Coarsely resolved topography along protein folding pathways

Ariel Fernández
Instituto de Matemática, Universidad Nacional del Sur, Consejo Nacional de Investigaciones Científicas y Técnicas, Avenida Alem 1253, Bahía Blanca 8000, Argentina and James Franck Institute and the Department of Chemistry, The University of Chicago, Chicago, Illinois 60637

Konstantin S. Kostov and R. Stephen Berry[a)]
James Franck Institute and the Department of Chemistry, The University of Chicago, Chicago, Illinois 60637

(Received 13 September 1999; accepted 21 December 1999)

The kinetic data from the coarse representation of polypeptide torsional dynamics described in the preceding paper [Fernandez and Berry, J. Chem. Phys. **112**, 5212 (2000), preceding paper] is inverted by using detailed balance to obtain a topographic description of the potential-energy surface (PES) along the dominant folding pathway of the bovine pancreatic trypsin inhibitor (BPTI). The topography is represented as a sequence of minima and effective saddle points. The dominant folding pathway displays an overall monotonic decrease in energy with a large number of staircaselike steps, a clear signature of a good structure-seeker. The diversity and availability of alternative folding pathways is analyzed in terms of the Shannon entropy $\sigma(t)$ associated with the time-dependent probability distribution over the kinetic ensemble of contact patterns. Several stages in the folding process are evident. Initially misfolded states form and dismantle revealing no definite pattern in the topography and exhibiting high Shannon entropy. Passage down a sequence of staircase steps then leads to the formation of a nativelike intermediate, for which $\sigma(t)$ is much lower and fairly constant. Finally, the structure of the intermediate is refined to produce the native state of BPTI. We also examine how different levels of tolerance to mismatches of side chain contacts influence the folding kinetics, the topography of the dominant folding pathway, and the Shannon entropy. This analysis yields upper and lower bounds of the frustration tolerance required for the expeditious and robust folding of BPTI.
[S0021-9606(00)50711-9]

I. INTRODUCTION

A general problem in both materials science and molecular biology is this: What are the underlying laws that determine how, and even whether, molecular systems find their ways to ''special'' structures, such as the physiologically active forms of biomolecules and the crystalline forms of nanoscale materials, and how can our understanding of these laws be used to control the ''structure-seeking'' processes that molecular systems undergo? One approach to this problem is to find a link between the topography of the multidimensional potential-energy surfaces (PES) exhibited by molecular systems and the tendency of these systems to form glasses or selected structures.[1–3] This topography is determined by the set of minima, transition states, and pathways on the PES.[4] The exponential growth of the number of these features with system size renders their exhaustive cataloging prohibitive for all but the smallest systems. This explosive increase in complexity mandates a reduced description of the PES. Such a representation may be achieved by using statistical samples of the stationary points of the PES in the form of linked sequences of minima and the saddles that connect them.[5] The rate coefficients for transitions between adjacent minima along these sequences may obtained from transition state theory.[6,7] This kinetic information is then incorporated in a master equation which describes the dynamics of transitions between different minima, and thereby connects topography and dynamics.[5,7] The master equation approach has the advantage that it offers a direct connection between the dynamics and the state-to-state kinetics on the underlying PES. In addition, it is far less computationally intensive than molecular-dynamics simulations, while at the same time can describe the average dynamical behavior of the system over a range of temperatures, and, perhaps most important, for time intervals orders of magnitude longer than molecular dynamics can achieve.

The procedure described above has been successfully applied to characterize the PES topography of a variety of clusters.[2,3] These studies demonstrate that the glass-forming vs structure-seeking propensity of a system is determined by the sawtooth vs staircase topography of its PES. For example, the topography of the glass-former Ar_{19} exhibits a sawtooth pattern with small energy changes between adjacent minima compared to the saddle heights. In contrast, the topography of $(KCl)_{32}$, which forms a crystal, is staircaselike with large energy differences between adjacent minima. Staircase potentials exhibit highly collective motions, in contrast to the few-atom rearrangements on the sawtooth surfaces.[2,9] Master equations constructed from statistical samples extracted from the complete set of stationary points for small clusters can reliably capture the essential features of the relaxation dynamics for these systems.[8] These results

a)Electronic mail: berry@rainbow.uchicago.edu

suggest that a link between the generic properties of PES and structure-seeking propensities may be found in other molecular systems. One such incompletely understood process with central biomedical importance is the ability of proteins to fold into their native structures. Recent research shows that the PES topography of a ''three-color 46-bead'' model protein that forms a β-barrel is similar in many respects to those of the structure-forming clusters.[9]

A different approach to modeling the structure-seeking properties of proteins is proposed in the preceding paper (paper I).[10,11] This method combines several ingredients to retain all essential features of proteins in a minimalist model readily amenable to simulations and analysis.[12] A central simplification is the coarse-grained treatment of the polypeptide dynamics in terms of transitions between a discrete set of backbone torsional states.[13] At regular time intervals the conformational state of the polypeptide backbone is inspected for the formation (or destruction) of torsional patterns that correspond to elementary secondary structure motifs such as α-helices, β-sheets, loops, or any combinations thereof. This pattern recognition operation also includes an analysis of the complementarity of the hydrophobic/polar side-chain contacts and identifies consensus windows of residues that belong to a given structural element. Three classes of residues emerge: Free residues not engaged in any structural motif, residues engaged only in secondary structure, and those engaged in tertiary structure. The transition frequencies of the backbone dihedral angles of amino acids identified as participating in structural motifs are reduced to values compatible with experimental measurements of torsional motions in such motifs. The computational cycles simulating backbone dihedral angle dynamics and pattern recognition are repeated to follow the time evolution of the topological state of the polypeptide chain. This iterative algorithm is based on recognizing the topological compatibility of different torsional states and side chain contacts with secondary and tertiary structure motifs, while the precise configurational geometry is irrelevant. Rather than deriving the protein dynamics explicitly from an underlying PES, the method of structural motifs focuses on the kinetics of the formation and transitions between topological states of the chain which represent secondary and tertiary structure motifs. The present paper is a first step towards bridging the topological and topographical approaches by analyzing the new method in terms of the concepts developed previously for studying the topography of multidimensional PES.

Paper I demonstrates that the combination of discretized torsional dynamics and pattern recognition yields the native structure of the protein BPTI, starting from a random coil state.[10,11] The method of structural motifs makes it possible to evaluate transition rates between the different topological states of BPTI during the evolution from the unfolded to the folded state. This work uses this kinetic information with the principle of detailed balance to characterize the energy landscape explored by the dominant folding pathways. This analysis is complemented by computing the Shannon entropy which gives a measure of the changing number of significantly populated topological states as folding proceeds.

The companion paper incorporates tolerance to incorrect conformations and to hydrophobic/polar (h/p) mismatches in the pattern recognition procedure. The results demonstrate that some degree of such tolerance tolerance accounts for the robust and expeditious folding of BPTI to its active structure. What are the bounds on error tolerance that produce an expeditious search for the native structure? The present work addresses this question by examining in greater detail the influence of different levels of frustration on the folding kinetics, the topography of the dominant folding pathway of BPTI, and the Shannon entropy. Section II briefly reviews the model. The results are discussed in Sec. III, while the last section presents the conclusions.

II. METHODOLOGY

The algorithm for long-time torsional dynamics is described in detail in the companion paper and, therefore, we give only a brief summary here.

The primary amino acid sequence of the polypeptide chain is represented by a reduced set of amino acids characterized by their (1) hydrophobicity, and (2) accessible torsional states of the Φ, Ψ backbone dihedral angles.[14] Three kinds of residues are distinguished based on hydrophobicity: Hydrophobic, neutral, and polar (or hydrophilic). The amino acids are also grouped into four torsionally distinct classes—*L*-alanyl-like, glycine, proline, and residues preceeding proline—which have different accessible basins in the Ramachandran maps of the Φ, Ψ dihedral angles. The Ramachandran map of any amino acid consists of a discrete and small number of basins of attraction[14] (a maximum of four for the glycine residue) and, hence, a natural simplification is to view the Φ, Ψ coordinates modulo the basin of attraction to which they belong. The validity of this assumption for providing a correct description of the dynamics is supported by the fact that the dominant factor which restricts the accessible range of the Φ, Ψ angles is the presence of the C_β atoms in the side chains. Analysis of the available crystal and NMR (nuclear magnetic resonance) structures of proteins shows that the conformations of 90% of all residues except Gly fall in one of three low-energy Ramachandran basins.[15] Since these basins are generally insensitive to the conformations of other residues, individual amino acids will be in their locally favorable conformations for most of the folding process. Thus, the model contains eight types of amino acids (12 if ''neutral'' residues are distinguished) each of which can be in one of the discrete set of at most four torsional states determined by its torsional class.

The rigid peptide bonds prevent correlations in the Φ, Ψ transitions of different residues and hence the dynamics of the polypeptide chain is modeled as a sequence of independent (Φ_i,Ψ_i) transitions in the discretized torsional space. The state of the chain at time t is characterized by the variable $\{R(\mathbf{y}(t),n)\}_{n=1,\ldots,N}$, which can take values between 1 and 4 depending both on the torsional types of the amino acids as determined by the primary sequence, and on their torsional states at that particular time (see paper I for a more detailed discussion). The interbasin transition frequencies are chosen from temperature-dependent Gaussian distributions centered about $\tau_f = 10^{-11}$ s for ''free'' residues not part of any secondary structure, about $\tau_s = 10^{-7}$ s for residues en-

gaged in secondary structure, and about $\tau_t = 10^{-3}$ s for residues participating in tertiary structure. The experimental justification of these transition times is discussed in the companion paper. It is in principle possible to derive these transition times theoretically for a given model by considering the relaxation kinetics on the PES as described by a master equation.

The participation of residues in secondary structure is determined by searching at fixed time intervals of 64 ps for torsional patterns (consensus regions) in windows of at least six adjacent amino acids. The interval of 64 ps was chosen because it is the mean first passage time to form a consensus region of six residues starting from an arbitrary conformation.[16] For example a right-handed α-helix requires a window of residues with $R(\mathbf{y},n)=3$ and the appropriate hydrophobicity pattern, while a β-sheet is characterized by a sequence of extended conformations marked by $R(\mathbf{y},n)=1$. Once a group of residues are identified as part of secondary structure, their mean interbasin transition frequencies are reduced from 10^{-11} to 10^{-7} s. Similarly, tertiary structure forms when secondary structure elements are connected with a loop (a region of dominant curvature that enables long-range interactions), whereupon the mean interbasin transition frequency for the residues involved is lowered to 10^{-3} s.

The topological state of the polypeptide chain is described in terms of contact patterns. A contact pattern (CP) is defined as a cluster of kinetically related structures that interconvert within 64 ps. A CP may contain two or more consensus windows separated by free residues. An elementary contact pattern transition involves consensus fulfillment in a single window of L adjacent amino acids. Two contact patterns are connected if and only if they are linked by an elementary torsional transition. Since the probability of simultaneous events is vanishingly low, all elementary transitions between CPs involve only one consensus window. The evolution of the topological state of the chain is thus represented by the pattern recognition procedure as a series of kinetically determined transitions between different contact patterns.

The transitions between two CPs may involve either the formation of a new consensus window, or the growth or dismantling of an existing window. First, we consider transitions between two CPs caused either by the formation of new consensus windows or the growth of existing ones. In the absence of frustration all hydrophobic and hydrophilic (h/p) contacts and torsional angles must be correctly matched. In this case the transition rate W_{ij} for formation of CP_i from CP_j involving L_{ij} residues is simply $\tau_f^{-1}(\frac{1}{2})^{L_{ij}}$. This follows from the independence of the Ramachandran transitions of different residues and the equal probability for transitions between the correct and incorrect states.

A central feature of the model is the tolerance both to local frustrating conformations and to mismatches in the contacts between hydrophobic and polar amino acids (h/p contacts). The preceeding paper discusses the importance of frustration tolerance in achieving the active structure of BPTI. In this paper we explore in greater detail the influence of different degrees of tolerance to h/p mismatches on the ability of BPTI to form its active structure *vis-a-vis* the accessible PES topographies explored by the dominant folding pathway.

With frustration allowed, the transition rate W_{ij} for formation of CP_i from CP_j becomes

$$W_{ij}=\sum_{F=0}^{L_{ij}}\sum_{X=0}^{M_{ij}} 2^{-(L_{ij}-F)}\binom{L_{ij}}{F}\binom{M_{ij}}{X}\tau_f^{-1}P(i,j,F,X). \quad (2.1)$$

$2^{-(L_{ij}-F)}$ is the probability of obtaining $L_{ij}-F$ residues in the Ramachandran basin compatible with CP_j. The binomial factors give the number of possible distributions of F torsionally incompatible residues amongst L_{ij} residues, and similarly for the X h/p contact mismatches amongst the M_{ij} possible contacts. The function $P(i,j,F,X)$ represents the probability that CP_i will be formed in the presence of F torsional and X h/p mismatches and is equal to unity for $F=X=0$. We study different forms of $P(i,j,F,X)$, which will be specified explicitly in the next section.

Next we turn to destructive CP transitions. A consensus window is dismantled if a consecutive stretch of residues with length greater or equal to 30% of the total length of the window is in the incorrect conformation, a requirement that is justified in Ref. 17. Therefore, the rate W_{ji} for the destructive transition between two CPs in the absence of frustration is given by

$$W_{ji}=2^{-(1/3)L_{ij}(\frac{2}{3})}L_{ij}\tau_s^{-1}. \quad (2.2)$$

In the presence of frustration the rate of CP dismantling is accelerated since some torsional mismatches already exist. These rates are estimated directly from the simulations.

The energies of the different topological states are obtained from the transition rates by applying the principle of detailed balance

$$W_{ij}P_j^{\mathrm{eq}}=W_{ji}P_i^{\mathrm{eq}}, \quad (2.3)$$

where

$$P_i^{\mathrm{eq}}=\frac{D_i\exp(-\beta U_i)}{Z}. \quad (2.4)$$

with $Z=\Sigma_j D_j\exp(-\beta U_j)$, and $\beta=1/kT$. D_i is the degeneracy of CP_i which may be easily found from the number of Ramachandran basins $q_n(i)$ available to the N_f free residues in CP_i

$$D_i=\prod_{n=1}^{N_f} q_n(i). \quad (2.5)$$

The energy difference between two CPs may then be obtained from

$$U_i-U_j=\frac{1}{\beta}\ln\left(\frac{D_iW_{ji}}{D_jW_{ij}}\right). \quad (2.6)$$

Typically the detailed balance principle is used to obtain the kinetics compatible with a given energetic profile, whereas here we do the opposite since the kinetic information is readily accessible from the dynamical simulations performed in paper I. Effective barriers for the transitions between CPs along these folding pathways are obtained by equating the transition rates specified above with an Arrhenius expression

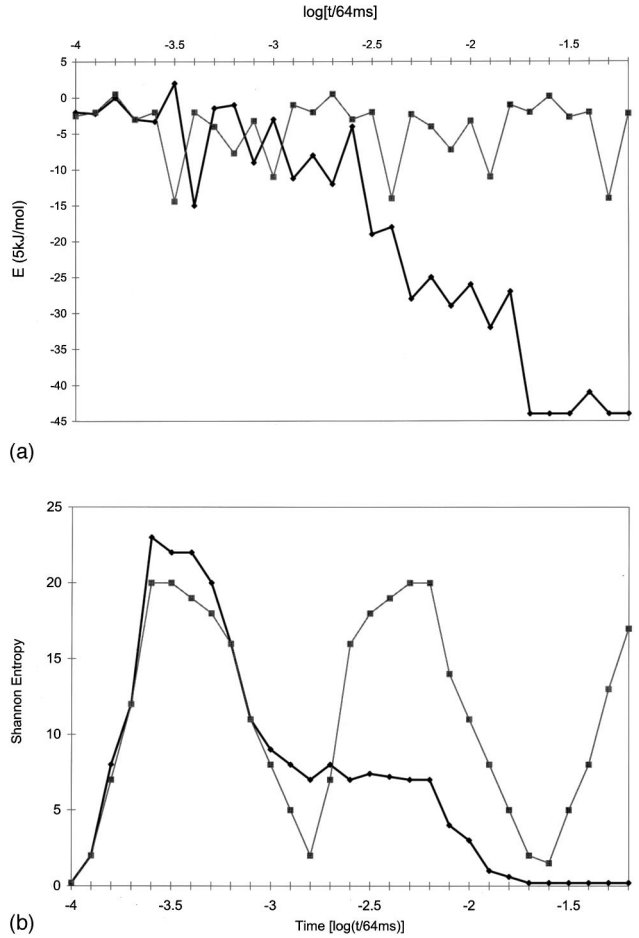

FIG. 1. (a) Topography of the PES of the dominant folding pathway of BPTI expressed as a sequence of minima and effective saddles. Simulations with frustration tolerance (solid line) and no tolerance (dashed line) are shown. (b) Time evolution of the Shannon entropy for the frustration tolerance (solid line) and zero tolerance (dashed line) cases.

for the height of the activation barrier. This allows the construction of a cross-section of the PES along the dominant folding pathways of BPTI.

III. RESULTS AND DISCUSSION

A. Topography along the dominant folding pathway

Figure 1(a) displays the topography of the PES obtained from Eq. (2.6) along the dominant folding pathway of BPTI as it evolves from a random coil initial state. A sequence of minima and effective barriers is presented, with a few exceptions where the barriers could not be resolved. The energies of the minima correspond to the average energies of the kinetically related structures that constitute a given CP. Since the model is entirely kinetically determined, this trajectory simply represents the most probable folding pathway of BPTI. The solid line corresponds to a run which includes tolerance to h/p contact mismatches as discussed in Sec. III B. Initially a series of misfolded states are formed and dismantled with no discernible pattern in the topography. Subsequently, a series of staircase transitions lead to a nativelike intermediate with low energy. This intermediate exists for an appreciable amount of time as it undergoes structural refinements, which eventually result in the formation of the native structure of BPTI accompanied by a large drop in energy. The most probable folding pathway of BPTI displays an overall monotonic decrease in energies of its local minima, with some staircaselike regions.

B. Shannon entropy

A full understanding of the folding process requires knowledge not only of the most probable folding pathway displayed in Fig. 1(a), but also of the set of alternative pathways that the protein may take as folding proceeds. One measure of the diversity of pathways available to the protein is the Shannon entropy associated with the time-dependent probability distribution over the ensemble of CPs. The Shannon entropy is defined as

$$\sigma(t) = -\sum_{i \in CP} P_i(t) \ln P_i(t), \tag{3.1}$$

where $P_i(t)$ is the probability of finding the system in CP_i at time t and the sum is over all contact patterns that are combinatorially possible for BPTI. $\sigma(t)$ is maximum if there is equal probability for the occurrence of all CPs, and zero if only one CP is populated. Thus, the Shannon entropy provides a measure of the number of populated CPs during the folding process. The time dependence of the probability vector reflects the exploration of the PES topography. The $\{P_i(t)\}$ in Eq. (3.1) are obtained from a numerical solution of the master equation $\mathbf{P}(t) = \exp(t\mathbf{W})\mathbf{P}(0)$, by diagonalizing the matrix of transition rates $\mathbf{W}$.

It is important to stress that the contact patterns represent *topologically* different states and hence the Shannon entropy in this context is different from the configurational entropy. For example, the topological description maps all possible random coil configurations into one state. Likewise, the native state also represents one topological state although its configurational entropy is clearly much smaller than that of the random coil.

The time evolution of the Shannon entropy is depicted in Fig. 1(b). This plot complements the information presented in Fig. 1(a), which focuses only on the maximum probability states, by depicting the time-dependent spread in probability over the CPs. $\sigma(t)$ is zero at the random coil initial condition when all amino acids are in the free state. Initially $\sigma(t)$ grows rapidly as the chain explores a wide variety of CPs. Clearly if $\sigma(t)$ is large the multitude of possible folding pathways may dominate over the most probable one. Consequently, in these time intervals the representation of the folding process may require several pathways, not just one dominant pathway. After this initial exploration misfolded states are formed and dismantled, accompanied by a large reduction in $\sigma(t)$. The Shannon entropy remains at a plateau value during the formation of the nativelike intermediate. Finally, the rearrangements of this intermediate to form the active structure of BPTI are accompanied by a reduction of $\sigma(t)$ to zero.

C. Frustration dependence

The companion paper demonstrates that incorporating frustration tolerance in the model is essential for the formation of the active structure of BPTI. Here we consider in greater detail different levels of tolerance to h/p mismatches.

The limit of zero-frustration requires perfect matching of all h/p contacts in order to allow the formation of a CP and the corresponding renormalization of transition frequencies for the residues that participate in this pattern. The simulations carried out with this model fail to reach the native structure of BPTI. Instead, partially folded states form and dismantle periodically. This behavior is clearly seen in Figs. 1(a) and (b), where the dashed lines represent the topography followed by the dominant pathway and the Shannon entropy for the zero-frustration version of the algorithm. The transient formation of partially folded structures is mirrored by periodic high-lying minima in the energy, and a periodic change of $\sigma(t)$. Clearly, imposition of overly rigid requirements for CP formation leads to a dramatic reduction in the number of ''correct'' pathways available to the protein. Making correct downhill steps towards the native structure requires exhaustive conformational searches and any secondary structure motifs that may form have a high probability of dismantling before they can grow or combine with other motifs to form tertiary structure. Moreover, the native states of proteins, which inherently have some degree of frustration, will never be reached with such an algorithm.

The other limit of allowing an arbitrary amount of frustration is equally unacceptable. In this case everything is possible and every conformation becomes equally likely. Clearly, the amount of frustration that represents the behavior of a real protein lies between these two limits. The preceeding paper expresses the frustration-dependent probability $g(i,j,X)$ of a $\mathrm{CP}_j \rightarrow \mathrm{CP}_i$ transition by the logistic function

$$g(i,j,X) = [1+\exp(-\arctanh\{2[M_{ij}-X]/M_{ij}-1\})]^{-1}, \qquad (3.2)$$

which smoothly bridges the two limits by decreasing as frustration increases. This function satisfies the limiting conditions $g(i,j,0)=1$ and $g(i,j,M)=0$ (see Fig. 7 of paper I). An identical function is used to describe the tolerance to torsional mismatches, and the final probability of a constructive $\mathrm{CP}_j \rightarrow \mathrm{CP}_i$ transition involving L_{ij} residues with F torsional and X contact mismatches is given by the product, $P(i,j,F,X)$, of the two logistic functions. Paper I demonstrates that incorporating frustration tolerance into the pattern recognition procedure is essential in order to obtain the native structure of BPTI.

In order to establish the minimum and maximum levels of frustration required to produce an expeditious search for the native structure we use a step function for the frustration-dependent probability of CP formation

$$g(i,j,X) \equiv \theta_{ij}(X) = \begin{cases} 1 & \text{if } X/M_{ij} \leqslant X_c \\ 0 & \text{if } X/M_{ij} > X_c \end{cases}. \qquad (3.3)$$

The formation of a CP is prohibited if the fraction of total contacts that represent h/p mismatches exceeds the threshold value X_c.

Using the function

$$g(i,j,X) = \begin{cases} [1+\exp(-\arctanh\{2[M_{ij}-X]/M_{ij}-1\})]^{-1} & \text{if } X/M_{ij} \leqslant X_c \\ 0 \quad \text{if } X/M_{ij} > X_c \end{cases}, \qquad (3.4)$$

yields results identical to those obtained with the step function.

We have studied the folding behavior of BPTI for a variety of threshold values X_c. Figure 2 displays representative examples of the most important cases. Simulations with $0 < X_c < 0.2$ show behavior qualitatively similar to that according to the zero-frustration algorithm ($X_c = 0$) and fail to reach any low energy state of BPTI. The expedient and reproducible folding depicted in Fig. 1 is achieved only in the approximate range $0.2 \leqslant X_c < 0.3$. When $0.3 \leqslant X_c < 0.4$ the active structure of BPTI is also reached, albeit considerably slower. The solid line in Fig. 2(a) shows that in this case the initial period of formation of folded structures is longer, and is followed by a slow descent in energy to the native state. The topography that is observed during folding clearly depends on how the energy surface is explored, i.e., on the allowed level of frustration. If X_c is increased to 0.4 a single final state can no longer be reached reproducibly: different runs reach different final states. This is evidenced in Fig. 2(b) by the high Shannon entropy exhibited by the system at long times. While the $X_c = 0.4$ simulations occasionally reach the active structure, they frequently converge to structures with energies *lower* than the native state, as depicted in Fig. 2(a), dashed line. Similar results are obtained as X_c is increased beyond 0.4.

The choice of tolerance limit of course affects the rates of passage along the landscape. Consequently, through the ''backward'' inference of well-to-well energy differences from these rates, increasing the h/p mismatch tolerance dramatically transforms the PES landscape. On one hand, the number of possible CPs is greatly increased as new structural elements with imperfect contact matching become possible. The energy differences between existing CPs are also affected, since the changes in the $\mathrm{CP}_j \leftrightarrow \mathrm{CP}_i$ transition rates affect the relative energies of the two CPs as obtained from the detailed balance principle. Therefore, the CP energies of runs with different levels of frustration are not directly comparable. While the lower-than-native energy of the final state in Fig. 2(a) appears to be an artifact of the model, it is important to characterize this and related states. An analysis of

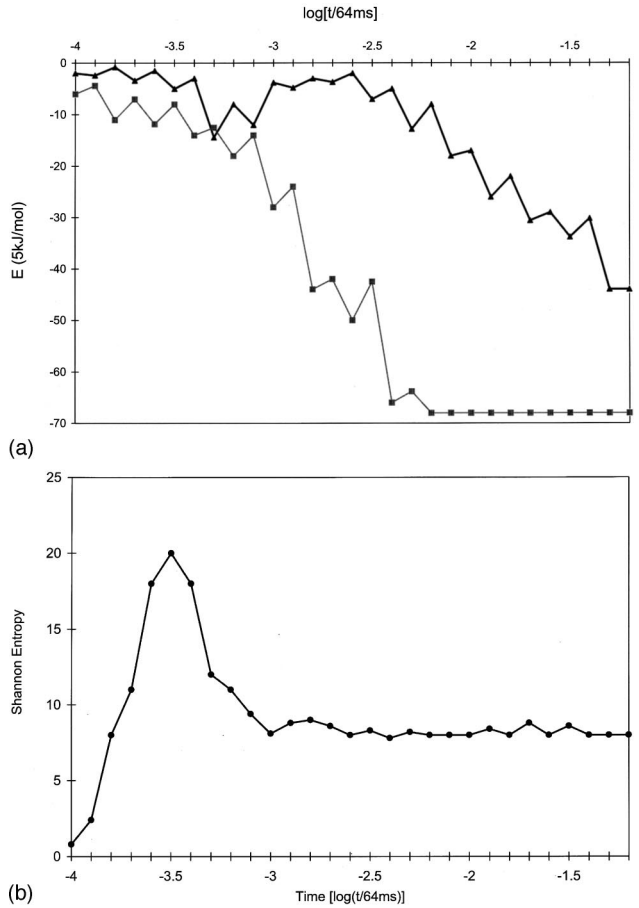

FIG. 2. (a) Frustration dependence of the folding pathways: solid line $X_c = 0.3$, dashed line $X_c = 0.4$. (b) Evolution of the Shannon entropy for $X_c = 0.4$.

the contact matrix of this state shows that it has a compact structure characterized by formation of the non-native (5,30) Cys–Cys disulfide bond with the concurrent breaking of the (30,51) and (5,55) disulfide bonds. The reshuffling of disulfide bonds is made possible by the reducing solvent conditions implicit in the present study. The breaking of the (30, 51) and (5,55) bonds releases the *C*-terminal α-helix, which then becomes free to move and form a tertiary interaction with the β-sheet region, thereby transforming the oblate native state into a globular structure. These findings are consistent with experimental evidence that BPTI forms a compact structure under reducing conditions.[18–20] In cases of very complex potential surfaces, such as that of BPTI, it may be possible to use the range of tolerance level that produces adequate folding as a device to *infer* the topography of the surface. A detailed characterization at the CP level of the compact states reached at high frustration levels will be given in a forthcoming publication.

D. Discussion

It would be instructive to compare these results with more exhaustive studies[6,9,21] of PES landscapes in a variety of systems which construct full databases or statistical samples of the minima, saddle points, and pathways on these PES. Recent work on clusters[2,5,8] and simple protein models[9] has demonstrated that structure-seeking systems have a staircaselike PES topography, while glass-formers exhibit a sawtoothlike topography. It is of interest to see if the paradigms established for these systems are also valid for the present model. Unfortunately, a direct comparison is prohibited by the coarse-grained nature of the torsional dynamics and since here we can only compute effective saddles. Nonetheless, it appears that portions of the landscape explored by the dominant folding pathway display the staircaselike topography characteristic of structure-seekers.

It is important to distinguish this dynamical model from the other computer simulations of protein folding. For example, molecular-dynamics or Brownian dynamics simulations involve numerical solutions of the Newton or Langevin equations of motion to find the time dependence of the protein coordinates.[22] In contrast, the present model does not explicitly include coordinates to describe the geometrical configuration of the polypeptide. Instead, the topological state of the system is specified by the sequence of Ramachandran basins occupied by the amino acids in the protein. Since the dynamics is approximated modulo a discrete set of Ramachandran basins, the dynamical state of the protein evolves on a lattice. This lattice, however, is conceptually different from the three-dimensional lattices used in many recent studies of protein folding.[23] While in the latter lattice models each lattice node represent the spatial position of one residue, each node of the lattice used here specifies the local-equilibrium state of the whole chain. The time evolution of the chain configuration involves transitions between the nodes of this lattice.

The lattice that emerges as a consequence of treating the dynamics modulo the Ramachandran basins provides a ground for comparison of the present model and the rotational isomeric states (RIS) model used in polymer physics. The RIS model was initially proposed to compute equilibrium properties of (bio)polymers[14] such as the radius of gyration by averaging over a discrete set of conformers. Subsequently RIS-like approaches employing a master equation to describe the transitions between the different conformers[24–26] have been used to describe polymer dynamics. More recently a generalized RIS model has been employed to obtain equilibrium averages that serve as the inputs to a mode-coupling theory of alkane dynamics.[27] While our coarse representation of torsional dynamics is similar in many respects to the RIS models, there are substantial differences. The most notable difference is that while the RIS models describe geometrical states of the polypeptide, the present treatment is topological in nature. Another distinction is the pattern recognition algorithm, with the concomitant concepts of frustration tolerance and frequency renormalization. The combination of these ideas with a topological rather than geometrical representation of the polypeptide chain allows the selection of the structurally relevant states out of the enormous number of geometrically possible conformational states. The parallels and differences between the RIS model and our approach as well as the validity of a RIS-like approach to calculating dynamical properties of polypeptides will be addressed in future work.

Another approach to folding of BPTI has focused on the

formation of sulfur–sulfur bonds as key stabilizing factors of intermediates along the folding path, as well as of the native structure.[28–30] That approach as well as this have been able to account for the folding of this species. What factors are causes and what are effects and consequences remains to be seen as more is learned about the actual interresidue forces and the way guided-diffusive searches occur. By inferring more information about the topographies of pathways from multiple topological searches, we hope in time to be able to help resolve that issue.

E. Conclusions

We use the detailed balance principle to obtain from the state-to-state transition rates the PES topography explored by the dominant folding pathway of BPTI. This pathway displays the clear signature of a good-structure seeker: An overall monotonic decrease in energy with a large number of staircaselike steps. The complementary analysis of the PES topography and the Shannon entropy suggests that the folding of BPTI proceeds in several stages, consistent with prior knowledge about this system.[28–30] In the early stage misfolded states are formed and then dismantled. Eventually the protein finds its way downhill through a series of staircase transitions on the PES forming in the process a nativelike intermediate with constant Shannon entropy. In the last stage, the structure of the intermediate is refined to form the native structure. The degree of tolerance to h/p contact mismatches greatly influences the folding process. An expeditious and robust search for the native state of BPTI is achieved only in a small tolerance range. Low frustration tolerance fails to yield any low-energy state of BPTI, while at high tolerance a variety of final states are reached without any reproducibility.

ACKNOWLEDGMENTS

K.K. is supported by a postdoctoral fellowship from the Program in Mathematics and Molecular Biology. A.F. thanks the J. William Fulbright Foreign Scholarship Board for a Grant that made this collaboration possible. R.S.B. acknowledges the support of a Grant from the National Science Foundation for his part in this work.

[1] O. M. Becker and M. Karplus, J. Chem. Phys. **106**, 1495 (1997).
[2] K. D. Ball *et al.*, Science **271**, 963 (1996).
[3] R. S. Berry, Int. J. Quantum Chem. **58**, 657 (1996).
[4] D. J. Wales, J. Chem. Soc., Faraday Trans. **89**, 1305 (1993).
[5] R. E. Kunz and R. S. Berry, J. Chem. Phys. **103**, 1904 (1995).
[6] K. D. Ball and R. S. Berry, J. Chem. Phys. **109**, 8541 (1998).
[7] K. D. Ball and R. S. Berry, J. Chem. Phys. **109**, 8557 (1998).
[8] K. D. Ball and R. S. Berry, J. Chem. Phys. **111**, 2060 (1999).
[9] R. S. Berry, N. Elmaci, J. P. Rose, and B. Vekhter, Proc. Natl. Acad. Sci. USA **94**, 9520 (1997).
[10] A. Fernández and R. S. Berry, J. Chem. Phys. **112**, 5212 (2000), preceding paper.
[11] A. Fernández, K. Kostov, and R. S. Berry, Proc. Natl. Acad. Sci. USA **96**, 12991 (1999).
[12] A. Fernández and A. Colubri, J. Math. Phys. **39**, 3167 (1998).
[13] A. Fernández, Phys. Chem. Chem. Phys. **1**, 861 (1999).
[14] P. J. Flory, *Statistical Mechanics of Chain Molecules* (Wiley-Interscience, New York, 1969).
[15] *Protein Folding*, edited by T. E. Creighton (W. H. Freeman and Company, New York, 1992).
[16] R. Zwanzig, A. Szabo, and B. Bagchi, Proc. Natl. Acad. Sci. USA **89**, 20 (1992).
[17] A. Fernández and A. Colubri, Phys. Rev. E **60**, 4645 (1999).
[18] D. Amir and E. Haas, Biochemistry **27**, 8889 (1988).
[19] E. E. Gussakovsky and E. Haas, FEBS Lett. **308**, 146 (1992).
[20] V. Ittah and E. Haas, Biochemistry **34**, 4493 (1995).
[21] M. A. Miller, J. P. K. Doye, and D. J. Wales, J. Chem. Phys. **110**, 328 (1999).
[22] C. L. Brooks III, M. Karplus, and B. M. Pettitt, Adv. Chem. Phys. **71**, 1 (1988).
[23] *Protein Folding*, edited by T. E. Creighton (W. H. Freeman and Company, New York, 1992).
[24] C. Hall and E. Helfand, J. Chem. Phys. **50**, 4831 (1982).
[25] G. J. Moro, J. Chem. Phys. **94**, 8577 (1991).
[26] G. J. Moro, J. Chem. Phys. **97**, 5749 (1992).
[27] M. Guenza and K. F. Freed, J. Chem. Phys. **105**, 3823 (1996).
[28] T. E. Creighton, J. Mol. Biol. **113**, 275 (1977).
[29] T. E. Creighton, Prog. Biophys. Mol. Biol. **33**, 231 (1978).
[30] C. J. Camacho and D. Thirumalai, Proc. Natl. Acad. Sci. USA **92**, 1277 (1995).