Modeling Compact Denatured States of Proteins†

Eaton E. Lattman

Department of Biophysics and Biophysical Chemistry, Johns Hopkins Medical School, Baltimore, Maryland 21205-2185

Klaus M. Fiebig and Ken A. Dill*

Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-1204

Received September 29, 1993; Revised Manuscript Received March 7, 1994®

ABSTRACT: We propose a model for the conformations of compact denatured states of globular proteins: that they are broad ensembles of chain backbone conformations that involve common localized hydrophobic clustering and helical contacts, depending on the amino acid sequence. We construct representative ensembles for chain lengths up to 136 monomers on three-dimensional cubic lattices using the "hydrophobic zippers" method (Fiebig & Dill, 1993). We find that model conformations with radii of gyration about 20% larger than native conformations commonly have bimodal distributions of $P(r)$, of the pairwise interatomic distances, r, and Kratky plots in agreement with recent small-angle X-ray scattering (Sosnick & Trewhella, 1992; Flanagan et al., 1992; Kataoka et al., 1993; Flanagan et al., 1993) experiments on three different proteins. We also find that the lattice model of the Shortle 1–136 fragment of staphylococcal nuclease does not appear capable of forming a single hydrophobic core by hydrophobic zippering, consistent with experiments.

What are the compact denatured conformations of globular proteins? Although compact denatured states have extensive conformational diversity, several properties are well-known (Ptitsyn, 1991; Ptitsyn & Semisotnov, 1991; Kuwajima, 1989; Dill & Shortle, 1991). They are up to 10–20% larger in radius than native molecules. They have secondary structure, often native-like (Baum et al., 1989; Roder & Wüthrich, 1986), but little native-like tertiary structure (Jeng et al., 1990). And they bind the fluorescent probe ANS (Ptitsyn, 1991), which is considered to indicate some hydrophobic clustering. Hydrogen-exchange NMR experiments show that some protons are partly protected from solvent (Baum et al., 1989; Hughson et al., 1990).

Recent small-angle X-ray scattering (SAXS) experiments (Sosnick & Trewhella, 1992; Flanagan et al., 1992; Flanagan et al., 1993) show an additional property of compact denatured states, namely a *bimodal* distribution $P(r)$ of the pairwise distances r between atoms. This behavior is observed in the Shortle 1–136 fragment (Shortle & Meeker, 1989) of staphylococcal nuclease and thermally denatured ribonuclease A (see Figure 1). Since *unimodal* $P(r)$ curves are characteristic of both native (Sosnick & Trewhella, 1992; Flanagan et al., 1992) and unfolded states (Sosnick & Trewhella, 1992) (generally obtained in strong denaturants), the SAXS data provide important new evidence bearing on how these compact denatured states differ from native and unfolded states. In particular, a $P(r)$ distribution with a bimodal or shoulder-like shape indicates some clustering of amino acids.

What chain conformations could lead to these properties? One model, referred to as the "molten globule" model (Ptitsyn & Semisotnov, 1991; Shakhnovich & Finkelstein, 1989; Karplus & Shakhnovich, 1992), proposes that the chain backbone has the conformation of the native protein and that the disorder in the compact denatured state arises only within the side chains. Here we use the term "molten globule" to refer specifically to that model, although other authors use the term molten globule more broadly. But it seems unlikely that the molten globule model could account for bimodal SAXS $P(r)$ curves, because that model does not appear to involve sufficiently large perturbations from the native structure to be consistent with the clustering of scattering centers.

We develop here a different model of highly compact denatured states. We propose that highly compact denatured states are broad ensembles of different backbone conformations, with considerable hydrophobic clustering. Based on this same idea, an earlier mean-field model was used to predict aspects of the salt/pH phase diagram of apomyoglobin (Stigter et al., 1991; Alonso et al., 1991), in reasonable agreement with the experiments of Goto and Fink (1990). In the present work, we generate compact chain conformations for sequences of amino acid modeled as hydrophobic (H) and polar (P) monomers (Lau & Dill, 1989, 1990; Chan & Dill, 1991) on the three-dimensional cubic lattice, using the hydrophobic zippers method (Fiebig & Dill, 1993; Dill et al., 1993). We find this model to be consistent with properties of the compact denatured states, including the SAXS clustering data.

THE MODEL

We model proteins as specific sequences of H and P monomers (Lau & Dill, 1989; Chan & Dill, 1991; Shortle et al., 1992) on the three-dimensional simple cubic lattice. Each chain configuration is a self-avoiding walk on the lattice. Each amino acid can only occupy one lattice site, and no two amino acids may reside on the same lattice site. To represent sequences of the proteins modeled here—crambin, cytochrome *c* (apo form), ribonuclease A (with no disulfides), and the staphylococcal nuclease 1–136 fragments (Shortle & Meeker, 1989)—we define "H" (hydrophobic) amino acids to be Ala, Ile, Leu, Met, Phe, Trp, Tyr, Val, and Cys and "P" (polar) residues as all others. The resulting HP sequences are shown in Table 1. Our aim here is not to attempt microscopically accurate predictions of the structures of these proteins. The lattice model is too crude to represent bond angles and interactions realistically and too approximate to represent native structures adequately. Rather, our aim in comparing these sequences is to learn how the $P(r)$ behavior depends on the monomer sequence, the compactness, and the extent of hydrophobic clustering.

Conformations are constructed by the "hydrophobic zippers" strategy (Fiebig & Dill, 1993; Dill et al., 1993), a confor-

† We thank the NIH for support.
® Abstract published in *Advance ACS Abstracts,* May 1, 1994.

0006-2960/94/0433-6158$04.50/0

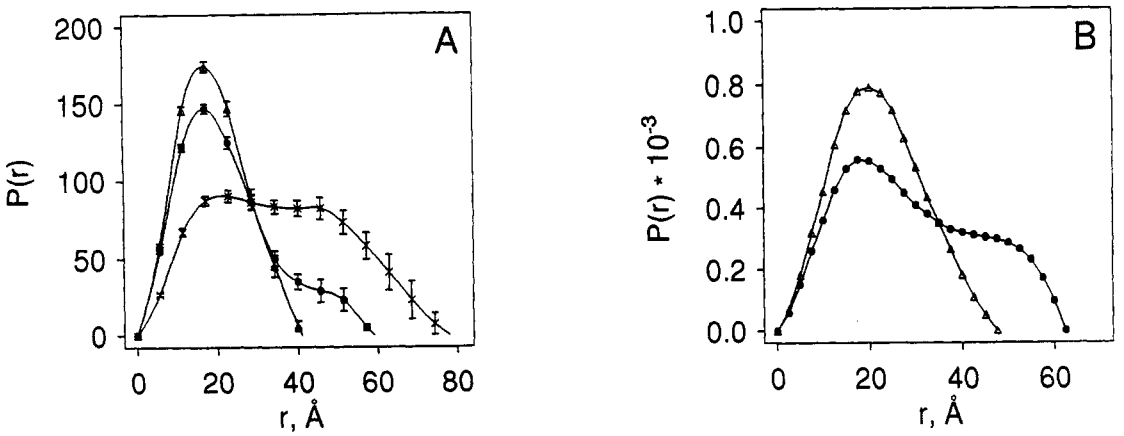

FIGURE 1: Bimodal $P(r)$ curves observed for ribonuclease A (panel A, from Figure 3 of Sosnick & Trewhella (1992)) and the 1–136 fragment of staphylococcal nuclease (panel B, Figure 4 of Flanagan et al. (1992)). The ribonuclease A data was measured under reducing conditions at temperatures 45 °C (△), 51 °C (●), and 57 °C (×). $P(r)$ curves for the nuclease fragment (panel B) were obtained in the presence (△) and absence (●) of Ca^{2+} and inhibitor pdTp (3′,5′-biphospho-2′-deoxythymidine).

Table 1: HP Sequences of the Four Model Proteins

Crambin, 1crn (46 Residues): PPHHHPHHHPPPHPHHPHHPPHPHHHHPHPPHHHHHPHPHHPPHHP
Cytochrome *c* (apo form), 3cyt (103 Residues): PPHHPPPPPHHPPHHPHPPHPPPPPPPHPPPHHPHPHPPPPPPHPPHPHPPHPPPPPHHH- PPPPHHPHHPPPPPHHPPPPHHHHPHPPPPPPPPHHHHHPPHP
Ribonuclease A, 7rsa (124 Residues): PPPHHHPHPPPPHPPPPPHHPPPPHHPPPHHPPPPHPPPPHPPHPPHHPPPHHPHPHHHP- PPPHHHPPPPPPHHPPHPPHPHPPHPPPPPPPHPPHHHPPPPHPPPHHHHHPPPPHHPH- PHPHPH
1–136 Staphylococcal nuclease fragment, 2sns (1-136): HPPPPPHPPPPHPHHPHHPPPPHPHHHPPPPHPHPHHHHPPPPPPPPPPHPPHPPPHP- HHPPPHHPPHPPHPHPHPPPPPPPPHPPPHHHHHHPPPHHPPHHHPPPHHPHHHHHP- PPPPPPPPHPPPPHPHPPPP

mational construction procedure that generates diverse conformations of any specified number of HH contacts, up to the maximum or near-maximum number. It has been proposed that hydrophobic zippers may describe the folding pathways of globular proteins (Dill et al., 1993). The hydrophobic zippers strategy is an algorithm that sequentially assembles hydrophobic contacts, leading to a compact chain conformation with one or several hydrophobic cores. The algorithm leads to an ensemble of different conformations of different numbers of hydrophobic contacts. A few of these conformations have maximal or near maximal numbers of hydrophobic contacts and a single compact hydrophobic core. These are native or near-native conformations of the model sequence. However, the greater preponderance of conformations have fewer contacts than this. The zippers process terminates at "end states", most of which are trapped configurations which are unable to find additional viable hydrophobic contacts. At this point, the chain has reached an "entropy catastrophe", since the finding of an additional HH contact is either impossible or it involves an extensive conformational search. Alternately the hydrophobic zippers process can be terminated at a given number h of HH contacts to yield intermediate conformations. For model crambin, which is short, we use zippers to generate conformations over the full range of compactness, and for larger model proteins (ribonuclease A, cytochrome *c*, staphylococcal nuclease 1–136 fragment), zippers lead to end states that have many of the properties of compact denatured states.

To generate representative ensembles of conformations with the hydrophobic zippers procedure we proceed via two steps.

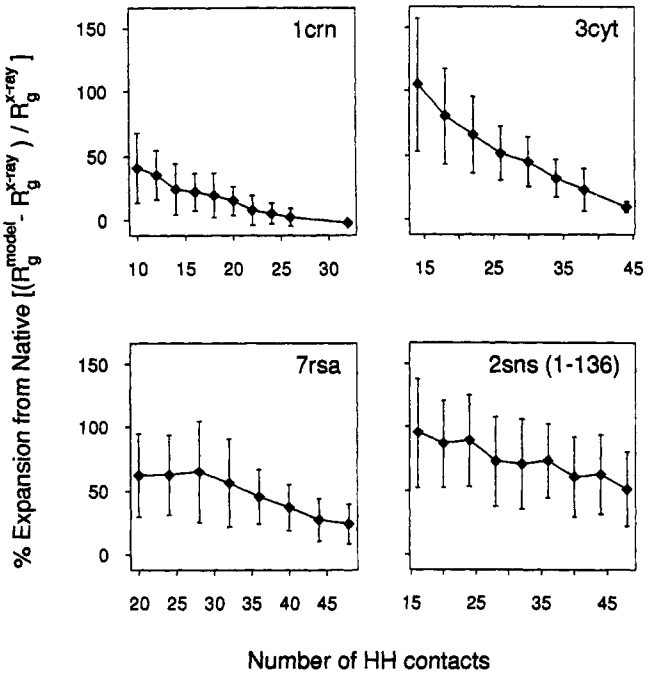

FIGURE 2: Radii of gyration of hydrophobic zipper intermediate and endstate conformations: percent expansion relative to the native structure from PDB C_α coordinates vs number of HH contacts for crambin (1crn), cytochrome *c* (3cyt), ribonuclease A (7rsa), and the 1–136 fragment of staphylococcal nuclease (2sns). For cytochrome *c* and ribonuclease A the compact states reached by hydrophobic zippers are about 10–20% larger than native. For the staphylococcal nuclease fragment, R_g is larger. In each panel the data at the largest h is obtained from an ensemble of conformations which are predominantly hydrophobic zipper end states. Ensembles with larger h values were not generated due to their prohibitive computational cost. All other data points are derived from hydrophobic zipper intermediate conformations. Each point on the figure represents an average over 400 conformations constructed from 20 contact maps (20 conformations from each contact map). Error bars indicate one standard deviation.

First a contact map (matrix of HH contacts) is generated with the hydrophobic zippers procedure (Fiebig & Dill, 1993), one contact at a time. The procedure chooses contacts randomly among those that are "T-local" at the given stage of partial construction (Fiebig & Dill, 1993). A "T-local" contact is defined by a graph-theoretic procedure (Fiebig & Dill, 1993; Dill et al., 1993), but physically it approximately corresponds to any contact whose formation causes the least loss of conformational entropy relative to the already partially constructed structure. We set in advance a target h value.

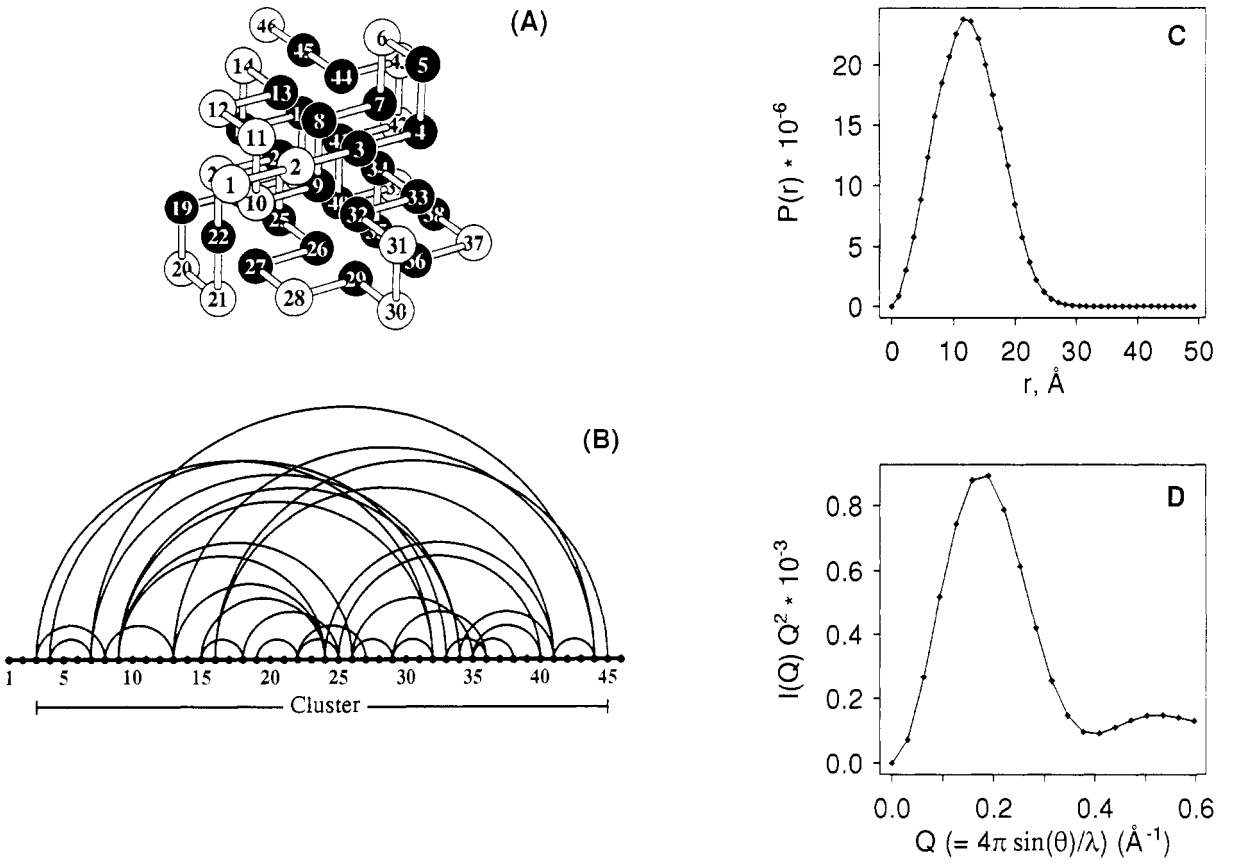

FIGURE 3: Model crambin (1crn) conformation (A) of high compactness (h = 32). The $P(r)$ curve (C) and Kratky plot (D) show unimodal behavior. On the "polymer graph", B, the straight line represents the covalently linked monomers. The curved links above the line represent noncovalent spatial contacts among monomers. For example, in configuration A, monomers 4 and 7 are neighbors, as are 13 and 44, etc. The single hydrophobic cluster of this conformation is indicated by the multiple curved links connecting residues 3–45.

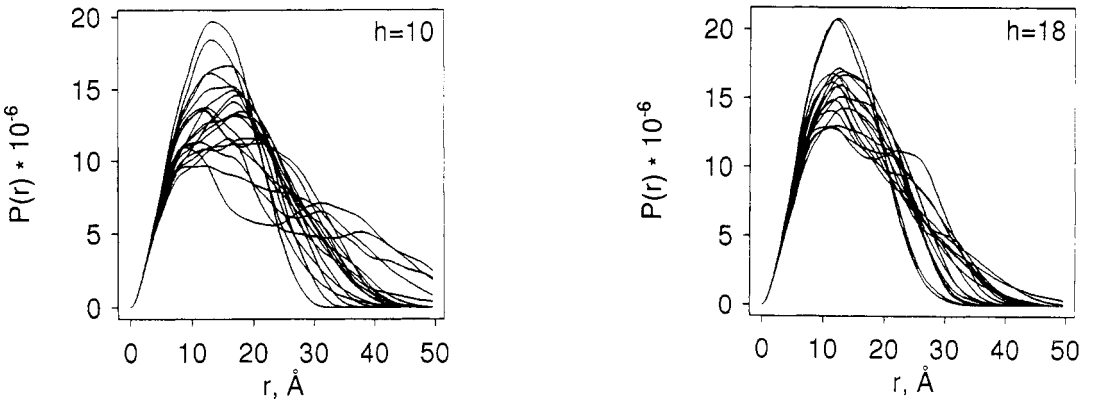

FIGURE 4: Representative $P(r)$ curves of model crambin (1crn) conformations with h = 10 and h = 18 HH contacts; h = 32 is shown in Figure 3.

HH contacts are then sequentially assembled until either h HH contacts have been formed, or until it is no longer possible for the algorithm to find additional HH contacts. In the latter case the resulting contact map is discarded. Thus an ensemble of contact maps is generated where each contact map has the given number, h, of HH contacts. Some conformations are end states (if the algorithm could not have proceeded further), and some are not. Finding conformations becomes computationally difficult at large h, and a large fraction of such conformations are end states.

The second step involves using the contact maps generated in step 1 to construct conformations on the lattice. For each contact map we construct an ensemble of about 20 conformations which satisfy all contact constraints of the given contact map. The algorithm used is a standard "backtracking" algorithm (Reingold et al., 1977) which was modified to (a) include HH contact constraints and (b) ensure that successively generated conformations are not correlated. The procedure checks contact constraints whenever a monomer is placed on the lattice. If placement violates any contact constraints, then the algorithm must backtrack. The lack of correlation between generated conformations was verified by a conformational distance measure described below.

CALCULATION OF $P(r)$ CURVES

From the conformations assembled in this way, we then compute the distributions of $P(r)$, the probabilities that amino acid centers are separated by a distance r, using the relation

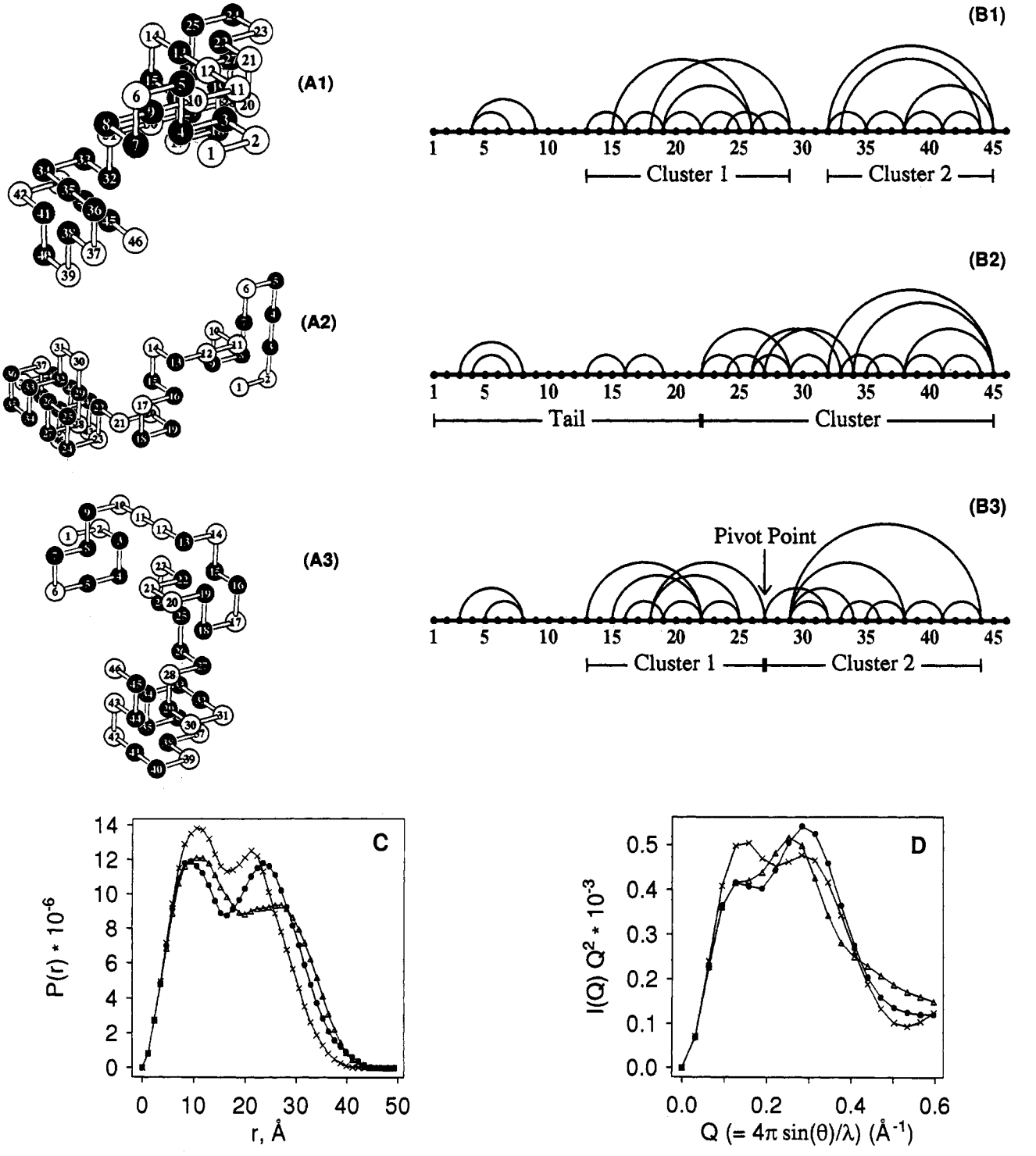

FIGURE 5: Examples of model crambin conformations (A) that give rise to bimodal $P(r)$ curves (C). The corresponding polymer graphs and Kratky plots are shown in B and D, respectively, conformation 1 (●), 2 (△), 3 (×).

$$P(r) = \frac{1}{2\pi^2}\int_0^{Q_{max}} I(Q)Qr \sin(Qr)\, dQ \qquad (1)$$

$I(Q)$ is the diffracted intensity at a distance $2\pi/Q$ from the origin of reciprocal space. The scattering vector Q is equal to $4\pi \sin(\theta)/\lambda$, where θ is the Bragg angle and λ is the wavelength of the radiation. The integral was evaluated as a discrete sine transform using a fast Fourier transform program. Integration was carried out to a maximum scattering vector Q_{max} of $\pi/2$ Å^{-1}. $I(Q)$ is calculated from the coordinates of the lattice model using the program CALCROT (Lattman, 1989), in which the intensity is expanded in spherical harmonics. In order to establish a reasonable unit of distance, necessary for comparison with the SAXS experiments, we set the spacing between lattice sites to be 5 Å. This spacing gives approximately the correct volume of the compact chain compared to the X-ray crystal structure. The value of this quantity is not critical. We used a scattering factor of the form

$$f(Q) = 1 + \cos\left(\frac{\pi Q}{Q_{max}}\right), \quad 0 \le Q \le Q_{max} \qquad (2)$$

for all lattice monomers, where Q_{max} is the largest scattering vector in reciprocal space for which $I(Q)$ is being evaluated. This representation of $f(Q)$ smoothly approaches zero at Q_{max},

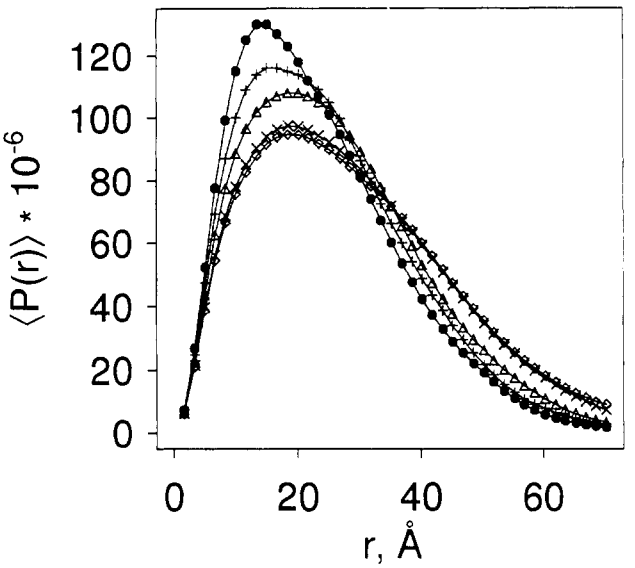

FIGURE 6: Model staphylococcal nuclease fragment (1–136): Averages ⟨ ⟩ of $P(r)$ are over 80 different conformations with a given h = 48 (●), 40 (+), 32 (△), 24 (×), 16 (◇).

which eliminates artifactual ripples in $P(r)$. Principal axis directions and the moments of inertia about them were calculated as described by Goldstein (1980).

RESULTS

The following properties indicate that for model ribonuclease A, cytochrome *c*, and the staphylococcal nuclease 1–136 fragment, the predominant end states of hydrophobic zipper processes resemble compact denatured states. For the small protein crambin we study the full range of compactness.

(*1*) *Radii Slightly Larger than Native.* We find that the predominant end states of hydrophobic zippers have radii of gyration, R_g, only slightly greater (about 10–20%) than the maximally compact states (see Figure 2). The only exception was our simulation of the staphylococcal nuclease 1–136 fragment, for which the end state radii were about 50% larger than the maximally compact state. In contrast, as seen in Figure 3, model crambin can reach very compact conformations (h = 32). Radii of gyration of the PDB structures were calculated using the C_α coordinates. Radii of model conformations were calculated assuming a lattice spacing of 5 Å, such that the most compact model crambin (h = 32) conformations have a radius of gyration about equal to that of the X-ray structure. Because this lattice spacing is approximate, the radii of the model chains are only approximate.

(*2*) *Bimodal $P(r)$ Curves.* We find that over a narrow range of intermediate chain compactness, $P(r)$ curves of individual conformations commonly show the same bimodal behavior as is observed in the SAXS experiments. For our crambin-like sequence, the maximum number of HH contacts we found was h = 32. For all the conformations with 32 HH contacts, the $P(r)$ curves resemble those of the native protein SAXS data insofar as they are very similar to each other, are all unimodal, and their radii of gyration range only from 10 to 12 Å, which is the native-like radius for the lattice model of this protein (see Figure 3). For the looser hydrophobic zipper intermediate conformations, with h = 20 HH contacts, the $P(r)$ curves were usually broader. Some curves were bimodal. For conformations with h = 18 HH contacts, about half of the $P(r)$ curves were bimodal, having either a pronounced shoulder or two peaks. These may therefore be representative of the compact denatured states studied in the SAXS experiments. For conformations even less constrained, h = 10 HH contacts, the $P(r)$ curves vary widely, and few give bimodal curves that resemble the experimental data. Our open conformations look about as disordered as those of the model of Calmettes et al. (1993) of phosphoglycerate kinase in strong denaturants. Thus bimodal distributions are not observed if a lattice chain conformation is too compact or too open, but only if it falls within a narrow range (for crambin h = 16–20) of intermediate compactness. Representative $P(r)$ curves for h = 10 and h = 18 HH contacts are shown in Figure 4.

The prediction of bimodal behavior is not proof that the model is correct, but is only an indication that it is consistent with experiments. The scattering function cannot be directly inverted to give a unique model. Other models can also fit the data. For example, we found that the data are also consistent with (i) a simple dumbbell model of two nonoverlapping spheres or (ii) a model based on scattering centers located at the vertices of a regular triangle or tetrahedron.

What other configurational properties predict whether a $P(r)$ distribution will be bimodal or unimodal? We tested several conformational properties and found no predictive power from the radius of gyration, principal axes of rotation, or moments of inertia. We also calculated matrices of distances

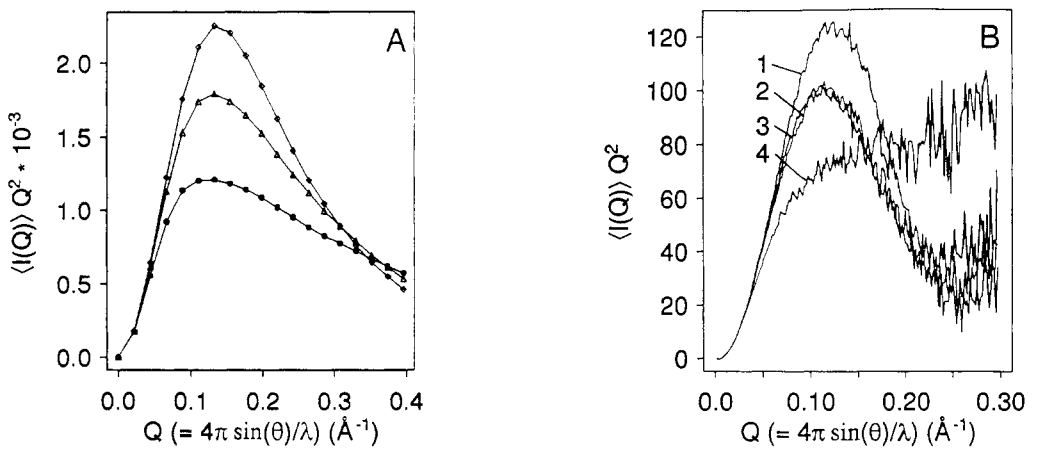

FIGURE 7: Cytochrome *c*, three states: (A) model and (B) experiments shown in Figure 1 of Kataoka et al. (1993). Kratky plots of the unfolded state (h = 2 (●)), the compact denatured state (h = 26 (△)), and a near native state (h = 38 (◇)) of model cytochrome *c* are averaged over different conformations of given h. Experimental (B) curves: native state (curve 1), NaCl-induced (curve 2) and acetylation-induced (curve 3) compact denatured states, acid-unfolded state (curve 4).

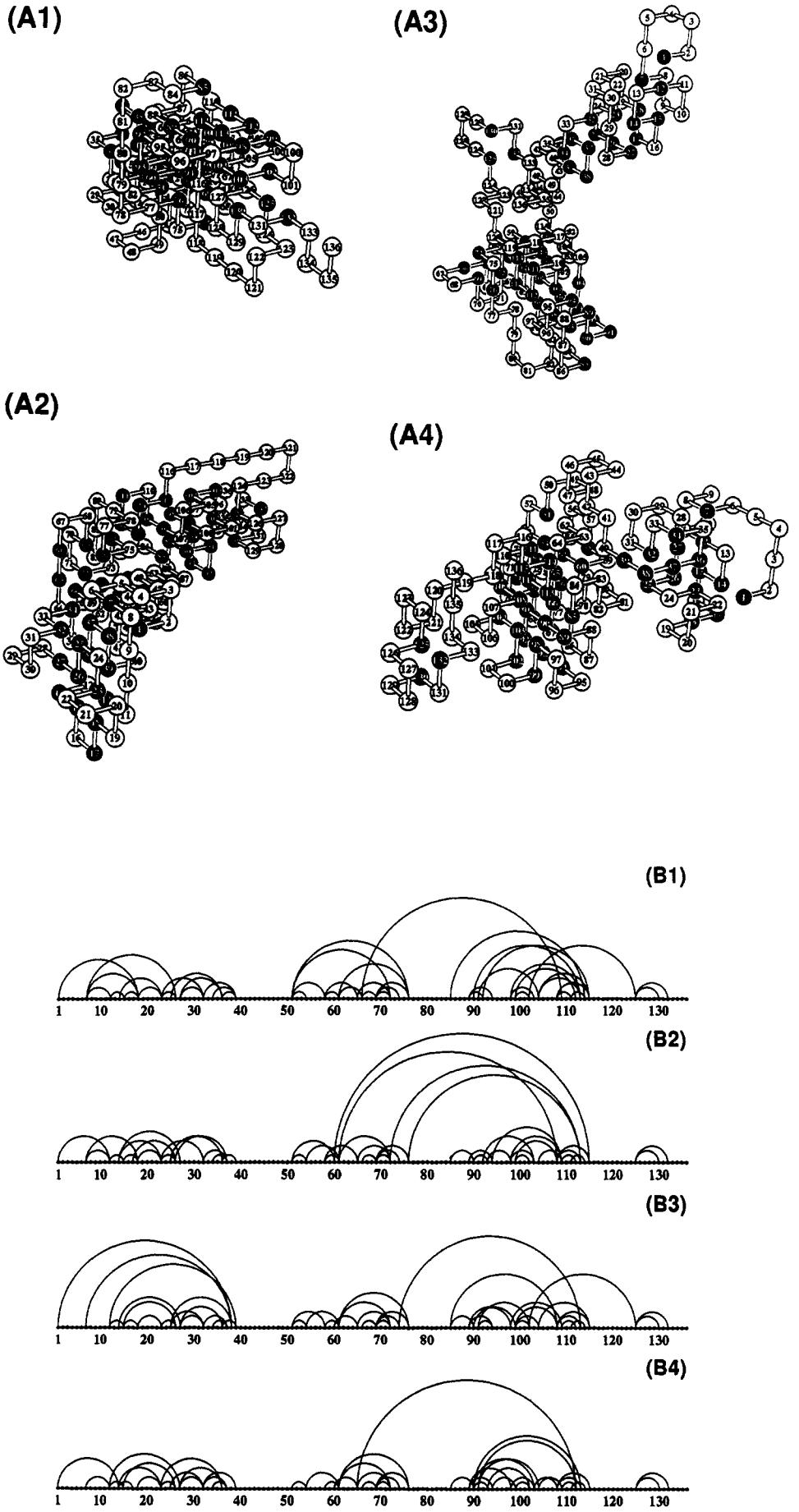

FIGURE 8: Polymer graphs of model conformations of the Shortle staphylococcal nuclease 1–136 fragment ($h = 48$). The conformations are different, but the hydrophobic clustering and the local contacts are quite similar.

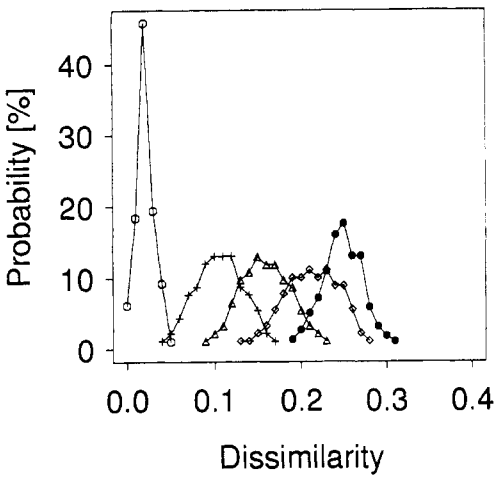

FIGURE 9: Dissimilarity distributions of conformations as a function of h, the number of HH contacts. Conformations that have different contact maps display greatest dissimilarity (●). The more HH contacts (h = 10 (◇), 16 (△), 20 (+), 32 (○)) there are in an ensemble of configurations with identical contact map, the greater is the similarity among their conformations.

d_{ij} between residues i and j. These too are not very predictive of bimodal vs unimodal $P(r)$'s, although in a few extreme cases, where $P(r)$ had two distinct peaks, two "islands" can be observed in the distance matrix.

One property that correlates well with bimodal vs unimodal behavior is hydrophobic clustering. This is shown in Figures 3 and 5 with the help of "polymer graphs" (defined in Figure 3) (Fiebig & Dill, 1993). Unimodal distributions tend to arise from conformations that have a single hydrophobic cluster (Figure 3), whereas bimodal curves tend to result from multiple hydrophobic clusters (Figure 5). But we also find a few exceptions to these generalizations, particularly in conformations of low compactness.

There are some caveats in comparing the lattice model conformations to real proteins. First, the simple HP potential function is very crude and does not contain other specific interactions. Sosnick & Trewhella (1992) observe a broadened unimodal $P(r)$ distribution for ribonuclease A under nonreducing conditions at 67 °C. We do not model disulfide bonds in this study. They also observe a bimodal distribution for the reduced protein in strong denaturant. Second, it is possible that bimodal $P(r)$ behavior observed in experiments may partly arise from artifacts in the scattering models used to Fourier transform the intensities to produce $P(r)$'s (Flanagan et al., 1993). To avoid these potential artifacts it is advantageous to have experimental data that are reported directly as $I(Q)Q^2$ vs Q (Kratky) plots (Glatter & Kratky, 1982). While the lattice model often shows individual chain conformations with bimodal $P(r)$'s, our limited calculations show that more extensive conformational averaging tends to smooth the distributions, see Figure 6. However, until it is possible to get full Boltzmann averages at a given compactness from hydrophobic zippers, our conclusions about conformational averages are limited.

Figure 7 compares Kratky plots for model cytochrome c to experimental data reported by Kataoka et al. (1993). Here we compute averages $\langle I(Q)\rangle$, where the brackets indicate averaging over different conformations of a given number h of HH contacts. The model $I(Q)$ data agree qualitatively with the experimental curves at small Q ($Q < 0.15$), where the experimental data are most reliable. The scale of the y-axes of the experimental and theoretical curves differ by an arbitrary instrumental factor. However, the maxima of the model Kratky curves are slightly shifted toward larger Q when compared to the experimental maxima. This is probably due to our use of an approximate lattice spacing. In the $Q > 0.2$ region the h = 38 (◇) and h = 26 (△) model cytochrome c conformations give SAXS Kratky curves in good agreement with the experiments on the native (curve 1) and compact denatured states (curves 2 and 3). However the model unfolded state (h = 2 (●)) shows a decrease in $\langle I(Q)\rangle Q^2$ with respect to Q whereas the experimental data of the acid unfolded state increases (curve 4). This discrepancy is most likely due to the model's limited resolution. Small spatial shapes, such as the roughness of the protein surface at atomic resolution, contribute to $\langle I(Q)\rangle Q^2$ at large Q, but are not represented by the low-resolution lattice model. Thus the model will underestimate $\langle I(Q)\rangle Q^2$ at large Q.

(3) Structural Similarities among the Compact States of a Given Sequence. Even though each hydrophobic zipper is generated by random decisions, a given amino acid sequence zips to end states that often have certain properties in common. Figure 5 shows three different conformations of the crambin sequence, all of which have h = 18 HH contacts and give bimodal $P(r)$ curves. All three conformations have similar hydrophobic clusters and similar secondary structures. For example, residues 3–9 are in a short antiparallel sheet topology: in Figures 5A1 and 5A3 it is twisted, whereas in Figure 5A2 it is straight. The residues 32–44 have many helix-like turns in all three conformations. All three conformations have one or two nonlocal contacts of a monomer in the region of amino acids 29–34 with monomers 44 or 45. Of course, this comparison of only three conformations cannot necessarily be generalized to a larger ensemble, nor, as we noted before, do these results necessarily apply to real crambin, since the lattice model is crude. They simply indicate the predicted structural similarities among compact conformations of a given sequence. In Figure 8 we compare four polymer graphs of different conformations of the staphylococcal nuclease 1–136 fragment. Each of the polymer graphs has 48 HH contacts. Two distinct clusters of HH contacts form in all four hydrophobic zippers. The first cluster, residues 1–39, is connected to the second one, residues 51–115, by a string of polar monomers, residues 40–50. The second cluster can be further subdivided into two subclusters, residues 51–75 and 85–115, which are again separated by a run of polar (P) residues. It is apparent that most local contacts (small arcs on the polymer graph) are similar in all four structures. However, nonlocal contacts (large arcs) within and between the clusters show more variation. Thus, even though hydrophobic zipper construction paths involve many random decisions, they lead to conformations with considerable similarities in hydrophobic clusters and local contacts but differences in the combination of nonlocal contacts which constitute the particular clusters.

(4) Compact Conformations Have Broad Conformational Diversity. Despite the similarities, the conformations are actually quite diverse. How different are they? We compared them using a simple conformational distance measure described only briefly here and in more detail elsewhere (Yee & Dill, 1993). A dissimilarity score is obtained by correlating two contact matrices, using a measure S between two sets of coordinates c and c':

$$S(c,c') = \frac{\sum_{i>j} |d_{ij}^{-2} - d'_{ij}^{-2}|}{\sum_{i>j} |d_{ij}^{-2} + d'_{ij}^{-2}|} \qquad (3)$$

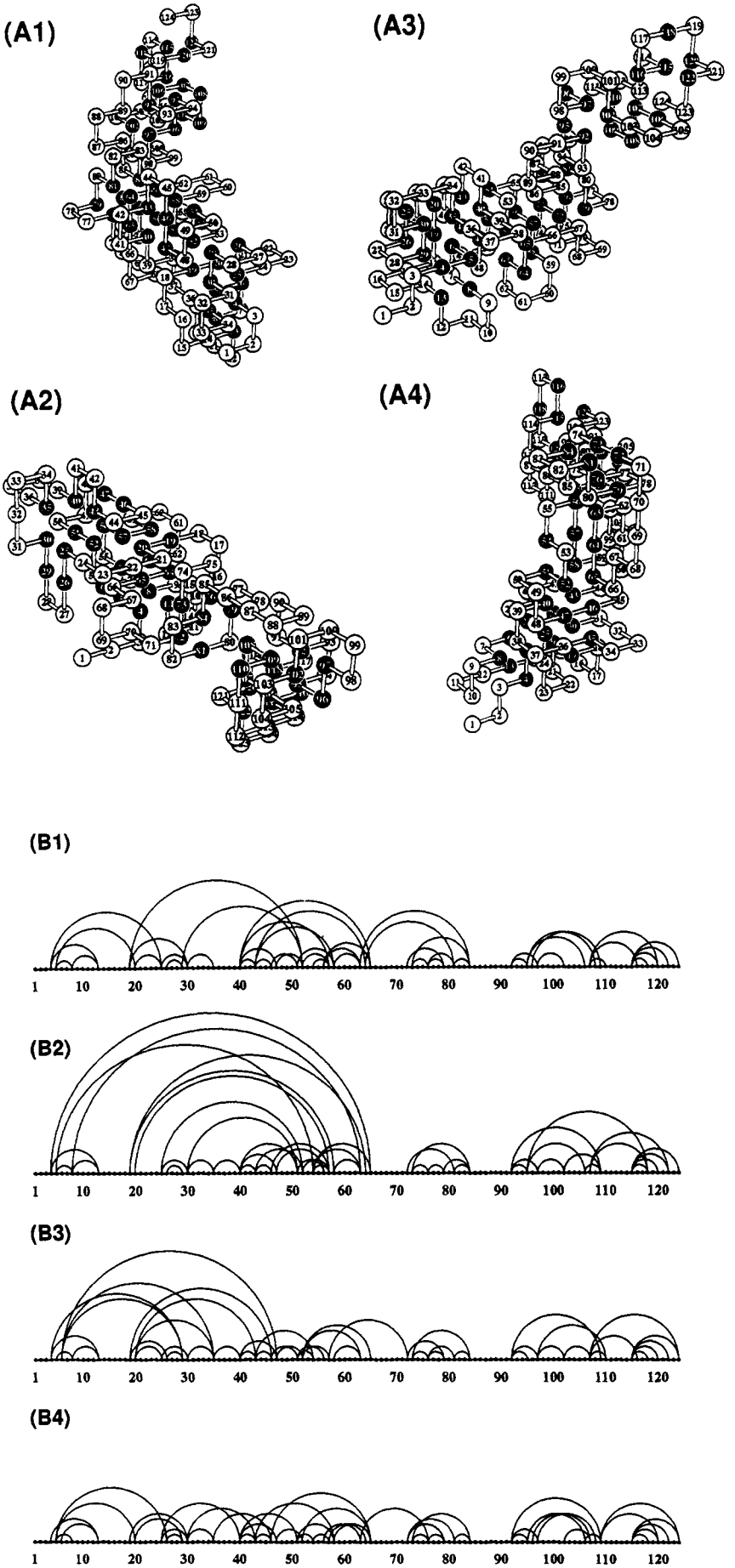

FIGURE 10: Polymer graphs of model ribonuclease A conformations with $h = 48$ HH contacts.

where d_{ij} and d'_{ij} are the Euclidean distance between residues i and j of coordinate sets c and c', respectively.

Consider the similarities of highly compact model crambin conformations that have up to $h = 32$ HH contacts. We first compare different conformations that all have a given contact map, then different conformations from different contact maps (each with $h = 20$ HH contacts). Not surprisingly, we find the distance between all pairwise conformations *within a given* contact map are much smaller than conformational distances *between two* different contact maps (see Figure 9). But as compactness decreases (h becomes smaller), Figure 9 also shows that there ultimately becomes as much conformational diversity *within a given* contact map as there is *between two* contact maps. Figure 9 shows wide conformational diversity among conformations with $h = 20$ HH contacts. In addition, the various sets of 20 HH contacts are not, in general, subsets of the 32 HH contacts that define the native (or near-native) state; there are generally some nonnative contacts. In short, we find that hydrophobic zippers lead to a relatively large ensemble of conformations with different contact maps, but with much recognizable similarity, particularly in local conformations and hydrophobic clustering.

(*5*) *Compact Conformations Depend on Amino Acid Sequence.* The compact conformations generated by hydrophobic zippers depend on the amino acid sequence. Figure 10 shows four polymer graphs of ribonuclease A zipper states with $h = 48$ HH contacts. Comparison of Figures 8 and 10 shows how amino acid sequence affects the spacing of hydrophobic clusters. In particular, polar runs of amino acids often separate hydrophobic clusters.

CONCLUSIONS

We propose a model for compact denatured configurations of globular proteins. We generate compact nonnative conformations using the hydrophobic zippers strategy. This strategy sequentially introduces additional HH contacts into a growing conformation by an opportunistic process. We believe this process also corresponds roughly to the sequence of events in the fast kinetic steps of folding. Most zippers reach compact nonnative end state conformations. We study the properties of these end state conformations. We find that they are slightly less compact than the native state and that many of them bear remarkable similarity to each other, despite the many random choices made in each zipper process. In particular, hydrophobic residues are often clustered, often involve the same hydrophobic amino acids in the sequence, and often have considerable similarities in the helical contacts. Despite these general similarities, the actual conformations are widely different. In these respects hydrophobic zipper end state conformations appear to resemble compact denatured states of globular proteins.

ACKNOWLEDGMENT

We thank Mikio Kataoka, John Flanagan, Don Engelman, and David Shortle for helpful advice.

REFERENCES

Alonso, D. O. V., Dill, K. A., & Stigter, D. (1991) *Biopolymers 31*, 1631–1649.
Baum, J., Dobson, C. M., Evans, P. A., & Hanley, C. (1989) *Biochemistry 28*, 7–13.
Calmettes, R., Roux, B., Durand, D., Desmadrill, M., & Smith, J. C. (1993) *J. Mol. Biol. 231*, 840–848.
Chan, H. S., & Dill, K. A. (1991) *J. Chem. Phys. 95*, 3775–3787.
Dill, K. A., & Shortle, D. (1991) *Annu. Rev. Biochem. 60*, 795–825.
Dill, K. A., Fiebig, K. M., & Chan, H. S. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 1942–1946.
Fiebig, K. M., & Dill, K. A. (1993) *J. Chem. Phys. 94*, 3475–3487.
Flanagan, J. M., Kataoka, M., Shortle, D., & Engelman, D. M. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 748–752.
Flanagan, J. M., Kataoka, M., Fujisawa, T., & Engelman, D. M. (1993) *Proc. Natl. Acad. Sci. U.S.A. 32*, 10359–10370.
Glatter, O., & Kratky, O. (1982) *Small Angle X-ray Scattering*, Academic Press, New York.
Goldstein, H. (1980) *Classical Mechanics*, Addison Wesley, Reading, MA.
Goto, Y., & Fink, A. L. (1990) *J. Mol. Biol. 214*, 803–805.
Hughson, F. M., Wright, P. E., & Baldwin, R. L. (1990) *Science 249*, 1544–1548.
Jeng, M. F., Englander, S. W., Elove, G. A., Wand, A. J., & Roder, H. (1990) *Biochemistry 29*, 10433–10437.
Karplus, M., & Shakhnovich, E. I. (1992) in *Protein Folding* (Creighton, T. E., Ed.) pp 127–195, W. H. Freeman, San Francisco.
Kataoka, M., Hagihara, Y., Mihara, K., & Goto, Y. (1993) *J. Mol. Biol. 229*, 591–596.
Kuwajima, K. (1989) *Proteins 6*, 87–103.
Lattman, E. E. (1989) *Proteins 5*, 149–155.
Lau, K. F., & Dill, K. A. (1989) *Macromolecules 22*, 3986–3997.
Lau, K. F., & Dill, K. A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 638–642.
Ptitsyn, O. B. (1991) *FEBS* Lett. *285*, 176–181.
Ptitsyn, O. B., & Semisotnov, G. V. (1991) in *Conformations and Forces in Protein Folding* (Nall, B. T., & Dill, K. A., Eds.), Chapter 10, pp 155–168, American Association for the Advancement of Science, Washington.
Reingold, E. M., Nievergelt, J., & Deo, N. (1977) *Combinatorial Algorithms: Theory and Practice*, Prentice-Hall, Englewood Cliffs, NJ.
Roder, H., & Wüthrich, K. (1986) *Proteins 1*, 34–42.
Shakhnovich, E. I., & Finkelstein, A. V. (1989) *Biopolymers 28*, 1667–1680.
Shortle, D., & Meeker, A. K. (1989) *Biochemistry 28*, 936–944.
Shortle, D., Chan, H. S., & Dill, K. A. (1992) *Protein Sci. 1*, 201–215.
Sosnick, T. R., & Trewhella, J. (1992) *Biochemistry 31*, 8329–8335.
Stigter, D., Alonso, D. O. V., & Dill, K. A. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 4176–4180.
Yee, D. P., & Dill, K. A. (1993) *Protein Sci. 2*, 884–899.