INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE

H4.SMR/642 - 9

College on Methods and Experimental Techniques in Biophysics

28 September - 23 October 1992

Protein Folding Funnels

J.N. ONUCHIC
University of California, U.S.A.

These are preliminary lecture notes, intended only for distribution to participants.

Biophysics

Protein folding funnels:

A kinetic approach to the sequence–structure relationship

Peter E. Leopold[1], Mauricio Montal[1,2], and José Nelson Onuchic[1*]

Departments of Physics[1] and Biology[2]

University of California, San Diego

La Jolla, CA 92093–0319

Communicated by John Hopfield, _____, 1992 (received 31 October 1991)

* to whom correspondence should be addressed

ABSTRACT

A lattice model of protein folding is developed to distinguish between amino acid sequences that do and do not fold into unique conformations. While Monte Carlo simulations provide insights into the long–time processes involved in protein folding, these simulations cannot systematically chart the conformational energy surface that enables folding. By assuming that protein folding occurs after chain collapse, a kinetic map of important pathways on this surface is constructed through the use of an analytical theory of probability flow. Convergent kinetic pathways, or "folding funnels", guide folding to a unique, stable, native conformation. Solution of the probability flow equations is facilitated by limiting treatment to diffusion between geometrically-similar collapsed conformers. Similarity is measured in terms of a reconfigurational distance. Two specific amino acid sequences are deemed foldable and nonfoldable because one gives rise to a single, large folding funnel leading to a native conformation and the other has multiple pathways leading to several stable conformers. Monte Carlo simulations demonstrate that folding funnels calculations accurately predict the fact of and the pathways involved in folding specific sequences. The existence of folding funnels for specific sequences suggests that geometrically–related families of stable, collapsed conformers fulfill both kinetic and thermodynamic requirements of protein folding.

A first functionally–important action of any protein is to fold. While the general process of protein folding is not fully understood, it has been known since the ribonuclease refolding experiments of Anfinsen that small globular proteins fold in the absence of any catalytic biomolecules. From this fact, it was surmised that folded proteins exist in the global minimum free energy state [1]. Soon thereafter, Levinthal used a simple counting argument to demonstrate that Anfinsen's thermodynamic hypothesis implied a paradoxical result: a typical protein has far too many conformations to permit a thorough search for the global minimum [2,3].

Since the statement of Levinthal's celebrated paradox, several groups have attempted to preserve the concept of the global minimum by arguing that compactness [4] or native state propensities [5] reduce the effective size of conformation space. Neither requirement resolves the paradox because compactness alone lessens but does not eliminate the conformation counting problem and native state propensities cannot be justified from amino acid sequence analysis. Accordingly, a plausible resolution to the Levinthal paradox might be obtained using a kinetic rather than a thermodynamic approach to the folding problem. Folded proteins are not (necessarily) equilibrium entities; rather, they are metastable with lifetimes longer than (or perhaps about equal to) their functionally–significant lifetimes. To overcome the Levinthal problem, proteins must have a broad conformational focussing property to direct folding to a unique, locally–stable, kinetically–accessible "native" conformation.

Here, we introduce the concept of "protein folding funnels," a kinetic mechanism for understanding the self–organizing principle of the sequence–structure relationship. This concept follows from a few general considerations (Figure 1). First,

proteins fold from a random state by collapsing and reconfiguring [4]; second, reconfiguration occurs diffusively and follows a general drift from higher energy to lower energy conformations; and third, reconfiguration occurs between conformations that are geometrically similar, i.e., global interconversions are energetically prohibitive after collapse, so local interconversions alone are considered. We define the folding funnel as a collection of geometrically–similar collapsed structures, one of which is thermodynamically–stable with respect to the rest, though not necessarily with respect to the whole conformation space. Since it is tightly coupled to the ability of a protein to fold, the existence of a folding funnel for a given protein depends on its primary structure. An amino acid sequence having a single sizeable folding funnel leading to a unique, stable conformation is said to be "foldable." Conversely, "non-foldable" amino acid sequences have multiple folding funnels and do not exhibit a single minimum.

Three major simplifications are required to describe the long–time kinetic behavior of a protein. First, a lattice model is used to restrict discussion to the most important degrees of conformational freedom. Each residue is located on a vertex of the lattice. Second, consistent with the low resolution structural description of the lattice model, the chemical identity of the protein is depicted in a simplified N–letter hydropathic code, where some number N of amino acid types is chosen to enable an arbitrary degree of heterogeneity for residue–residue interactions. Each residue may represent clusters of real amino acids. Finally, the dynamics of the folding process can be treated as a collection of discrete "hopping" processes between local minima. Since conformational dynamics is governed by the same residue–residue interactions

that drive the initial conformational collapse, it is possible to divide motion into two types: relatively fast, sterically–constrained chain motions that do not involve the breaking of many chain–chain contacts, and slow motions across energy barriers associated with larger structural fluctuations and the breaking of many chain–chain contacts.

The time scale separation proposed for the interconversion process can be used to divide collapsed conformations into "compactness cells." Within a compactness cell, conformations interconvert quickly; between compactness cells, conformations interconvert slowly. The rates for slow interconversion are found by calculating the average time the protein spends in any cell before a thermal fluctuation moves the chain over a conformational barrier into another cell. The division of conformation space into compactness cells is valid for any model for protein structure, whether on– or off–lattice. A difficulty arises when attempting to define compactness cells for specific sequences. To show how such difficulties can be managed through an approximation technique, we now focus on a simple lattice model.

Statement of Model

Lattice Model and Energy Function. A protein is represented as a 27-unit chain on a simple cubic lattice. The protein is assigned a chemical identity in the form of a primary sequence $S = \{s_1, \ldots, s_{27}\}$ where each site i on the chain is assigned a residue from a list of N residue types, $s_i \in \{a_1, \ldots, a_N\}$. The residue types do not correspond directly to the naturally occurring amino acids but are chosen to provide

a wide range of chemical differentiation. It is not clear how much heterogeneity is necessary for the careful organization implicit in the process of folding, but in the present case $N = 20$, allowing a wide range of residue contact energies.

Folding is driven by nonbonded contact energies, interactions between residues that inhabit adjacent lattice sites but are not covalently bonded to each other. The energy E of a given conformation is calculated by summing the values of the energies over all nonbonded contacts in the lattice:

$$E = \sum_{i,j}^{27} \epsilon(s_i, s_j)\ \Delta(\vec{r}_i - \vec{r}_j) \tag{1}$$

where r_i and r_j denote the locations of residues i and j and $\Delta(\vec{r}_i - \vec{r}_j) = 0$ unless residues i and j are on adjacent vertices of the lattice. Contact energies $\epsilon(a_i, a_j)$ between residue types a_i and a_j are chosen to have an average of $\sim -2k_BT$ with an effective deviation of about k_BT. Native nonbonded contacts — those contacts observed in the unique, stable, accessible native state — are among the most stable of all nonbonded contacts with an average energy of $\sim -3k_BT$.

Compactness cells. Folding begins with a structural collapse from a random coil state to a dense state. Here "dense" denotes a conformation with few interior vacant lattice sites. While the number of dense states is far smaller than the number of random coil states, dense conformations are still too numerous to determine computationally. However, the subset of dense conformations with no internal vacancies — the so–called "maximally–compact conformations" – - is countable. For the 27-mer (Figure 2), there are 103,346 geometrically–distinct, maximally–compact conformations, each shaped like a 3×3×3 cube [6,7].

The maximally–compact cubes perform a two-fold function in describing interconversion. First, they act as markers among the dense conformations. Each dense conformation is assumed to be a structural fluctuation around some specific cubic conformation. Second, maximally-compact conformations are strong candidates for local energy minima. Energy barriers between maximally-compact conformations are large because many contacts must be broken in fluctuating from one conformation to another. Taken together, the maximally compact local minimum conformation and its dense affiliates constitute a "compactness cell."

Folding process. The picture of folding that emerges can be stated simply. A collapsed conformation undergoes structural fluctuations among a series of dense states en route to the native state. At each moment, the conformation is assigned to the nearest maximally–compact conformation. Folding is the process of diffusion between distinct compactness cells and can be measured in terms of the number of native contacts that exist at any given time. Since each compactness cell contains a maximally–compact conformation, folding may be viewed in lower resolution as discrete hopping between maximally–compact states (Figure 1).

Here arises the central question: If we can understand the diffusional relationships between the lowest energy maximally– compact conformations, can we predict if a given amino acid sequence is foldable? First, it is necessary to examine the relationship between geometry and interconversion among the maximally–compact conformations.

Reconfigurational distance. Two maximally–compact conformations will interconvert through small structural fluctuations if the conformations are geometrically

similar. The greater the geometrical dissimilarity between the conformations, the larger the structural fluctuation required for interconversion to occur. Thus, it is useful to introduce a quantitative measure of geometrical distance between two cubic conformations. We define a reconfigurational metric to represent the minimum number of residues that need to move in order for one conformation to fluctuate into the other given a set of fundamental lattice moves. In many cases, interconversion between conformations requires the motion of residues solely to "get out of the way" of other residues. This definition anticipates the kinetic problem to be faced. The reconfiguration distance satisfies the criteria of a metric, *viz.*, $\rho(\alpha,\beta) = \rho(\beta,\alpha)$, $\rho(\alpha,\alpha) = 0$ and $\rho(\alpha,\beta) \geq 0$, where α and β are two arbitrary conformations. The reconfigurational distance between maximally–compact conformations α and β is used to determine which pairs of low energy maximally-compact conformations are likely to interconvert.

Sequence–structure design criteria. Sequences S1 and S2 (Table 1) were selected from a 20-letter hydropathic residue set to demonstrate foldability and nonfoldability, respectively. (Parameter set $\epsilon(a_i,a_j)$ is available upon request.) S1 was selected to stabilize a particular conformation (Figure 2), which was chosen because 30 other maximally–compact conformations were within a reconfiguration distance of 10. The native structure of S1 exhibits two 2×2×2 subdomains in opposite corners of the cube. Each subdomain has four possible maximally–compact arrangements, so the native structure possesses a high proclivity to local geometrical rearrangement among the 15 neighboring cubic conformations. S1 was picked to stabilize *not only* the native structure but also the local conformational permutations of the subdomains. These other

"nearby" maximally–compact conformations are among the most stable of the overall conformation space, but they are sufficiently unstable with respect to the native state as to have negligible steady–state populations, confirming the two-state criterion imposed by calorimetric studies of folding [8]. S1 contains information for *both* the "local" thermodynamic stability of the native structure and its conformational proximity to several other stable states.

Sequence S2 is a random set of monomers drawn from the same 20-letter set. Although no structural or reconfigurational criteria were used to determine S2, it does have a distinct global minimum. Thermodynamically, it possess a unique structure, but it will be shown that this structure is kinetically inaccessible.

Results and Discussion

Lattice kinetics: simulations. The kinetic behavior of sequences S1 and S2 was simulated at room temperature using the Metropolis algorithm. During the simulation, single residues on the chain were randomly moved to nearby vacant lattice sites with the constraint that the covalent chain bonds remain intact. In addition, pairs of residues were allowed to move in single 90° rotary motions.

Starting from a random conformation, sequence S1 was shown to achieve its global minimum conformation in 81 of 200 trials each lasting up to 500,000 iterations. In only 14 cases, the final conformation had less than half of the native nonbonded contacts. In all other cases, it is supposed that longer simulations would have succeeded in folding the recalcitrant intermediates. S2 failed to achieve its global

minimum conformation even once during 10 trials with 500,000 iterations. None of the final conformations had more than 14 nonbonded contacts in common with the lowest energy structure. Thus, in a time frame $\leq$500,000 iterations, S1 has a unique, stable, accessible conformation and sequence S2 does not.

In addition to proving that S1 is a foldable sequence, the simulations validate the collapse–reconfiguration picture of folding. The number of chain–chain nonbonded contacts and native nonbonded contacts are plotted versus iteration number for a simulation in Figure 3. In the first 1000 iterations, the collapse phase is evident as total number of nonbonded contacts increases quickly. In the subsequent reconfiguration phase, (34000 iteration in this example) the total number of nonbonded contacts fluctuates as the number of native nonbonded contacts increases. Maximally–compact nonnative intermediates occur during a large fraction of this time. When the native state occurs after about 35000 iterations, it remains stable with minor fluctuations until the trial ends after 5 million iterations.

Lattice kinetics: interconversion calculations. The temporal dominance of the reconfiguration phase evident in Figure 3 suggests that during the slow stage of folding all motions occur as interconversion between collapsed conformations. It is possible to simplify the complex set of trajectories onto a discrete number of compactness cells.

Before the kinetic map of compact conformation space can be drawn, individual rates between neighboring compactness cells must be calculated. These rates are determined through a probability flow equation formulated for the microscopic motion of the chain (the master equation). Using pairs of maximally–compact conforma-

tions as the starting and ending points — or reactant and product pairs — diffusion between compactness cells is calculated assuming the same microscopic kinetic rules used by the lattice kinetic simulation. The intermediate states are given by explicit local conformations of the chain as it fluctuates between reactant and product conformations.

The intermediate states are determined through an exhaustive enumeration of "loop structures" (Figure 4). Loops consisting of fewer than 10 residues and 500 intermediate loop structures are considered. Longer loops are ignored because global interconversions are not thought to participate in compact reconfiguration. Loop intermediates with more nonbonded contacts in common with the reactant's maximally–compact conformation are assigned to the reactant's compactness cell. Similarly, the product's compactness cell consists of loop intermediates containing a majority of product cube's nonbonded contacts. The initial distribution of probability in the reactant compactness cell is chosen to obey a thermal distributution. When half of the probability seeps over the barrier into the reactant compactness cell, the average time for interconversion has elapsed. Details of the calculation for average reaction times are given in the Appendix.

Folding funnels. For sequences S1 and S2, rates between pairs of the 1000 lowest energy maximally–compact conformations were calculated. In Figure 5, the maximally–compact conformations are represented by dots organized by energy along the y-axis and listed by arbitrary conformer number along the x-axis. The global minimum is conformer number 1. Interconversion times between these conformations are $\leq$1 ms. Solid (dashed) lines between the maximally–compact conformers indicate the likeli-

hood for interconversion from the higher (lower) energy to the lower (higher) energy conformer.

A single deep folding funnel dominates the interconversion diagram for S1. For S2, a myriad of smaller and shallower folding funnels appears. Thus sequence S1 will fold and S2 will not. These predictions are consistent with the simulation. The lowest energy conformation of S1 exhibits structural and energetic similarity among the several other most stable conformations. A close comparison of the lattice kinetics simulation (Figure 3) and the folding funnel calculation (Figure 5A) shows that simulated folding trajectories diffuse down the folding funnel (Figure 6). The longest dwell periods occur in conformations relatively close to the native conformation. For sequence S2, however, the "nearest neighbor cubes" are of substantially higher energy (Figure 5B) so the conformational energy landscape appears more like a volcano than a funnel.

Perspective and Concluding Remarks

The existence of folding funnels in simple lattice models of folding suggests that directed self-assembly of proteins can be predicted using enumeration as well as simulation techniques. The occurrence of folding funnels contradicts previous notions that folding must follow a specific pathway to overcome the Levinthal counting problem [9]. Rather, folding may well proceed as a convergence of multiple pathways among families of interconvertible dense conformations, as suggested by the jigsaw model [10].

This work is related to several perspectives on folding that have been put forth in recent years [4,6,11,12,13,14]. In a mean field spin model of protein folding, Bryngelson & Wolynes showed that, during the last stages of folding, conformational energy replaces conformational entropy as the dominant repository of free energy [15,16] To give their mean field spin model a kinetic interpretation, they mandated that every spin state or "conformation" of the protein be kinetically connected to a state of greater polarization or "folding," which they called the principle of minimal frustration [15]. The kinetic interpretation of minimal frustration applied to polymers is not obvious. Here, we submit that minimal frustration of proteins is defined by the existence of a folding funnel: an amino acid sequence has a large number of dense conformations which are kinetically connected to conformations of greater folding.

Recently, Shakhnovich, *et al*, [17] reported that three of 33 randomly interacting heteropolymers of length 27 with stable global minima were kinetically foldable. The parameter regime for nonbonded contact energies used here is identical; however, our perspective differs considerably on the issue of sequence specificity: we present two amino acid sequences based on a 20–letter hydropathic code — the foldable sequence possesses relationships between sequence and structure, while the nonfoldable sequence exhibits no such concordance.

Folding funnels contain the key thermodynamic and kinetic requirements for folding: convergence of families of diffusion pathways to a unique , locally-stable, kinetically–accessible state. With further development of the reconfigurational metric, effective extension of these methods to larger proteins heretofore treatable only by simulations [18] is foreseeable.

Appendix

The kinetics of structural fluctuation between compactness cells is given by a continuous–time master equation in which each microscopic chain state of the interconverting segment of the loop is treated as an independent probability site. Rates for interconversion between compactness cells are calculated by designating the product cell's chain sites as absorbing sites. When the chain diffuses into an absorbing site, the interconversion reaction is said to have occurred.

The rate is given by the inverse of the mean first passage time for absorption into the product states. The quantity is derived from the master equation:

$$\frac{\partial \vec{p}(t)}{\partial t} = W\vec{p}(t) \qquad (A1)$$

where $\vec{p}(t)$ is an n–dimensional time–dependent probability vector ($\sum p_n = 1$) and W is the $n \times n$ transition matrix satisfying detailed balance. The microscopic transition rates $W_{\alpha,\beta}$'s from structure β to α are assumed to be dominated by the energy difference $\Delta E = E_\beta - E_\alpha$.

$$W_{\alpha,\beta} = \begin{cases} R_0 \exp\left(\frac{-\Delta E}{k_B T}\right) \Theta\left(\sigma_0 - \sigma(\alpha,\beta)\right), & \text{if } \Delta E \geq 0; \\ R_0, & \text{otherwise.} \end{cases} \qquad (A2)$$

The function $\sigma(\alpha,\beta)$ denotes the sum of the squares of the differences between α and β coordinates, and $\Theta\left(\sigma_0 - \sigma(\alpha,\beta)\right)$ is a step function that sets the matrix element $W_{\alpha,\beta} = 0$ when $\sigma(\alpha,\beta)$ is larger than a cutoff $\sigma_0 = 4$. Θ requires structural transitions to be local. The master equation is designed to mimic the restricted set of "moves" allowed in lattice dynamics simulations. The constant R_0 imposes an arbitrary time scale and is chosen to be 10^{11} sec^{-1} for the numerical examples cited.

Acknowledgements

We thank H.S. Chan, K.A. Dill, D.G. Covell, B. Gerstman, A. Grove, E. Shakhnovich, J. Skolnick, P.G. Wolynes and B.H. Zimm for discussions, D. Beratan and D. Fredkin for providing computational resources and D. Fredkin for analytical and numerical advice on solving the master equation. PEL was supported by a training grant from NIMH and partially by NSF #MCB-9018768. JNO is a Beckman Young Investigator. JNO is supported by the NSF #DMB-9018768. MM is supported by ONR #N00014- 89-J1469, NIMH #44638 and #00778, NIGMS #42340 and DAMD #17-89- C-9032.

S1:	j e d f q g i c n m h l q o p r s j c k t b t b a r s
S2:	a e a b a j a m p i q r s n a j k b p s o r q t l f i

Table 1 Foldable and nonfoldable primary sequences

References

1. Anfinsen, C.B., Haber, E., Sela, M., & White, F.H,(1961) *Proc. Natl. Acad. Sci. USA* **47**, 1309-1314.
2. Levinthal, C. (1968) *J. Chim. Phys.* **65**, 44-45.
3. Levinthal, C. (1969) in *Mossbauer Spectroscopy in Biological Systems*, eds. Debrunner, P., Tsibris, J.C.M., & Munck, E. (University of Illinois Press, Urbana), pp. 22-24.
4. Dill, K.A. (1985) *Biochemistry* **24**, 1501-1509.
5. Zwanzig, R., Szabo, A., and Bagchi, B. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 20-22.
6. Chan, H.S. & Dill, K.A. (1990) *J. Chem. Phys.* **92**, 3118-3135.
7. Shakhnovich, E.I. & Gutin, A.M. (1990) *J. Chem. Phys.* **93**, 5967-5971.
8. Privalov, P.L. & Gill, S.J., P (1988) *Adv. Prot. Chem.* **39**, 191-231.
9. Baldwin, R.L. (1990) *Nature* **346**, 409.
10. Harrison, S.C & Durbin, R. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 4028-4030.
11. Ueda, Y., Taketomi, H., G, N. (1978) *Biopolymers* **17**, 1669-1678.
12. Miller, R., Danko, C.A., Fasolka, M.J., Balazs, A.C., Chan, H.S., & Dill, K.A. (1992) *J. Chem. Phys.* **96**, 768-780.
13. Shakhnovich, E.I. & Gutin, A.M. (1990) *Nature* **346**, 773-775.
14. Covell, D.G. & Jernigan, R.L. (1990) *Biochemistry* **29**, 3287-3294.
15. Bryngelson, J.D. & Wolynes, P.G. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 7524-7528.

16. Bryngelson, J.D. & Wolynes, P.G. (1989) *J. Phys. Chem.* **93**, 6902-6915.

17. Shakhnovich, E., Farztdinov, G., Gutin, A.M., & Karplus, M. (1991) *Phys. Rev. Lett.* **67**, 1665-1668.

18. Skolnick, J. & Kolinski, A. (1990) *Science* **250**, 1121-1125.

Figure Captions

Figure 1. A schematic representation of the folding process. The denatured coil (A) collapses to a random dense structure (B), approximated by a set of maximally compact conformers (C). A reconfigurational distance is defined between compact states (D) and is used to sort pairs of cubes for calculation of the mean first passage time for interconversion (E). The kinetic structure of conformation space (F) shows a folding funnel leading to a unique, locally–stable, kinetically–accessible state.

Figure 2. The conformation of the native state of sequence S1 represented in a four–letter hydropathic code (for calculations a 20–letter code was used). Two "sub–domains" are present as eight- -residue 2×2×2 subcubes in the upper left octant and in the lower right octant. Sub–domains enhance kinetic accessibility by providing a large number of dense conformations that are geometrically similar to the native state.

Figure 3. The collapse (upper trace) and reconfiguration (lower trace) of the sequence S1 during a 35000–iteration lattice kinetics simulation. The protein collapses quickly, establishing most of the 28 possible nonbonded contacts in the first few thousand iterations. Folding, which is measured in terms of native nonbonded contacts, occurs more slowly. Just before the protein folds a fluctuation occurs, breaking all of the incorrect nonbonded contacts. At that point, the collapse and folding curves merge.

Figure 4. Interconversion between two maximally–compact conformations requires the enumeration of all microscopic loop structures. Here, 48 loop conformations of an 8–mer segment of the native conformation of S1 (Figure 2, residues 3 to 10) are explicitly drawn. Boxed conformers represent loop packings for the native state of Figure 2 and another maximally–compact conformation at a reconfigurational distance of 7. All other conformers are local fluctuations around these two "cubic" states. The shaded domain corresponds to the compactness cell of one cubic conformer. All other conformers constitute the compactness cell of the native loop conformer. The interconversion calculation is performed by permitting probability to diffuse along all possible connections between and within the compactness cells. Starting from a thermal distribution of probability over each compactness cell, the mean first passage time is defined as the period in which half of the probability seeps from one compactness cell to the other.

Figure 5. Folding funnels for the foldable sequence S1 (A) and nonfoldable sequence S2 (B). Line segments indicate submillisecond connections kinetic pathways between compactness cells (dots) associated with maximally-compact conformations. Solid (dashed) lines correspond to diffusion from higher (lower) to lower (higher) energy conformations. Arrows indicate the lowest energy maximally compact conformations. The folding funnel of S1 is evident in the convergence of kinetic pathways to the maximally–compact conformations with energy -81 k_BT. S2 shows no such self-organizing property. The reconfigurational distances between conformations in the

S1 folding funnel and the S1 lowest energy structure are ≤ 14. The reconfigurational distances between the low energy S2 conformations and its lowest energy structure are > 20.

Figure 6. The final steps of the folding trajectory of Figure 3 are mapped onto the compactness cells (large dots a-g) from the funnel diagram from Figure 5A. The simulated diffusive trajectory passes between these compactness cells in the order: (b,d)-d-(a,e,g)-(e,g)-e-g, where the parentheses indicate when particular noncompact conformations are assigned to a few compactness cells simultaneously. Conformations c and $\bar{f}$ are seen in the final stages of other simulations. The simulation results substantiate the predictions of the interconversion calculations that diffusive reconfiguration "down the funnel" is the mechanism of folding. The reconfigurational distances between the funnel conformations and its lowest structure are shown inside of the circles.

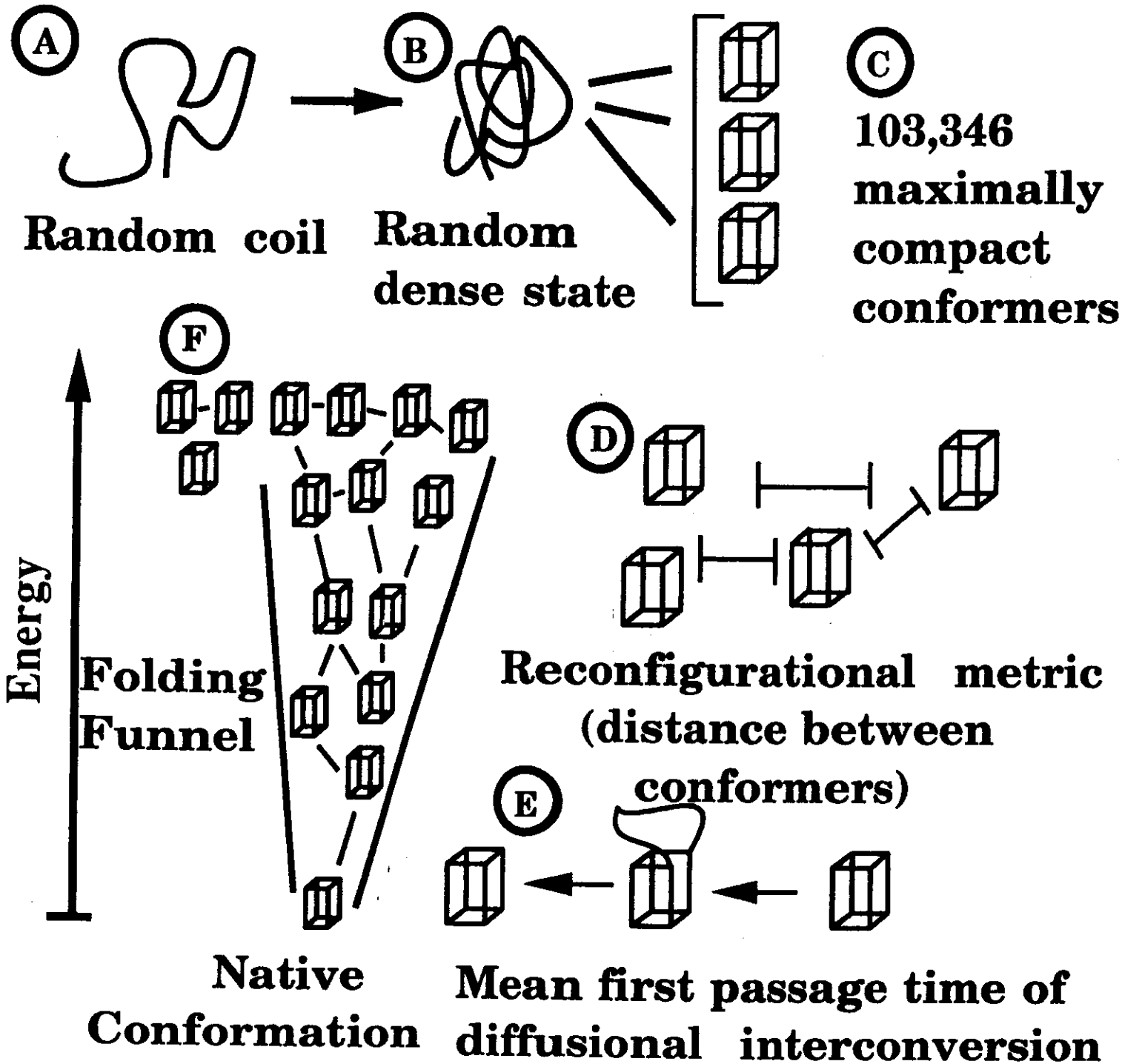
A
B
C
Random coil
Random dense state
103,346 maximally compact conformers
F
D
Energy
Folding Funnel
Reconfigurational metric (distance between conformers)
E
Native Conformation
Mean first passage time of diffusional interconversion

FIGURE 1

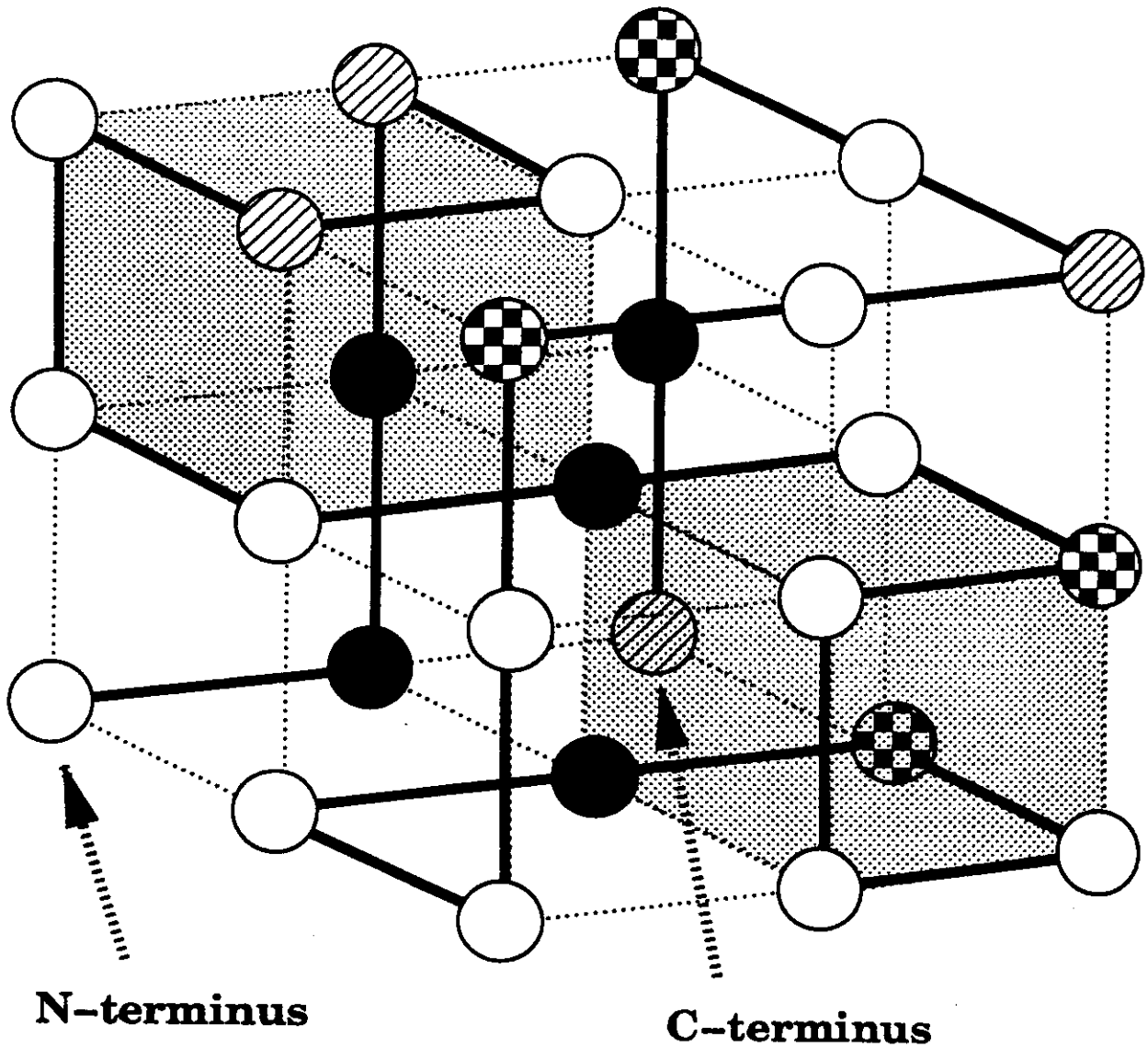

FIGURE 2

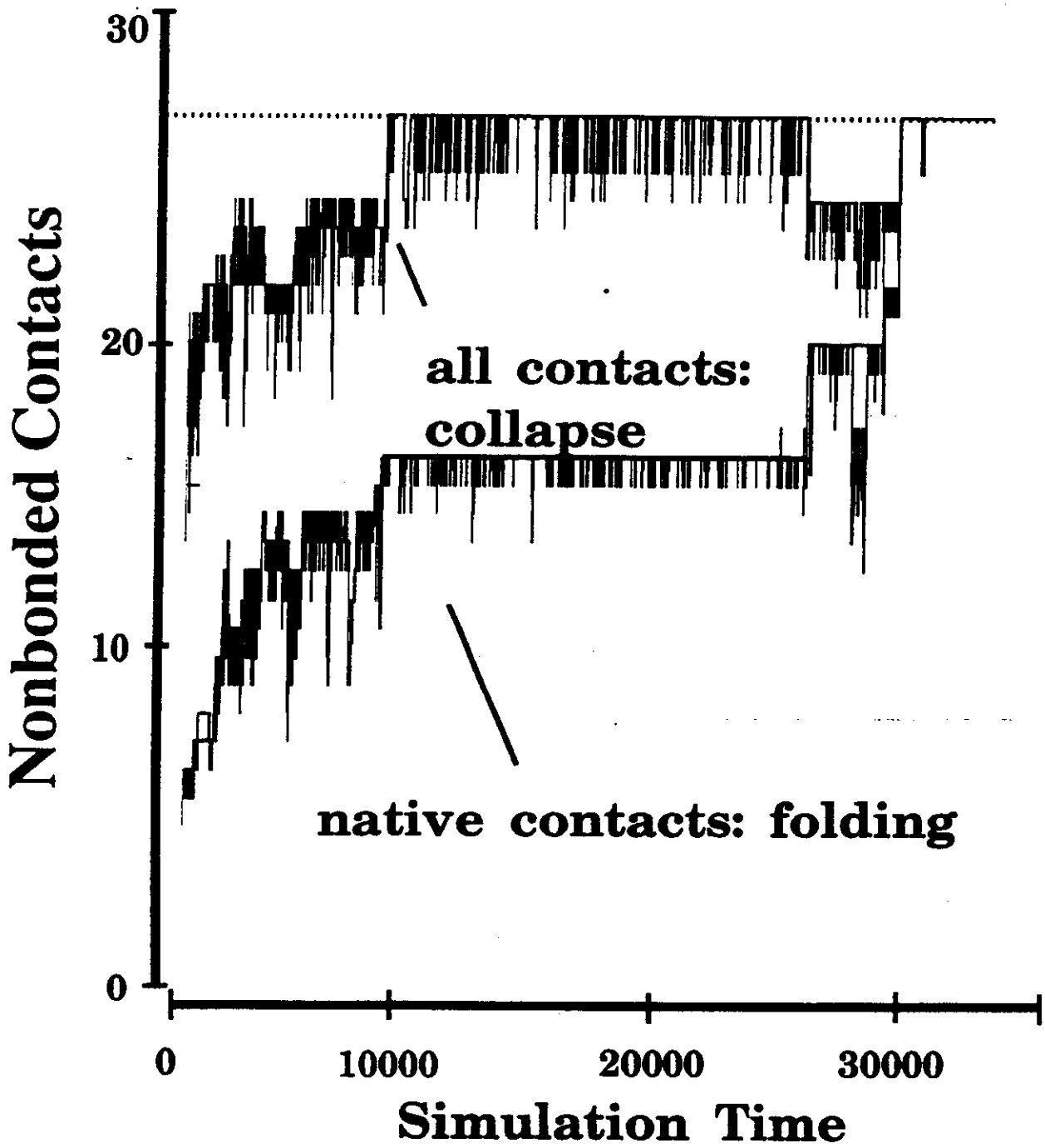
30
20
10
0
Nonbonded Contacts
all contacts:
collapse
native contacts: folding
0
10000
20000
30000
Simulation Time

FIGURE 3

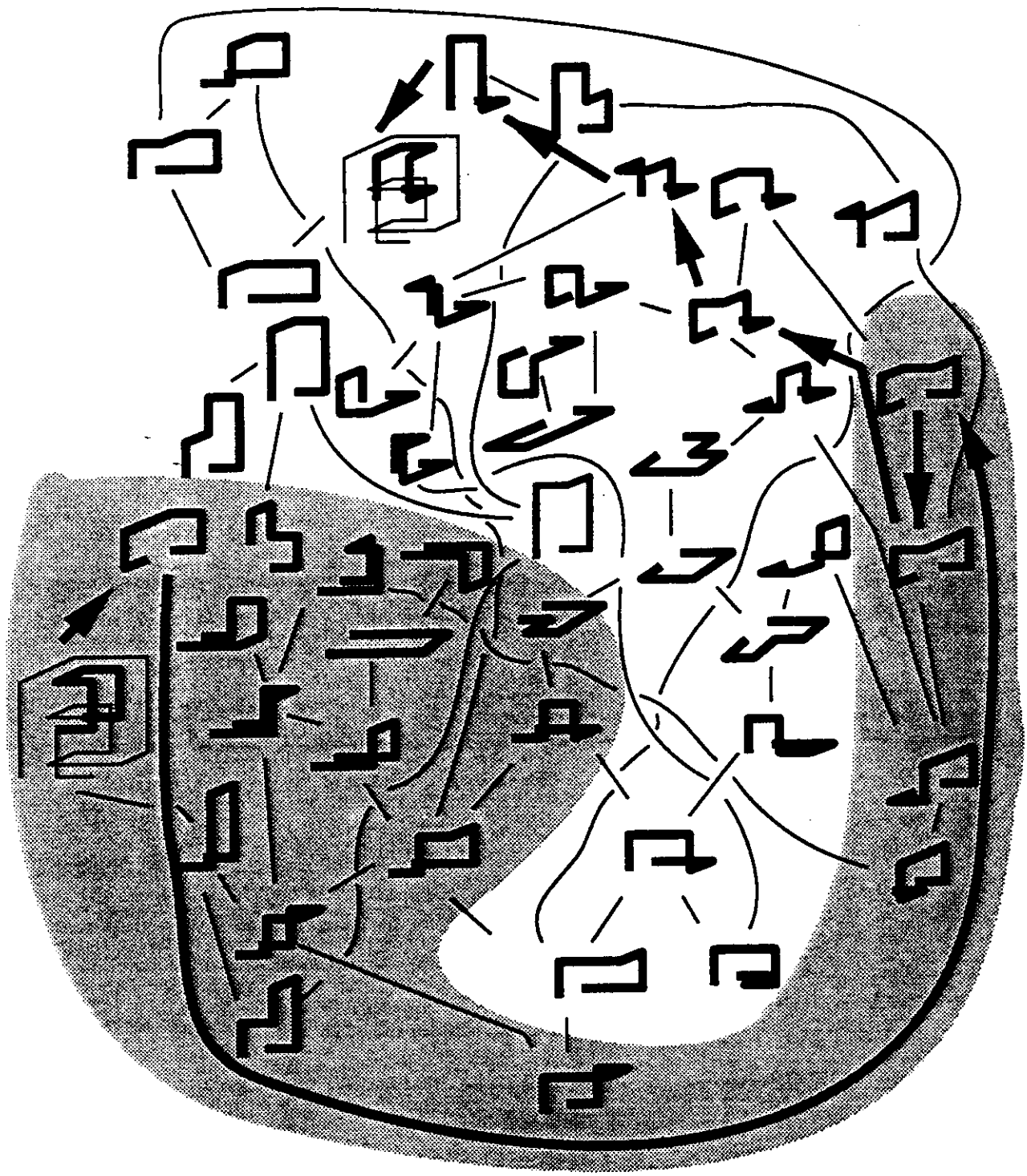

FIGURE 4

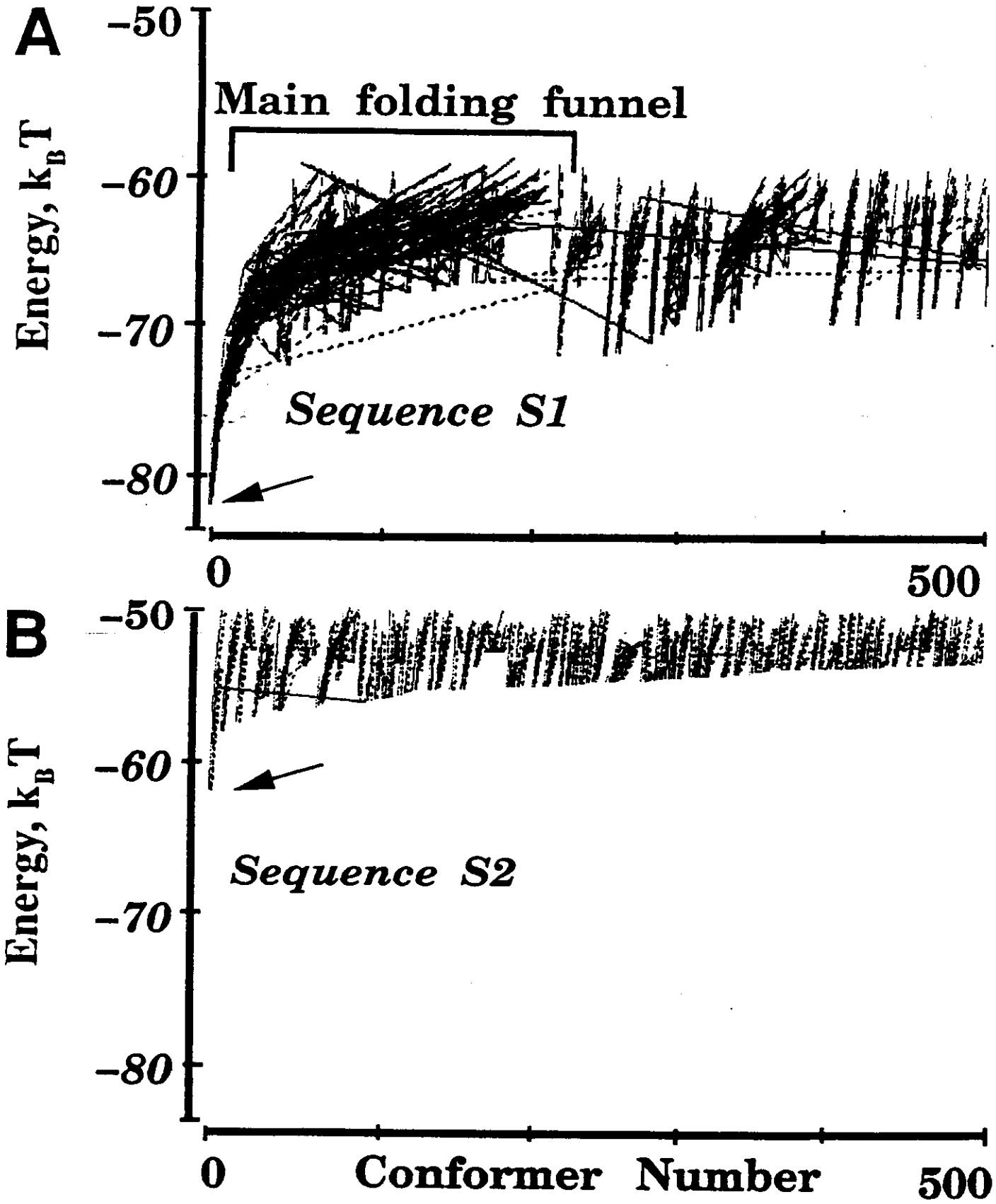
A
-50
Main folding funnel
-60
Energy, $k_B T$
-70
Sequence S1
-80
0
500
B
-50
-60
Energy, $k_B T$
Sequence S2
-70
-80
0
Conformer Number
500

FIGURE 5

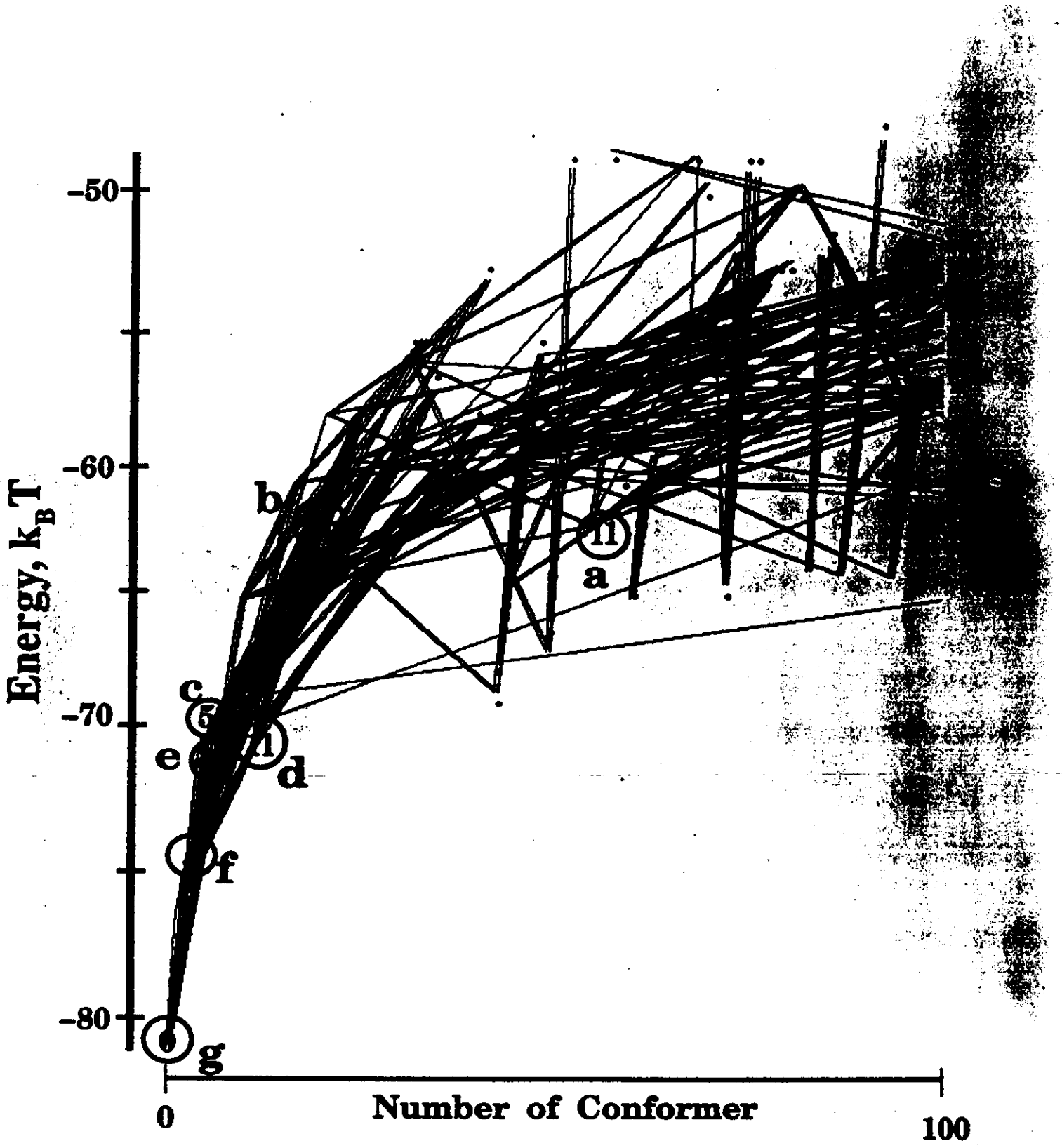

FIGURE 6