Kauzmann, 1959), hydrophobic interactions are shown to play a major role in protein stability, whereas the important effects of conformational entropy and electrostatic interactions, both opposing protein folding, are also recognized. In an alternative approach, Ben-Naim and co-workers (Ben-Naim, 1989; Ben-Naim et al., 1989a, b) argued that intramolecular hydrophilic interactions and direct interactions with water are expected to affect highly specific biochemical processes such as protein folding more effectively than hydrophobic interactions. From an analysis of long-range contact residue types, Miyazawa and Jernigan (1985) concluded that the strongest interactions were among hydrophobic residues, but the most specific ones were among the hydrophilic ones. This qualitative difference might lead one to expect a hydrophobic collapse before an optimal arrangement of polar residues.

There also could be another reason for placing polar residues on the exterior and, by default, hydrophobic residues in the interior. Proteins also have a network of polar hydrogen bonds in their interiors. Earlier, Wada and Nakamura (1981) suggested that helical dipoles, i.e., the interior electrostatics, are arranged in a highly favorable way. If the polar side chains were to be randomly embedded within such an interior, they would reduce the favorable interactions in the hydrogen bond network. Such a disruption would be especially

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

0006-3495/94/02/454/13 $2.00

large in the protein interior, which is generally perceived to be a medium of rather low dielectric constant and consequently stronger electric interactions. Nevertheless, we note that, in aqueous media, the net effect of interactions between amino acid residues, whether of hydrophilic or hydrophobic origin, is invariably the occurrence of intramolecular attractive potentials enhancing the close packing of the globular structure.

Generally, the thermodynamic picture that emerges from all studies is that of a delicate balance between energetic interactions in favor of the compact state and entropic driving forces that oppose compactness. This balance determines the preferential state, denatured, molten globule, or native, of the protein, under well defined conditions. In fact, calorimetric measurements indicate that the native states are only marginally stable relative to the unfolded states (Privalov, 1979).

Thus, two major factors need be considered in the crudest model of protein structures. The first is entropic and is associated with the conformational freedom of the macromolecule, which inherently favors a loosely packed organization of the residues. This is the same factor that manifests itself in the elastic response of polymer chains (Volkenstein, 1963; Flory, 1969). The second is the energetic/enthalpic interaction between residues, with two principal aspects. 1) Repulsive interactions operate if residues are too close to each other, i.e., van der Waals overlaps. In particular, spatial overlap of two or more residues is not possible physically. This effect is commonly referred to as the excluded volume effect in polymer statistics (Yamakawa, 1967). 2) Nearby non-bonded residues interact directly with one another through an attractive potential. This effective attraction is particularly evident for hydrophobic residues in an aqueous environment. It would not be so apparent for residues in good solvents, where monomeric units could have a favorable contact energy with solvent.

In the present work, the probability distribution of various packing densities was analyzed for various length chains, by simulating self-avoiding walks (SAWs), representative of protein chains, on a planar square lattice. The perturbation in the density distributions brought about by various strengths of attractive interactions was considered. The possible occurrence of well populated intermediate density states with substantial structural organization was investigated, as well as various characteristics of their organization. Two-dimensional contact maps were used to analyze structural similarities between different density regimes, just as the crystal structures of different proteins were compared. Two important questions asked were: 1) What and how well defined are the characteristics of intermediates? and 2) Do the structural characteristics of the intermediates persist into the native state?

The approach here is simple but includes the major factors of conformational entropy, excluded volume, and attractive interactions that predominantly affect globular structures of proteins. The important role of such low resolution models and computational methods for exploring the type and strength of factors which lead to stable tertiary structures in globular proteins have been studied recently (Jernigan, 1992). In contrast to the classical methods of polymer conformational statistics (Flory, 1969), which have been widely used as the basic model for molecular dynamics simulations of biological systems (Brooks et al., 1988), the atomic-level details of the structures are not considered in most such on-lattice simulations of proteins. Many of these studies have aimed at obtaining correct overall folds rather than focusing on all atomic details (Covell and Jernigan, 1990; Hinds and Levitt, 1992). In general, each regular site is identified with at most one amino acid residue, which implicitly takes into consideration the intrinsic excluded volume effect of real chains. It is noted, however, that in a further simplification to include more than one residue at one lattice point to consider larger proteins, multiple occupancy was allowed in recent lattice simulations by Hinds and Levitt (1992). The general fold and arrangements of intramolecular contacts corresponding to specific proteins were successfully captured in that low resolution approach. Thus, one main advantage of low resolution models appears to be the regularization and reduction of conformational space, thus improving the possibilities for understanding the factors affecting specific folding patterns or those essential to stabilize a particular structural motif.

In the case of short chains, a systematic analysis of all possible configurations in a geometrically approximate but numerically complete set that includes the native state is feasible by the complete enumeration approach. An excellent example is the work of Chan and Dill, whose exhaustive simulations of compact short chains in square (Chan and Dill, 1989a, b) and simple cubic (Chan and Dill, 1990a, b) lattices have provided important insights into the appearance of secondary structures, such as helices and sheets, in globular proteins. From their studies, dense packing and/or restricted conformational freedom emerges as a major factor responsible for the regular components of the internal architecture of globular proteins. Hydrogen bonding is thought to come into play only at the stage of stabilizing some of the regular patterns already driven by steric packing, in agreement with previous work (Jaenicke, 1987). Those lattice simulations are also notable in elucidating the preferential occurrence of shorter helices in compact globular proteins, in contrast to the more common longer helix behavior of homogeneous polypeptides in solution, and in providing estimates of the relative proportions of different types of secondary structures observed in the folded state.

MODEL AND METHODS

General approach

Polypeptide chains are represented as SAWs on a two-dimensional square lattice. A given protein molecule consists of N bonds connecting $N + 1$ units, indexed from 1 to $N + 1$, each occupying a lattice site. It is convenient to identify each unit sequentially with the individual amino acids in the sequence, as has been done in a large number of previous simulations of proteins on a lattice. Alternatively, each unit could be viewed as a group of several residues closely associated with each other, by hydrogen bonding or other specific interactions, so as to be considered as a single united entity. The position vector of the ith unit along the chain is denoted as $\mathbf{r}_i$. The units are connected by N virtual bonds, each of length d_o, equal to the size of lattice

cells. The geometry of the lattice restricts the available bond angles at each step to the three discrete values ± 90° and 180°. A given chain conformation is characterized by the set of position vectors corresponding to chain units and is represented as $\{\mathbf{r}\} \equiv \{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \ldots, \mathbf{r}_{N+1}\}$. Chain length refers to the total number of units in a given chain. All sterically allowable conformations were generated for chain lengths up to $N + 1 = 14$ units. Exhaustive simulations of all SAWs on regular lattices have been extensively used in polymer statistics (Barber and Ninham, 1970). Their application to the structural analysis of globular proteins has proven to be quite informative (Taketomi et al., 1975; Go and Taketomi, 1978, 1979a, b; Chan and Dill, 1989a, 1990b).

Intramolecular interactions

The binary interaction between non-bonded units i and j is given by the square well potential

$$E(\mathbf{r}_i, \mathbf{r}_j) = \epsilon_{ij}\delta(d_o - |\mathbf{r}_i - \mathbf{r}_j|) + \epsilon_\infty\delta(\mathbf{r}_i - \mathbf{r}_j) \quad (1)$$

Here $|\mathbf{r}_i - \mathbf{r}_j|$ is the magnitude of the separation vector $\mathbf{r}_i - \mathbf{r}_j$ between units i and j. δ is the Kronecker delta, which is equal to 1 if its argument is equal to 0 and is 0 otherwise. ϵ_∞ is the potential occurring in the case of a complete overlap of residues i and j, and here always has the hard sphere value of infinity. It accounts for the excluded volume effect present in SAWs. ϵ_{ij} is the energy parameter corresponding to the specific interaction between units i and j. A negative value is assumed for ϵ_{ij} to account for the general attractive interactions between residues, as described above for proteins. In the homogeneous system which is the subject of the present work, we do not distinguish between individual units; therefore, the subscript ij on ϵ_{ij} can be omitted. However, we choose to leave the subscripts for generality, anticipating the consideration of specific interactions to be treated in the accompanying paper. It is clear from the above equation that units interacting through this attractive potential are only those occupying neighboring sites on the lattice, although they are not directly (by a single bond) connected along the chains. Conforming with previous terminology, this type of contact is referred to as a topological contact. Topological contacts between units separated by k bonds along the chain sequence are referred to as contacts of order k, as previously defined (Chan and Dill, 1989a).

In analogy to the expression for binary interactions, the energy for three-body interactions among chain units i, j, and k is

$$E(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) = \epsilon_{ijk}\delta(d_o - |\mathbf{r}_i - \mathbf{r}_j|)\delta(d_o - |\mathbf{r}_i - \mathbf{r}_k|) \quad (2)$$

This expression represents the extra energy, in addition to that already existing between pairs, arising from the simultaneous linear contact of three units. The generalization of the above expression to n-body interactions is straightforward. In a two-dimensional square lattice with only nearest neighbor lattice points interacting, no direct many-body interaction involving more than $z = 4$ units, the lattice coordination number, is possible. Such a quaternary cluster of interactions has not been considered here. Second, as is shown, even ternary interactions are found to be of secondary importance, which does not justify extensions in such a direction.

A given chain with conformation $\{\mathbf{r}\}$ is then subjected to the overall conformational potential $E\{\mathbf{r}\}$

$$E\{\mathbf{r}\} = \sum_{i,j}[\epsilon_{ij}\delta(d_o - |\mathbf{r}_i - \mathbf{r}_j|) + \epsilon_\infty\delta(\mathbf{r}_i - \mathbf{r}_j)] + \sum_{i,j,k}\epsilon_{ijk}\delta(d_o - |\mathbf{r}_i - \mathbf{r}_j|)\,\delta(d_o - |\mathbf{r}_i - \mathbf{r}_k|) \quad (3)$$

The above summations were performed over all distinct non-bonded units participating in binary and ternary interactions.

Packing densities

The simulations in two dimensions were performed according to a chain growth algorithm which eliminates, at each step, all of the self-intersecting paths as well as all conformations related to each other by rigid body rotations (or mirror reflection). This leads to a substantial reduction in the conformational space compared with the random walk generations which lead to $O((z-1)^N)$ conformations. The decrease in the number of attainable conformations is the basis of the inherent free energy gain associated with the excluded volume effect. Likewise, the variations in the population of various packing density regimes give an indication of the entropic free energy gain or loss associated with the compactness or expansion of the chain.

Inasmuch as the probability distribution of various packing densities is of interest, the generated chains can be classified according to their compactness. The measure of packing density adopted in previous lattice simulations (Chan and Dill, 1989a; 1990a) was taken to be the number N_c of topological contacts in a given conformation, relative to that ($N_{c,max}$) occurring in the most compact state. This ratio was referred to as compactness. In the present work, a somewhat different approach was undertaken. Each chain was assigned a number density $\rho\{\mathbf{r}\}$ on the basis of its length and the spatial distribution of its individual units. Clearly, the analog of volume in two dimensions is simply the surface $S\{\mathbf{r}\}$ value for a particular conformation. For a chain of $N + 1$ residues in the conformation $\{\mathbf{r}\}$, $\rho\{\mathbf{r}\}$ is found from the ratio

$$\rho\{\mathbf{r}\} = (N + 1)/S\{\mathbf{r}\} \quad (4)$$

The estimation of $S\{\mathbf{r}\}$ used here is illustrated in Fig. 1. The method closely resembles that of Lee and Richards (Lee and Richards, 1971; Richards, 1977) for assigning surface areas/volumes to monomers. Accordingly, the surface of a given protein is determined from the loci of the center of a probe sphere that is rolled around the protein. The rolling ball is usually the size of a solvent molecule. Here, a ball the size of an average residue was adopted.

Three distinct conformations of a chain of 13 units are displayed in parts (a) to (c) of Fig. 1, and the contour of $S\{\mathbf{r}\}$ in all cases are drawn. By adopting the value $d_o = 1$, the corresponding densities are calculated as (a) 0.634, (b) 0.628, and (c) 0.621. On the other hand, the lowest attainable $\rho\{\mathbf{r}\}$ value for a chain of the same length is 0.480, corresponding to the fully extended state. It should be noted that the range of density values estimated by the present approach is relatively narrow compared with the differences in the dimensions of real expanded or collapsed chains, and real atoms for a complete residue placed at each point would impart a substantially increased roughness that would considerably change these surface areas. The results here should be interpreted with these aspects of the model in mind. The present estimation of the space allocated to a given chain is approximate but is found to provide a systematic measure of overall packing densities, being most effective in discriminating among different compact shapes of globular structures, as shown below.

In terms of the previous measures of compactness, the three configurations (a) to (c) would not be distinguishable, inasmuch as all three have the same number (six) of topological contacts, as shown by the dashed lines. This represents in fact the maximum number of contacts for a chain of 13 units. However, the present approach does afford the possibility of distinguishing among the three conformations by assigning different $\rho\{\mathbf{r}\}$ values in the order (a) > (b) > (c), which conforms with their relative surface contour lengths shown in Fig. 1. This latter quantity, which would be equivalent to the surface exposure in three dimensions, is generally recognized as an important characteristic of globular proteins. Other examples of configurations with identical numbers N_c of topological contacts but distinguishable number densities were observed in simulations, which support the adoption of the present approach as an appropriate measure for analyzing globule densities. Meanwhile, it is shown below that there is an exact linear relationship between the topological number N_c of contacts and the above number densities, provided that both quantities are averaged over the set of conformations ranging within well-defined packing density regimes.

In this way, a set of discrete density values can be obtained for each chain length. Table 1 summarizes the number N_ρ of discrete density values attainable by a given chain for the entire collection of accessible conformations. Those numbers clearly increase with N because of the larger variety of shapes possible with longer chains. The third column in Table 1 lists, on the other hand, the maximum number $N_{c,max}$ of topological contacts in each case. Both quantities, each of discrete densities or each number of the attainable topological contacts, may be used to characterize the chain compactness. The first is adopted here, whereas the second has been commonly used in previous work. From an examination of these values, it is clear that

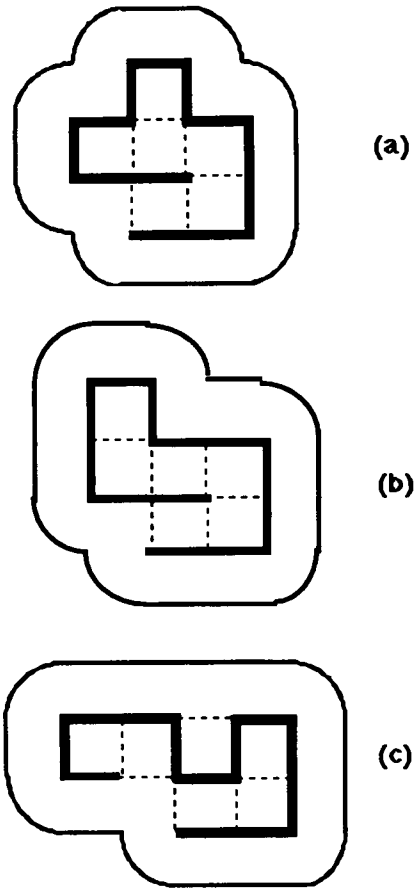

FIGURE 1 Schematic representation of three compact structures of a chain of $N + 1 = 13$ units on a square lattice. All chains have the same number (six) of topological contacts, indicated by the dotted lines. The curved contours indicate the perimeters enclosing the surface S allocated to the chains. They are used to obtain the surface densities (*a*) $\rho = 0.634$, (*b*) $\rho = 0.628$, and (*c*) $\rho = 0.621$ for the packing densities in the three conformations.

a substantially more detailed classification of various chain packing densities is possible by considering $\rho\{\mathbf{r}\}$ values rather than N_c values.

For a statistical analysis of the probability distribution of $\rho\{\mathbf{r}\}$, the total accessible density range for each chain length is divided into an equal number (seven) of intervals of size $\Delta\rho$. Chain conformations leading to distinct density intervals $\rho \pm \Delta\rho$ are grouped together to evaluate the probability distribution function $P(\rho)$ as

$$P(\rho) = Z^{-1} \sum_{\{\mathbf{r}\,|\,\rho\}} \exp(-E\{\mathbf{r}\,|\,\rho\}/RT) \tag{5}$$

where R is the gas constant and T is the absolute temperature. The notation $\{\mathbf{r}\,|\,\rho\}$ refers to the subset of conformations whose packing densities range in the particular interval $\rho \pm \Delta\rho$. The above summation is performed over all conformations belonging to that subset. Z is the conformational partition function evaluated as usual from the summation of the Boltzmann weights of all of the generated SAWs.

Contact maps

Two-dimensional maps of inter-residue separations have proven to be quite useful in reflecting secondary and tertiary structures of globular proteins (Liljas and Rossmann, 1974). The axes represent therein the indices of the amino acid residues ordered as in the primary sequence, and contours of equal separation between residues are drawn. The utility of this representation lies in detecting long-range topological neighborhoods between pairs of residues which are well separated along the chain sequence. In this respect, it contrasts to two-dimensional contour energy plots, which account only for the interdependence of locally connected bonds. Another principal advantage of the maps is that secondary structures such as α-helices and β-sheets stand out strikingly, thus allowing for a global analysis of structural similarities between different globular proteins. α-Helices appear as lines of slope -1 near the diagonal. Parallel β-sheets have slope -1 but are located farther from the diagonal, and antiparallel β-sheets have a slope of $+1$.

TABLE 1 Numbers of attainable discrete densities and maximum numbers of topological contacts per chain for different chain lengths

Chain length $(N + 1)$	$N\rho$	$N_{c,max}$
6	8	2
7	11	2
8	17	2
9	21	4
10	28	4
11	36	5
12	44	6
13	51	6
14	62	7

This type of representation has been used by Chan and Dill (1989a, b; 1990a, b) for analyzing the topological contacts between residues in the chains simulated on regular lattices. Accordingly, contacts are assigned free energies, depending on the reduction of the conformational space brought about by their occurrence. Those free energies are readily calculated from the ratio of the number of chains subject to a given contact to the total number of generated SAWs. In the present case of chains subject to intramolecular interactions, Boltzmann weighting factors are used to evaluate the probability of occurrence of particular conformations, leading to the expression

$$\Delta G(i,j) = -RT \ln\left[Z^{-1} \sum_{\{\mathbf{r}\,|\,[i,j]\}} \exp(-E\{\mathbf{r}\,|\,[i,j]\}/RT) \right] \tag{6}$$

for the change in free energy due to topological contact between units i and j. In analogy to the notation used in Eq. 5, $\{\mathbf{r}\,|\,[i,j]\}$ denotes the subset of conformations in which units i and j are in contact. The summation is performed over all conformations in this subset. The term in brackets represents the probability of occurrence of a conformation having topological contact between units i and j. This quantity is denoted as $P(i,j)$.

For purposes of characterizing the conformations of intermediates as well as compact states, the present approach may be extended to analyze the free energy changes associated with specific contacts in well defined density regimes. Thus considering a given density interval $\rho \pm \Delta\rho$, an expression analogous to Eq. 6 reads

$$\frac{\Delta G(\rho;i,j)}{RT} = -\ln\left[Z^{-1}P(\rho)^{-1} \sum_{\{\mathbf{r}\,|\,(\rho;i,j)\}} \exp(-E\{\mathbf{r}\,|\,(\rho;i,j)\}/RT) \right]$$
$$= -\ln[P(\rho;i,j)\,|\,P(\rho)] \tag{7}$$

Here $\{\mathbf{r}\,|\,(\rho;i,j)\}$ stands for the conformations having contact $[i,j]$, within the studied density interval. $\Delta G(\rho;i,j)$ represents the free energy change associated with contact $[i,j]$, provided that the density interval $\rho \pm \Delta\rho$ is presumed. Thus, the last expression in square brackets in Eq. 7 is the conditional probability of occurrence of a chain with contact $[i,j]$, in the specified density range.

RESULTS FROM CALCULATIONS AND DISCUSSION

Probability distribution of packing densities

To elucidate the relative contributions of the entropic and enthalpic effects on the probability distribution of packing

densities, simulations were carried out for two major cases: 1) systems subject to no other intramolecular interaction than volume exclusion (or a hard-sphere interaction potential with the interaction energy parameters ϵ_{ij} and ϵ_{ijk} taken to be zero), and 2) systems subject to attractive intramolecular interactions. The first is interpreted as the chain entropic contribution. The second yields the perturbations brought about by the introduction of homogeneous intramolecular attractive interactions. Inasmuch as we focus here on nonspecific interactions, a constant value of $\epsilon_{ij} = \epsilon$ is adopted for interactions among all units i, j of the chain. Ternary interactions will not be included in the calculations except in the last section below. Results are expressed in terms of dimensionless reduced energy parameters $\epsilon^* = \epsilon/RT$. In this way, the representation of interaction energies can be alternatively interpreted in terms of temperature.

The probability distribution functions $P(\rho)$ calculated for various chain lengths, $6 \leq N + 1 \leq 14$, by taking $\epsilon_{ij} = 0$, are shown in Fig. 2. Solid (odd) and open (even) circles represent the results of the simulations. They are connected by lines to guide the eye. On the right, the total number of distinct SAWs generated for each chain length are listed, which are in agreement with previous work (Chan and Dill, 1989a). The striking multi-modal shape exhibited by the shorter chains ($N \leq 9$) was the feature which initially motivated the present study. However, the latter is found to vanish gradually as larger sized chains are considered. The question of the possible preference for well defined density regimes, based upon a simple consideration of entropic and homogeneous intramolecular attractive effects, was posed.

The counterpart of Fig. 2 obtained in the presence of attractive interactions of magnitude $\epsilon_{ij}/RT = \epsilon^* = -0.84$ is shown in Fig. 3 for $9 \leq N + 1 \leq 14$. In general, the packing densities are shifted toward higher values, as a natural consequence of the attractive interactions favoring more compact structures.

A closer examination of the effect of various attractive interaction strengths is shown in Fig. 4, where the perturbation in the probability distribution of packing density for two distinct values of the reduced energy parameters is shown for a chain of 9 units. The heavy dashed curve represents the unperturbed distribution function. The light dashed and smooth curves correspond to the cases $\epsilon^* = -0.84$ and -1.68, respectively. It is clearly shown that the increase in the attractive interaction or decrease in temperature significantly changes the distribution function and enhances the probability of the denser states. The most compact state corresponding to the point on the curves most to the right does not so readily become the most probable due to its intrinsic low population. The value of $\epsilon^* = -0.84$ corresponds, for example, to an attraction of 0.5 kcal/mol at 300 K, which may be considered a rather weak interaction. However, its effect on the distribution function is dramatic.

These results invite attention to the intermediate density states that can be stabilized by the intramolecular attractive interactions. In the case of longer chains, the structured shape of the probability distribution function is relatively

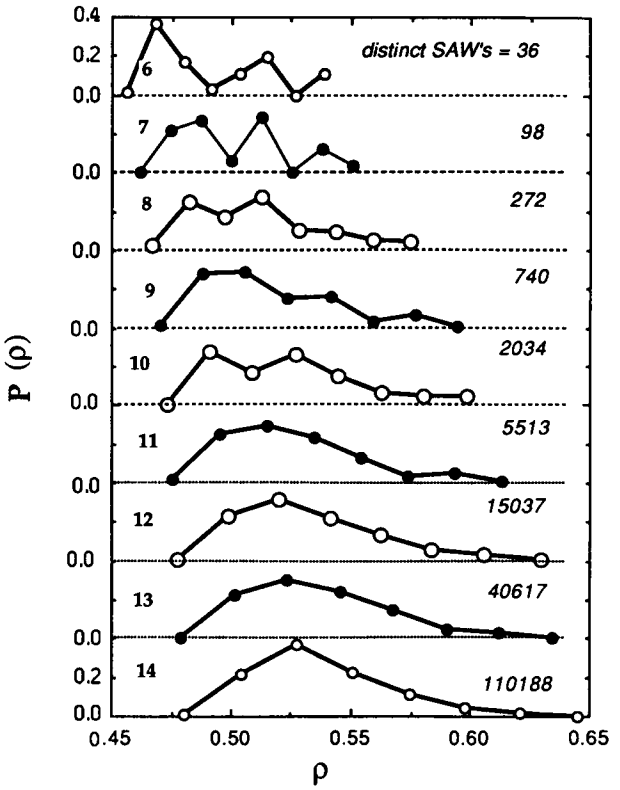

FIGURE 2 Probability distribution of densities for model protein chains of $6 \leq N + 1 \leq 14$ units in the absence of an intramolecular attractive potential. The complete number of distinct self-avoiding chains generated for each chain length is indicated on the right. Open (even) and solid (odd) circles represent results from simulations averaged over conformations in the seven density intervals within the full density range accessible to each chain. Points are connected to guide the eye.

smoothed, as illustrated in Fig. 5, but the appearance of a highly probable high-density domain, with packing efficiency close to that of the most compact state, is even more pronounced. In addition, for values used here, the most compact state is not the most favored state. The observations lend support to the simple hypothesis that favorable, even homogeneous, intramolecular interactions could be the main source of stabilization of intermediate compact but denatured states in proteins. It also supports the view that the most compact conformation may not be the most favorable.

The appearance of the above peaks in the distribution functions rests on a sensitive balance between entropic and enthalpic effects. The substantial decrease in the conformational space as one moves into a more compact regime may or may not be counterbalanced by intramolecular interactions, depending on the strengths of the latter. If we confine our attention to the two most compact density intervals for chains with $N > 10$, for example, the total number of accessible SAWs is found to decrease by a factor of about 5, as one moves into this higher density range. If the newly reached compact regime allows for the occurrence of an additional topological contact, which may lower the free energy by more than $-RT \ln 5$, then the most compact state will become preferred, because the attractive potential more than

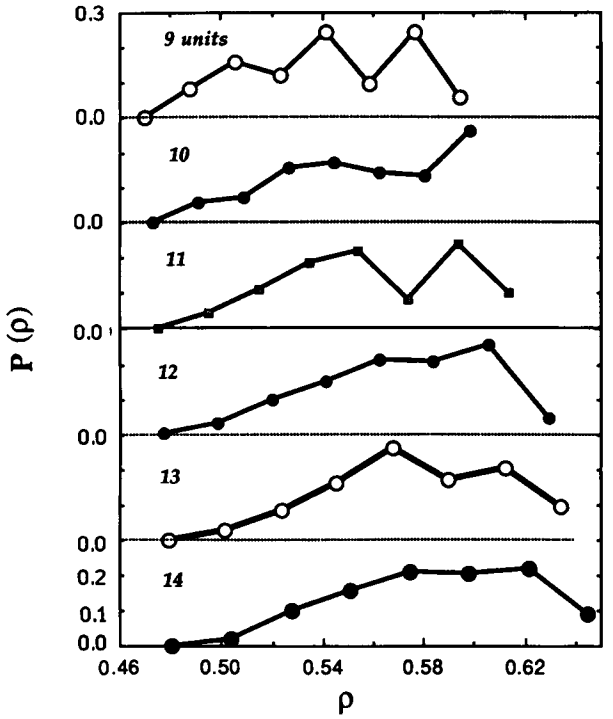

FIGURE 3 Probability distribution of densities for model protein chains of $9 \leq N + 1 \leq 14$ units in the presence of attractive interactions of 0.5 kcal/mol between residues at 300 K (see legend to Fig. 2). In general, the peaks are shifted toward larger values compared with Fig. 2, because of inter-residue attractions. The multi-modal character of the distribution functions corresponding to longer chains also is more pronounced.

overcomes the entropic effect opposing compact forms. Otherwise, the less dense structures will dominate. The attractive interactions of $\leq 1.68\ RT$ adopted in the above calculations

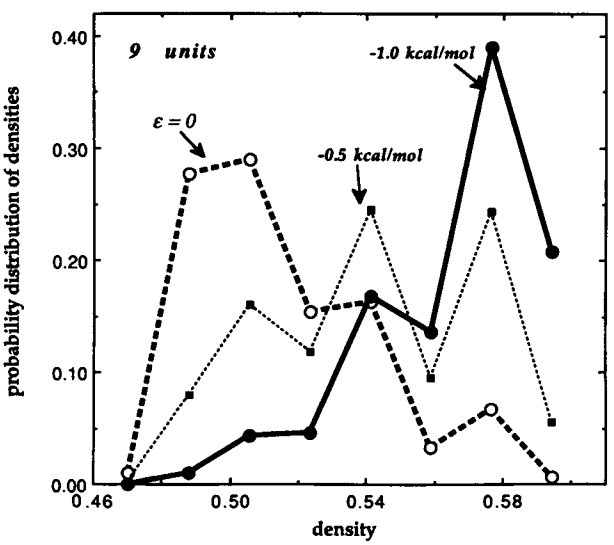

FIGURE 4 Influence of attractive interactions of various strengths on the probability distribution of densities for the model chains of 9 units. Curves are shown for $\epsilon = 0$, −0.5, and −1.0 kcal/mol, $T = 300$ K. Several peaks are observed, indicating intermediate density regimes can be stabilized by homogeneous attractive potentials.

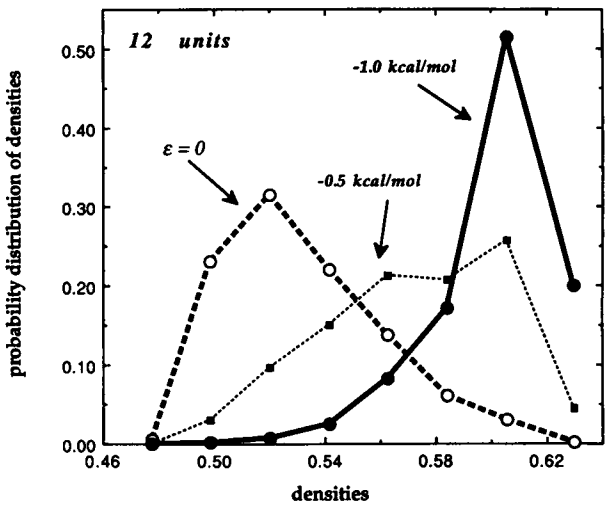

FIGURE 5 Influence of attractive interactions of various strengths on the probability distribution of densities for model chains of 12 units. Curves are shown for $\epsilon = 0$, −0.5, and −1.0 kcal/mol, $T = 300$ K. Next-to-highest density regime is enhanced by the increase in attractive interactions.

are large enough to enhance the probability of the two highest density regimes, yet relatively weak such that a preference for the less dense regime of the two prevails. Stronger attractive interactions would simply lead to a distribution function strictly increasing with packing density.

Topological contacts in various density regimes

The average number of topological contacts per chain occurring at various density intervals is shown in Fig. 6. The ordinate $\overline{N_c}(\rho)$ is calculated by considering all conformations ranging in a given density interval for chains of fixed length.

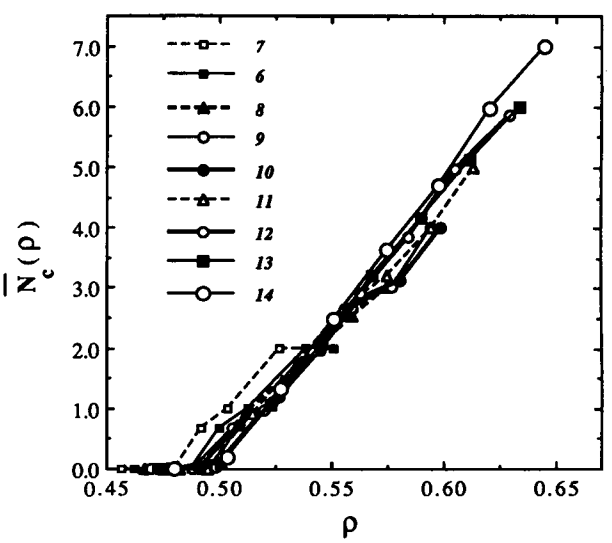

FIGURE 6 Increase in the average number $\overline{N_c}(\rho)$ of topological contacts in a given density regime with increasing density ρ. Results for various chain lengths, $6 \leq N + 1 \leq 14$, are almost indistinguishable, yielding the linear relationship between $\overline{N_c}(\rho)$ and ρ.

It is observed that results obtained for various chain lengths are almost indistinguishable. The resulting master curve indicates an approximately linear relationship between the number N_c of topological contacts and the packing densities $\rho\{\mathbf{r}\}$, provided that both are averaged over distinct density ranges. This analysis provides a direct validation of the use of the number of topological contacts for measuring packing density.

A more detailed analysis is achieved by classifying the topological contacts according to their order, i.e., the separation along the chain, between units participating in a given contact. Because of the particular geometry of the lattice, contacts between units separated by an even number of bonds are not possible. The lowest order contact that can occur is between units separated by three bonds. These are the only types of contacts observed at low densities. With a gradual increase in density, contacts of higher order come into play in a uniform way, as illustrated in Fig. 7 for the chain of 14 units. Here the abscissa indicates the seven distinct density intervals that were examined. The fraction of lowest order ($k = 3$) contacts diminishes as one moves into more compact regimes. Gradually, contacts of order $k = 5, 7, \ldots$ etc. appear, and their relative frequency also follows the same hierarchical order. The pie chart on the lower left corner of the figure shows the fraction of contacts of various orders occurring in the highest density regime. The probability distribution of contacts of various orders in the most compact state is further explored below.

The high probability of occurrence of lowest order topological contacts arises, in part, from the larger number of possible cases giving rise to them: in a chain of 14 residues as considered above, 11 pairs of residues may participate in contacts of order 3, while only a single pair, the terminal residues, are responsible for contacts of order 13. In this respect, the contributions of the more distant pairings are underestimated if their a priori probabilities of occurrence are disregarded. The above analysis gives information on the probability of contacts of various orders, in general, but does not give information on the relative participation of specific units in those contacts. Thus, although contacts of order 3 are highly probable, for example, not all of the units are expected to be involved with the same probability in those contacts. These features are most clearly analyzed with the help of contact maps.

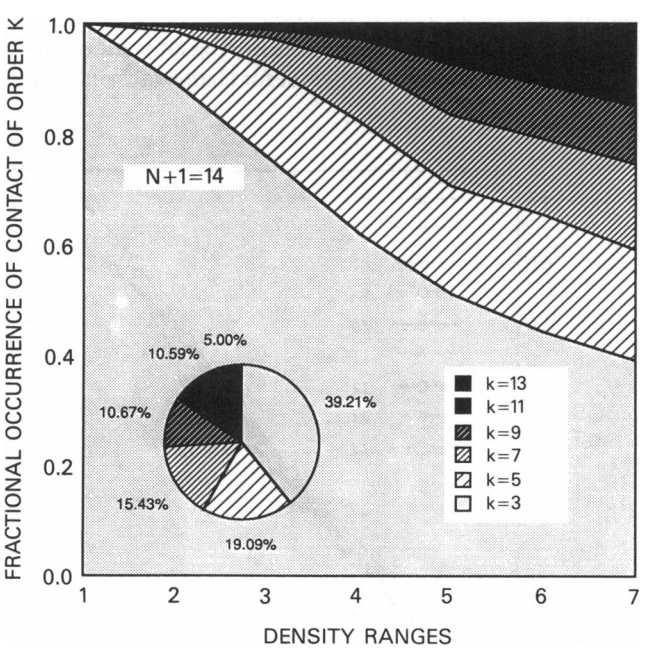

FIGURE 7 Fractional contribution of contacts of ranges of interaction in different density regimes for model chains of 14 units. Only odd-numbered orders are permitted because of the lattice geometry. The proportion of lower order contacts decreases with increasing density. The pie chart on the lower left indicates the fraction of contacts of various orders observed in the highest density regime.

Contact maps

We concentrated on the chain of 14 units because of its large number (110,188) of accessible distinct SAWs. This conformational space was considered large enough to yield statistically reliable results.

The two upper triangular maps in Fig. 8 display the change in free energy due to topological contacts, $\Delta G(i, j)$, resulting from (a) entropic and (b) entropic plus enthalpic effects. The reduced energy parameter ϵ^* is taken as 0 in (a). It is equal to -1.68 in (b). The residues with indices $1 \leq i \leq 11$ and $i + 3 \leq j \leq 14$ are shown in the ordinate and abscissa, respectively. These are the only units that may participate in topological contacts in a square planar lattice. Positions along the diagonal correspond to contacts of lowest order ($k = 3$). Contacts of higher order are located farther from the diagonal, their displacement increasing with their order. Darker regions in the maps indicate lower energy contacts. As mentioned above, contacts between units separated by an even number of bonds are not geometrically possible. For visual clarity, these contacts are smoothed by assigning mean values as the arithmetic average of the four neighboring points for inner ones with four neighbors, or two neighboring points along the edges for those involving either one of the terminal units. Alternative three- dimensional views of the same results are shown in Fig. 9, (a) and (b), which permit an easier visualization of the form of the free energy surface.

An examination of Figs. 8 and 9 reveals several important consequences of introducing intramolecular attractions. From the comparison of the two contact energy maps in parts (a) and (b) of Fig. 8, the dominance of contacts of order 3, along the diagonal, is found to be significantly attenuated. Contacts between terminal residues and any other residue in the chain now compete more efficiently with them. In particular, contacts between the two terminal residues are considerably enhanced. These are attributed to the joint effect of the high mobility or greater conformational freedom of the terminal units, supplemented by attractive interactions favoring compact forms. The new energy surface appears to have more structure, favoring specific contacts that were not so distinguishable in the absence of interaction energetics. However, a more important feature seen in the figures is the

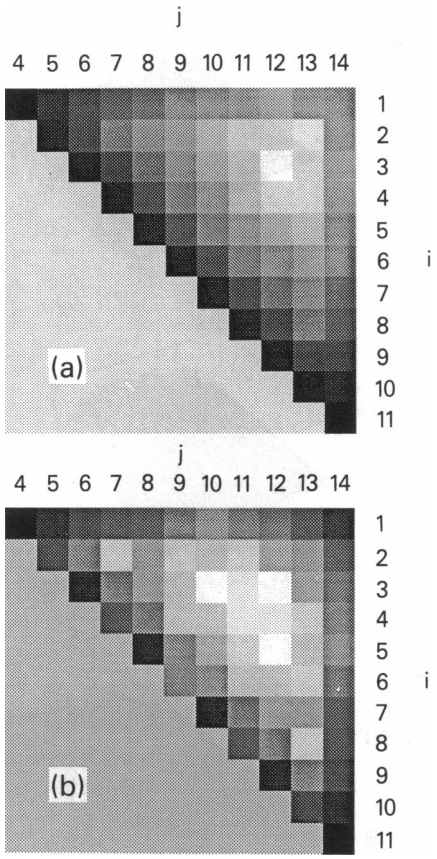

FIGURE 8 Maps of contact free energies $\Delta G(i, j)$ between residues $1 \leq i \leq 11$ and $i + 3 \leq j \leq 14$ for the complete set of 110,188 chains of 14 units subject to (*a*) entropic effects only and (*b*) entropic plus enthalpic effects with attractions of −1 kcal/mol between residues at 300 K. Residue indices are shown on both edges of the upper triangular maps. Darker regions indicate the lower energy contacts. Points along the diagonal represent contacts of lowest order between residues i and $i + 3$. Loci away from the diagonal correspond to more distant residues. The smooth distribution of energies in (*a*) becomes more specific in the presence of attractive interactions. An enhanced probability for contacts involving terminal residues is observed in (*b*). See legend to Fig. 9.

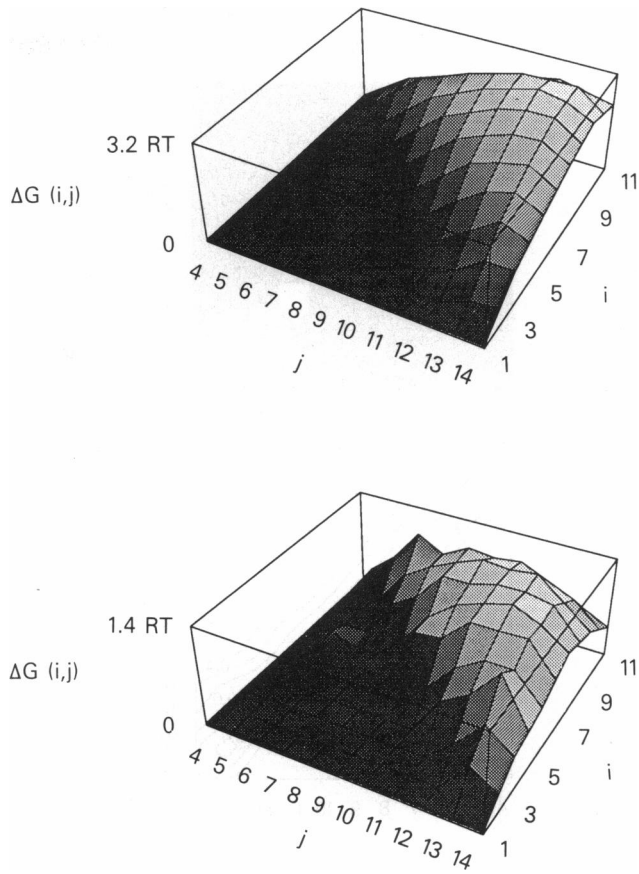

FIGURE 9 An alternative representation of the free energy surfaces $\Delta G(i, j)$ of Fig. 8 (*a*) and (*b*). A decrease by a factor of about 2 is observable in the height of the free energy surface upon introduction of attractive interactions between residues. The net effect is to reduce to a large extent the proportion of contacts between nearby residues (along the diagonal) and to enhance higher order contacts.

significant lowering of the free energy surface in (*b*) as compared with (*a*). This may be verified from the relative heights of the two surfaces in Fig. 9. The difference between the maximum and minimum values in the absence of intramolecular interactions is 3.1 *RT*. Inclusion of intramolecular attractions reduces this value to 1.3 *RT*. This decrease suggests that the homogeneous attractions between residues favor contacts, which are not quite favorable from the entropic point of view alone, at the expense of those occurring between nearest neighbors along the primary sequence. An analysis of this enthalpic contribution $\Delta H(i, j)$ to the free energy of contacts is carried out by taking the difference of the energy values in the two cases. The results are shown in Fig. 10. The upper part is the map of contact enthalpies and the lower part shows the corresponding isoenthalpic contours, ranging from −1.6 *RT* to 0.4 *RT* with 0.2 *RT* increments. The contacts of order 3 along the diagonal are subject to positive enthalpy values. It is clear that the enthalpic effects promote contacts between sequentially more distant residues, their impact being largest at the nearly maximally separated pair [2, 13].

Finally, from the comparison of the diagonals in the two maps of Fig. 8, we note that an even-odd type effect occurs in (*b*), favoring alternating contacts of order 3. For example, with an interaction energy, although the contact [1, 4] is highly probable, the next contact, between residues 2 and 5, has a considerably lower probability. The contact [1, 4] necessitates two consecutive bond angles of 90°. A given chain cannot simultaneously exhibit both [1, 4] and [2, 5] contacts due to square-planar lattice geometry. Inasmuch as the terminal residue has a larger degree of conformational freedom than the second, the probability of chains with the [1, 4] contact is larger and precludes the [2, 5] contacts. This explains the relative energies of the two highest points along the diagonal. On the basis of the decreased mobility of inner residues, one might expect to observe a lower probability of contact as subsequent points along the diagonal are considered. This is not the case. The contact [3, 6] is more favorable than [2, 5]. The origin of this behavior lies in the existence of strong correlations between contact pairs. In fact, contact [1, 4] is observed to be accompanied by [3, 6] in a large

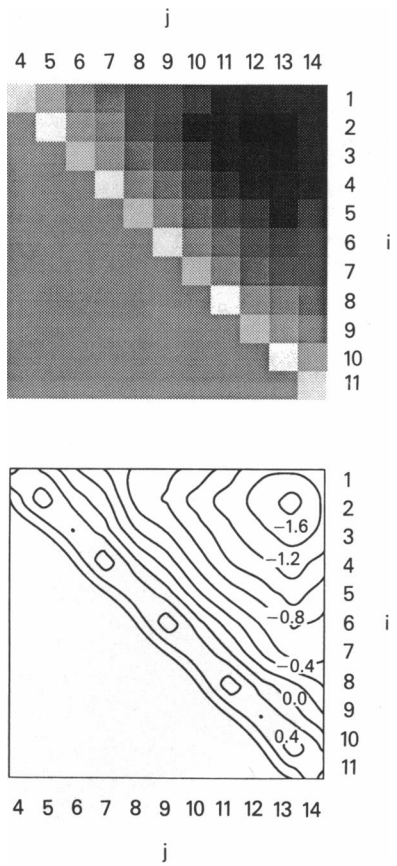

FIGURE 10 Enthalpy change $\Delta H(i, j)$ of contact $[i, j]$ due to attractive interactions between residues. Both contact map (*top*) and isoenergetic contours (*bottom*) are displayed for completeness. Contours are drawn at intervals of 0.2 *RT* ranging from −1.6 *RT* to 0.4 *RT*. Labels indicate reduced values, i.e., $\Delta H(i, j)/RT$. Homogeneous attractive interactions are found to lead to an unfavorable enthalpic effect on the association of nearby residues, while contacts between more distant pairs are favored.

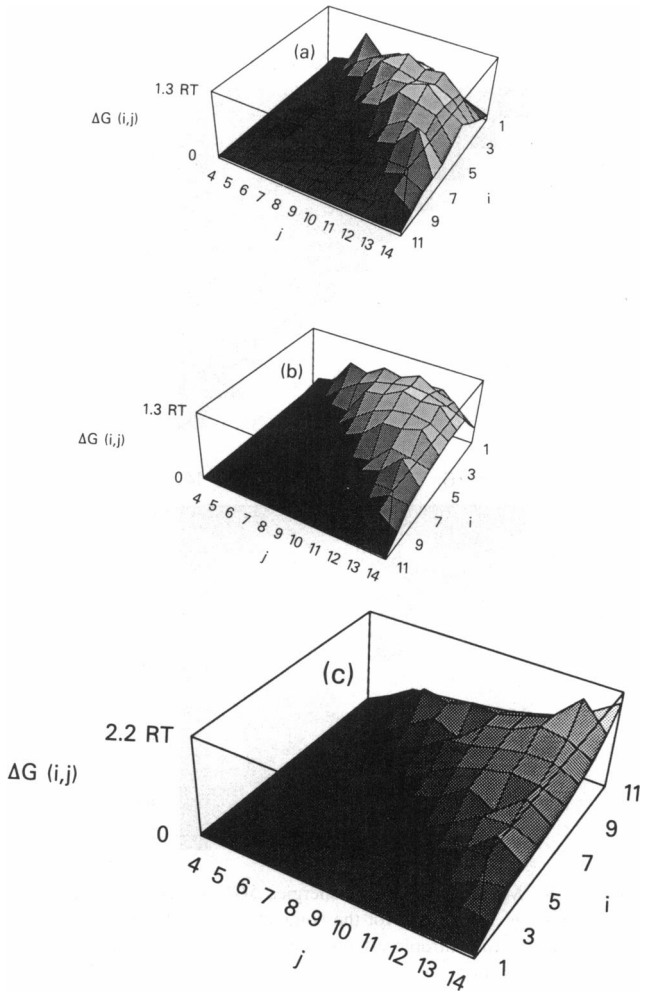

FIGURE 11 Comparison of the free energy surfaces $\Delta G(i, j)$ obtained in three highest density regimes for chains of 14 units subject to homogeneous attractive interactions of −1.0 kcal/mol between residues at 300 K. The respective densities are (*a*) 0.645, (*b*) 0.621, and (*c*) 0.598. See also legend to Fig. 12.

number of conformations. The simultaneous occurrence of those two contacts gives rise to a helical structure, one of the regular structures driven by compact state formation. Thus, *the introduction of even relatively mild attractive interactions, as presently performed, enhances the tendency of chain segments to be organized into regular patterns.* Correlations induced by specific attractive interactions are further explored in more detail in the accompanying paper.

We studied the similarities between topological contacts occurring in various density regimes. The three highest packing density intervals (ρ = 0.645, 0.621, and 0.598) were considered for N = 13. Fig. 11 displays the free energy surfaces corresponding to the cases (*a*) ρ = 0.645, (*b*) ρ = 0.621, and (*c*) ρ = 0.598, respectively. The close similarity between the two surfaces (*a*) and (*b*) is clearly apparent, whereas case (*c*) yields a free energy surface quite distinct both from qualitative and quantitative standpoints. Corresponding contact energy maps and contours are also presented as parts (*a*) to (*c*) of Fig. 12. Contours are drawn uniformly with 0.2 *RT* increments, starting from the lowest energy level (at [1,4]contact) up to 1.4 *RT* in each case, to facilitate a comparison of the results. The maximum energy (2.2 *RT*) occurs in part (*c*), at location [2, 13], as indicated. A closer examination of the surfaces in Fig. 12, (*a*) to (*c*), indicates that the two most compact structures preserve most of the same topological contacts, while the topology observed in the next slightly looser packing density regime exhibits substantial change. The most striking feature is the disappearance of higher order contacts and a more uniform frequency for the shortest range interactions, with a slight lowering of the chain density. It is noted that the average numbers of contacts per chain occurring in those domains are (*a*) 7, (*b*) 5.95, and (*c*) 4.71, as shown in Fig. 6. Thus, about one contact on the average is broken upon the decrease in the packing density as one moves from (*a*) to (*b*) and then (*c*). The passage from (*a*) to (*b*) does not bring about severe changes in topological contacts. However, a slight further expansion in chain dimensions from (*b*) to (*c*) leads to the disruption of the contacts between the most distant (in the primary sequence) residues. These contacts may, in fact, play a major role in stabilizing observed tertiary structure of the

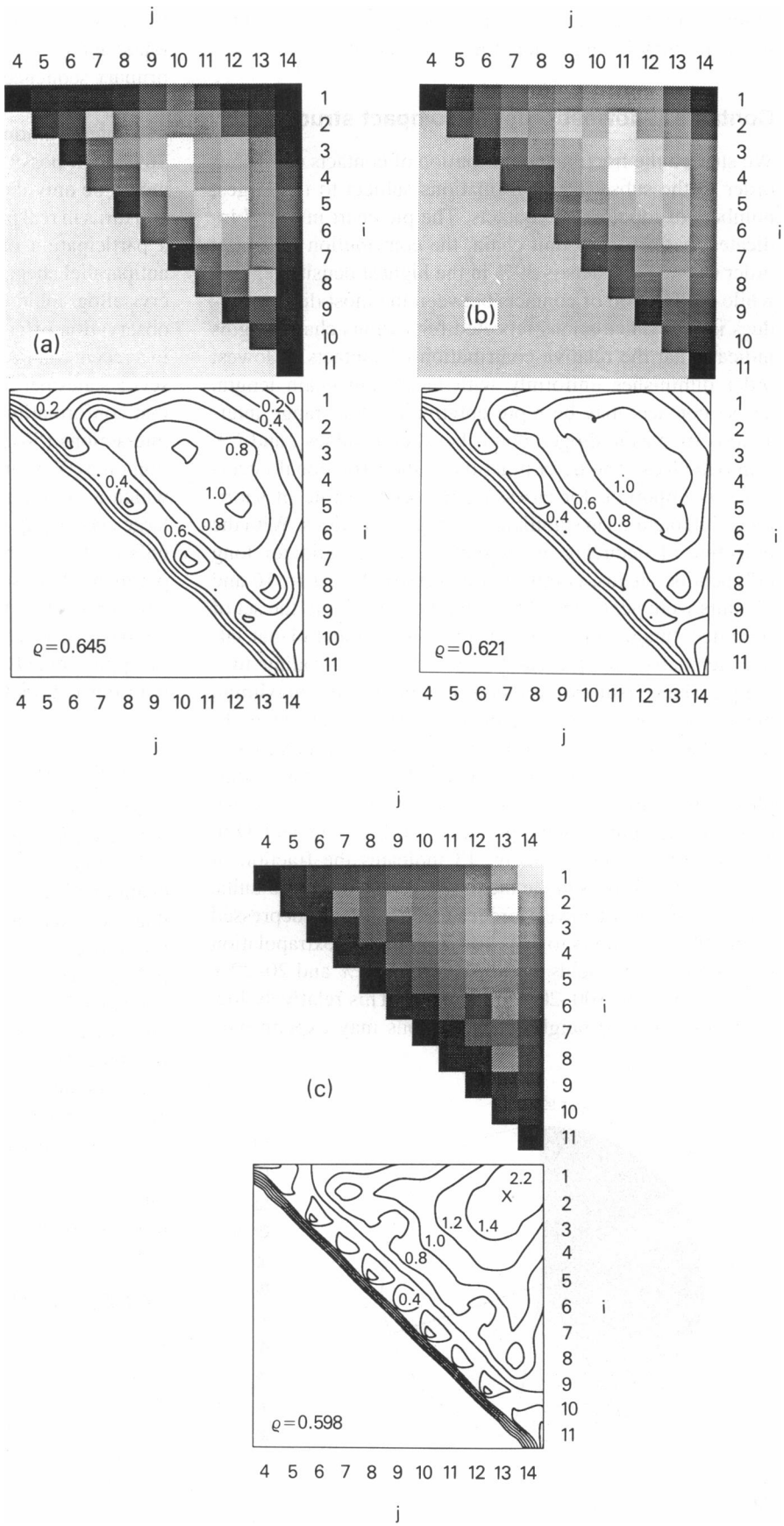

FIGURE 12 Contact maps and corresponding contour plots for the free energy surfaces $\Delta G(i, j)$ displayed in Fig. 11. The respective densities are (*a*) 0.645, (*b*) 0.621, and (*c*) 0.598. Contour labels are given in reduced units. (*a*) and (*b*) exhibit similar structural characteristics, whereas in (*c*) the probability of topological contacts between more distant residues is significantly decreased, as indicated by the high energy domain of 2.2 *RT* centered around the pair [2, 13].

globular proteins, and a major disruption of the folded molecule may be engendered near this point.

Contacts of lowest order in compact structures

We studied the fractional contribution of contacts of various order in the subset of conformations subject to the largest numbers of topological contacts. The pie chart in Fig. 7 indicates that, for a 14-unit chain, the contribution of lowest order ($k = 3$) contacts is 39% in the highest density regime, while the fraction of contacts between the most distant residues is 5%. Calculations repeated for various chain lengths indicate that the relative contribution of contacts of lowest order diminishes uniformly with increasing chain length. Those contacts may be representative of short-range interactions such as hydrogen bonds between residues i and $i + 3$ in α-helices. The determination of their fractional occurrence is important for assessing the relative role of short-range interactions in stabilizing compact structures. With the objective of deducing an asymptotic value valid for long polypeptide chains, compact structures of $N + 1 = 16$ and 25 units were generated. These numbers belong to the set of so-called "magic numbers" for two-dimensional planar lattice geometry and present the advantage of leading to a unique shape (square) for conformations with maximum numbers of contacts (Chan and Dill, 1989a). Thus, the problem reduces to the generation of all SAWs confined to 4×4 and 5×5 squares. 69 and 1081 SAWs result from a complete enumeration of all distinct conformations, for the two cases, in agreement with previous work (Chan and Dill, 1989a). The pie chart in Fig. 13 indicates the fraction of contacts of all orders in the most compact chains of 25 units. The probability of lowest order contacts is now depressed from 39% in 14 units to 30% for 25 units. An extrapolation to infinite chains yields a value of about 19% and 20–22% for proteins with 100–200 amino acids. This relatively low proportion of near neighbor interactions may explain why

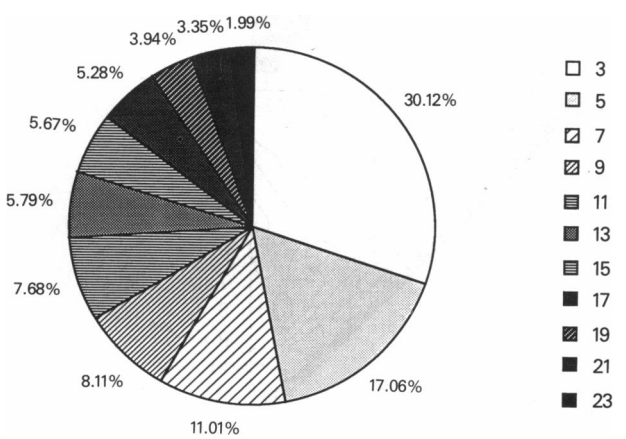

FIGURE 13 Distribution of topological contacts of various orders $3 \leq k \leq 23$ in the most compact state of a chain of 25 units. Local interactions between nearby units ($k = 3$) amount to 30% of all observed contacts in the set of 1081 distinct conformations. The proportion of higher order contacts decreases with their order.

theories based on intrinsic propensities alone (i.e., short range interdependence between closest residues along the primary sequence) are insufficient to account adequately for tertiary and secondary structures in proteins.

Contacts among near neighbors, which amount to about 20% of all possible contacts according to our extrapolation, would be only those among residues participating in helices and turns in real proteins. Kabsch and Sander (1984) reported a participation of 21.2% of residues in helices, 13.9% in antiparallel sheets and and 4.5% in parallel sheets, based on crystallographic observation of 62 globular proteins. This observation offers some support of our extrapolation. Also, in a recent rotational isomeric states statistical approach with seven states per residue, Rooman et al. (1992) demonstrated that conformational probability distributions based on local interactions (5–15 residues) yield some structural information on 44% of protein segments. Those are the segments in which a minimum energy conformation with a sufficiently large energy gap between other conformers exists, and about 68% of those minimum energy conformations match the experimentally observed folds of 69 proteins of well-resolved structures. Thus, the estimation of the correct tertiary fold on the basis of local interactions is possible with a probability of approximately 25%, which is in reasonable agreement with our extrapolation.

Influence of homogeneous ternary interactions

Contact free energy maps and corresponding contours were obtained (data not shown) by taking (*a*) $\epsilon_{ijk} = 0$ and (*b*) $\epsilon_{ijk} = -1.0$ kcal/mol at 300 K for the most compact state of a chain of 25 units. Inclusion of binary attractive interactions does not alter the maps, because all of the conformations in this subset are subject to the same number (i.e., 16) of topological contacts, and the relative statistical weights remain unchanged. The two cases exhibit only nearly indistinguishable differences: a small enhanced preference for contacts involving terminal residue(s) is discernible in case (*b*). The corresponding surface is somewhat more articulated than for case (*a*). Except for those two features, both maps indicate the same qualitative and quantitative characteristics. In this respect, we conclude that nonspecific homogeneous ternary interactions may have a relatively negligible effect on the topology of the compact structures themselves.

CONCLUSIONS

It is widely recognized that for a realistic understanding of protein stability and folding, the heterogeneity and specificity of intramolecular interactions need be considered. This is an essential requirement prescribed by the uniqueness of the most compact, i.e., native state. Likewise, a large portion of conformations occurring in the less dense, compact, but denatured regime might be biased by the same type of heterogeneous and specific interactions. Those effects were not considered in the present work. A general perspective was considered, with the objective of searching for the role of

homogeneous interactions possibly stabilizing intermediate density states and, having established their occurrence under suitable attractive interactions properly counterbalancing the adverse entropic effects, investigating the topological similarities between the most compact and less tightly packed stable structures.

The analysis invites attention to one or more intermediate density structures stabilized by attractive interactions of about 0.5–1.0 kcal/mol between residues. This is indicated by the distinct peaks appearing in the probability distribution functions for packing densities shown in Figs. 3–5. In general, the highest peaks are not located at the very highest density point but are shifted slightly toward less tightly packed regimes. The next-to-highest density state is preferred for the longer chains in the present results. These observations raise the question of whether the native state, or the most probable state, corresponds to the highest density form displayed in the figures or to somewhat lower density states, which exhibit relatively low free energy. Results here and our previous results (Covell and Jernigan, 1990) and those of Shakhnovich et al. (1991) suggest that the identification of native states with the most compact states or minimum energy states may not strictly hold. In fact, if kinetic accessibility is a factor affecting the type of equilibrium folded structures observed in proteins, compact but somewhat less dense regimes may well be a better starting point for investigating the protein folding-unfolding problem.

The occurrence of multiple peaks in the density distribution functions brings into consideration another question of whether it possible to identify the energetically favorable intermediate density states with the molten globule state. The relative frequency of those intermediate density states rests on a delicate balance between attractive interactions favoring compact forms and the entropic effect enhancing less dense states with higher conformational freedom. The present study suggests that the existence of stable compact but denatured states in globular proteins with packing density lower than the native state may be accounted for by a simple model of SAW chains with a homogeneous intramolecular attractive potential. The attractive interactions between residues may be viewed as an indirect consequence of the hydrophobic effect, inherently favoring condensation of the protein structure. In this respect, the present work brings into consideration the possible nonspecific role of hydrophobic interactions in stabilizing the molten globule state.

A comparison of the structural characteristics of model chains, with and without interactions between residues, indicates that the homogeneous attractive interactions between units have an adverse effect on the association of nearby units, while contacts between more distant residues along the primary sequence are enhanced. This interesting feature emerges from the contact enthalpy map displayed in Fig. 10. Another interesting observation was the close resemblance of the two most compact regimes, which distinctly disappeared upon a small decrease in packing density. The major structural difference arising in the lower density range was the substantial elimination of the topological contacts between residues located farther apart along the primary sequence. These types of disruptions may be critically important in inducing the complete denaturation of the molecule. Such a result is qualitatively consistent with a picture in which the range of interactions increases upon folding, if the density increases directly at every folding step. However, the present results are not consistent with a view (Miyazawa and Jernigan, 1982; Moult and Unger, 1991) that conformational characteristics (probably contact pairs) observed at states of lower density persist to states of higher densities. As a final remark, the present analysis indicates that contacts among near neighbors along the primary sequence amount to approximately 20% of all possible contacts in compact structures in typical proteins of about 100–200 amino acids. This relatively low proportion suggests that intrinsic propensities are not the major factor determining the tertiary structure of globular proteins.

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

1. Alonso, D. O. V., K. A. Dill, and D. Stitger. 1991. The three states of globular proteins: acid denaturation. *Biopolymers.* 31:1631–1649.
2. Barber, M. N., and B. W. Ninham. 1970. Random and Restricted Walks: Theory and Applications. Gordon & Breach, New York. 176 pp.
3. Ben-Naim, A. 1989. Solvent-induced interactions: hydrophobic and hydrophilic phenomena. *J. Chem. Phys.* 90:7412–7425.
4. Ben-Naim, A., K. Ting, and R. L. Jernigan. 1989a. Solvation thermodynamics of biopolymers. I. Separation of the volume and surface interactions with estimates for proteins. *Biopolymers.* 28:1309–1325.
5. Ben-Naim, A., K. Ting, and R. L. Jernigan. 1989b. Solvation thermodynamics of biopolymers. II. Correlations between functional groups. *Biopolymers.* 28:1327–1337.
6. Brooks, C. H. III, M. Karplus, and B. M. Pettitt. 1988. Proteins: a theoretical perspective of dynamics, structure and thermodynamics. *Adv. Chem. Phys.* 1–259.
7. Chan, H. S., and K. A. Dill. 1989a. Compact polymers. *Macromolecules.* 22:4559–4573.
8. Chan, H. S., and K. A. Dill. 1989b. Intra-chain loops in polymers. Effects of excluded volume. *J. Chem. Phys.* 90:492–509.
9. 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.
10. Chan, H. S., and K. A. Dill. 1990b. Origins of structures in globular proteins. *Proc. Natl. Acad. Sci. USA.* 87:6388–6392.
11. Covell, D. G., and R. L. Jernigan. 1990. Conformations of folded proteins in restricted spaces. *Biochemistry.* 29:3287–3294.
12. Dill, K. A. 1990. Dominant forces in protein folding. *Biochemistry.* 29:7133–7155.
13. Dill, K. A., and D. Shortle. 1991. Denatured states of proteins. *Annu. Rev. Biochem.* 60:795–825.
14. Dolgikh, D. A., R. I. Gilmanshin, E. V. Brazhnikov, V. E. Bychkova, G. V. Semisotnov, S. Y. Venyaminov, and O. B. Ptitsyn. 1981. α-lactalbumin. Compact states with fluctuating tertiary structure? *FEBS Lett.* 136:311–315.
15. Dolgikh, D. A., A. P. Kolomiets, I. A. Bolotina, and O. B. Ptitsyn. 1984. "Molten-globule" state accumulates in carbonic anhydrase folding. *FEBS Lett.* 165:88–92.
16. Finkelstein, A. V., and E. I. Shakhnovich. 1989. Theory of cooperative transitions in proteins. II. *Biopolymers.* 28:1681–1694.
17. Flory, P. J. 1969. Statistical Mechanics of Chain Molecules. Wiley Interscience, New York. 432 pp.

18. 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.

19. 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.

20. 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.

21. Hinds, D. A., and M. Levitt. 1992. A lattice model for protein structure prediction at low resolution. *Proc. Natl. Acad. Sci. USA.* 89: 2536–2540.

22. Ikeguchi, M., K. Kuwajima, M. Mitani, and S. Sugai. 1986. Evidence for identity between the equilibrium unfolding intermediate and a transient folding intermediate: a comparative study of the folding reactions of α-lactalbumin and lysozyme. *Biochemistry.* 25:6965–6972.

23. Jaenicke, R. 1987. Folding and association of proteins. *Prog. Biophys. Mol. Biol.* 49:117–237.

24. Jernigan, R. L. 1992. Protein folds. *Curr. Opin. Struct. Biol.* 2:248–256.

25. Kabsch, W., and C. Sander. 1984. On the use of sequence homologies to predict protein structure: identical pentapeptides can have completely different conformations. *Proc. Natl. Acad. Sci. USA.* 81:1075–1078.

26. Kauzmann, W. 1959. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* 14:1–57.

27. Kim, P. S., and R. L. Baldwin. 1982. Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu. Rev. Biochem.* 51:459–489.

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

29. Kuwajima, K., K. Nitta, M. Yoneyama, and S. Sugai. 1976. Three state denaturation of α-lactalbumin by guanidine hydrochloride. *J. Mol. Biol.* 106:359–373.

30. Lee, B., and F. M. Richards. 1971. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* 55:379–400.

31. Liljas, A., and M. G. Rossmann. 1974. X-ray studies of protein interactions. *Annu. Rev. Biochem.* 43:475–507.

32. Miyazawa, S., and R. L. Jernigan. 1982. Most probable intermediates in protein folding-unfolding with a non-interacting globule-coil model. *Biochemistry.* 21:5203–5213.

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

34. Moult, J., and R. Unger. 1991. An analysis of protein folding pathways. *Biochemistry.* 30:3816–3824.

35. Oghushi, M., and A. Wada. 1983. "Molten globule state": a compact form of globular proteins with mobile side-chains. *FEBS Lett.* 164: 21–24.

36. Privalov, P. L. 1979. Stability of proteins. Small globular proteins. *Adv. Protein Chem.* 33:167–241.

37. Ptitsyn, O. B. 1987. Protein folding: hypothesis and experiments. *J. Protein Chem.* 6:272–293.

38. Ptitsyn, O. B., and G. V. Semisetnov. 1991. The mechanism of protein folding. *In* Conformations and Forces in Protein Folding. B. T. Nall and K. A. Dill, editors. American Association for the Advancement of Science, Washington, DC. 155–168.

39. Richards, F. M. 1977. Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* 6:151–176.

40. Rooman, M. J., J.-P. A. Kosher, and S. J. Wodak. 1992. Extracting information on folding from the amino acid sequence: accurate predictions for protein regions with preferred conformation in the absence of tertiary interactions. *Biochemistry.* 31:10226–10238.

41. 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.

42. Shakhnovich, E. I., and A. V. Finkelstein. 1989. Theory of cooperative transitions in proteins. I. *Biopolymers.* 28:1667–1680.

43. 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. Peptide Protein Res.* 7:445–459.

44. Volkenstein, M. 1963. Configurational Statistics of Polymer Chains. Wiley Interscience, New York. 562 pp.

45. Wada, A., and H. Nakamura. 1981. Nature of charge distribution in proteins. *Nature* 293:757–758.

46. Yamakawa, H. 1967. Modern Theory of Polymer Solutions. Harper & Row, New York. 419 pp.