J. Indian Inst. Sci., July–August 1997, **77**, 339–352.

Design of rapidly folding protein-like heteropolymer chains and the cell dynamics: A lattice model study

SARASWATHI VISHVESHWARA*, INDIRA SHRIVASTAVA**, MAREK CIEPLAK*** AND JAYANTH R. BANAVAR****

*Molecular Biophysics Unit,Indian Institute of Science,Bangalore 560012,India;
**Molecular Biophysics Unit,Indian Institute of Science,Bangalore560012,India
(Present address: European Molecular Biology Laboratory, Meyerhefstr 1, 69012 Heidelberg, Germany)
***Institute of Physics,Polish Academy of Sciences,02-668 Warsaw,Poland;
****Department of Physics and Center for Materials Physics,104 Davey Laboratory,The Pennsylvania State University,University Park,PA 16802.

1. Introduction

Understanding the mechanism of protein folding is a fundamental problem in molecular Biology[1]. Key issues include the prediction of three-dimensional protein structure from one-dimensional sequence information and the inverse problem of designing sequences that fold into a desired three-dimensional structure. One of the puzzles of the dynamics of protein folding , called the Levinthal paradox[2], is the rapidity with which a polypeptide chain folds into the native state, even though an exhaustive search is ruled out due to the enormous number of all possible conformations. More generally, the dynamics of protein folding is one of the class of minimization of a fitness function characterized by complex structure and many local minima.

It is well known that lattice models, in spite of their simplicity, capture many of the features of protein folding. Using the computational facility at SERC, IISC, we have designed rapidly folding protein like heteropolymers in a 3-D lattice[3]. Secondly, the folding pathway is elucidated by cell dynamics in 2-D lattice[4]. The details of these investigations are presented in this paper.

2. Design of a rapid folder

A heteropolymer chain can be specified by the nature of the constituent monomers. In some of the lattice model studies, however, it is specified as interaction (contact) energy between a pair of constituent monomers. The interaction energies are generally derived from protein data analysis. A 27-residue polymer can have 156 interactions, excluding the two nearest neighbours. Generally, the interaction energies are attractive. However, a small percentage of interactions can also be repulsive. Because of attractive energies between the monomer units, a heteropolymer chain collapses to a compact form. When all

the monomers of a 27-residue polymer occupy the points on a $3 \times 3 \times 3$ cubic lattice, the polymer conformation is said to be maximally compact and will have 28 contacts. There are 1,03,346 unique maximally compact conformations on a cubic lattice and the energy of each conformation is the sum of 28 contact energies. The conformation with the lowest energy is termed as the native state and the contacts found in this state are native contacts. Those contacts which are not seen in the native conformation are non-native contacts. Rapid folding is measured in terms of foldicity, which is the fraction of the number of times a random polymer chain folds to its native state within a short specified time to the number of times attempted.

A simple model of a 27-bead self-avoiding chain on a cubic lattice is considered in the present study. To begin with, we use the overall attractive contact energies (similar to those of Miyazawa and Jernigan[5], denoted as *Bij*), similar to that of Sali, Shakhnovich and Karplus (SSK)[6,7]. The chosen values represent a gaussian distribution with a mean value (*B*0) around –2 with a heterogeneity factor (sigma) of 1. The conclusion from the SSK study is that as long as the native state is a pronounced global minimum on the potential surface and the temperature is high enough to overcome the barriers between local minima, the ground state is reached within 50 million time steps in a Monte Carlo (MC) simulation and the protein is said to have folded. The key issue is whether the pronounced global minimum is a necessary and sufficient condition for good foldicity.

A protein contains 20–25% of acidic and basic groups[8] that are ionized or protonated under physiological conditions. In compact conformations of heteropolymers, the repulsion between like charges, geometrical effects of excluded volume or an inability to satisfy all the hydrophobic and hydrophilic interactions simultaneously can give rise to frustrations. Wolynes and collaborators[9,10–13] have shown that the principle of minimal frustration distinguishes between natural proteins and random heteropolymers – in proteins, side chains contribute coherently to supersecondary structure and tertiary folds[14]. This principle leads naturally to a large stability gap, *i.e.*, a measure of the energy gap between the state with a structure similar to that of the native state and the lowest energy state among those that bear little resemblance to the native structure. In simple situations, the stability gap may be correlated with the energy gap[6,7] between the native and the first excited state. Goldstein *et al.*[12,13] have used the minimization of the stability gap of proteins of known sequences and structures as a means of determining the optimal interactions between amino acids and have successfully predicted folding structures for new sequences.

We have used an idea similar to the notion of strong disorder in spin glasses[15,16] to design protein sequences that are strongly folding[3]. When the Ising spin glasses are widely separated, frustration, or the inability of a spin to satisfy all the exchange interactions of its neighbours simultaneously, while present is irrelevant and the nontrivial ground state of the spin glass can be obtained trivially. Operationally, this is achieved by rank ordering the exchange interactions B_{ij}, in decreasing magnitude and arranging the mutual spin orientation to satisfy as many of these as possible in order of decreasing strength. This rank ordering idea is used in our protein design. However, a wide separation of monomer interaction is not required.

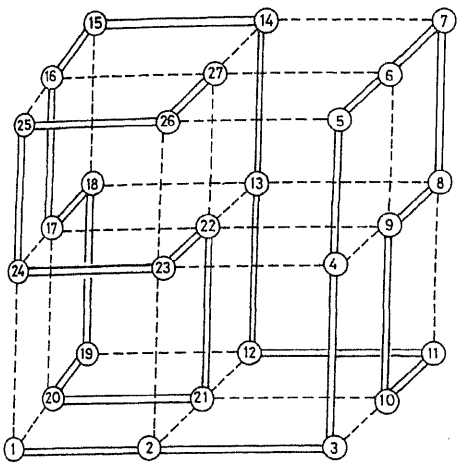


FIG. 1. An example of a compact conformation of a 27-bead polymer in a $3 \times 3 \times 3$ cubic lattice.

Our design procedure begins with the selection of a compact conformation as the target structure of the folded protein, such as the one given in Figure 1. The 156 B_{ij} values generated from a gaussian distribution similar to Miyazawa and Jernigan[5] contact energies are rank ordered in decreasing order of magnitude and the first 28 values are randomly assigned to the 28 native contacts and the remaining 128 values randomly to the non-native contacts of the chosen target native state. This ensures that the target structure is the ground state and also leads naturally to a large gap between the ground state and the first excited state. Our MC[3] folding simulations (identical to that carried out by SSK[6,7]) showed that the mean foldicity of the designed proteins is high (0.95) compared to the 0.13 obtained by SSK for randomly chosen sequences . Thus, our choice of maximizing the compatibility of the interacting monomers effectively ensures the stability of the native state over a range of temperatures and allows for kinetic accessibility of this state.

The situation when some of the B_{ij} values are repulsive cannot be investigated within the framework of the original SSK study[6,7], since the native state of randomly chosen contact energies of either sign (attractive or repulsive) is neither necessary nor likely to be maximally compact. Since an exhaustive search of only the maximally compact conformations is feasible, it is not possible to identify the ground state of the chain. Our simple idea of protein design by assigning the best interaction energies to the native contacts ensures that the ground state is maximally compact and known. Another set of MC folding simulations were carried out with interactions randomly attractive or repulsive with equal probabilities (chosen for simplicity). Strikingly, the measure of foldicity is 1 and the folding is much faster than in the purely attractive case[3].

Our previous work demonstrated that, as a general rule, the sequences fold as the native state energy is increased and fold rapidly when partially repulsive non-native interactions are present in the sequence. For all best interactions to be in the native state, however, is an extreme case, which may not be the situation in real proteins. Hence, in the pre-

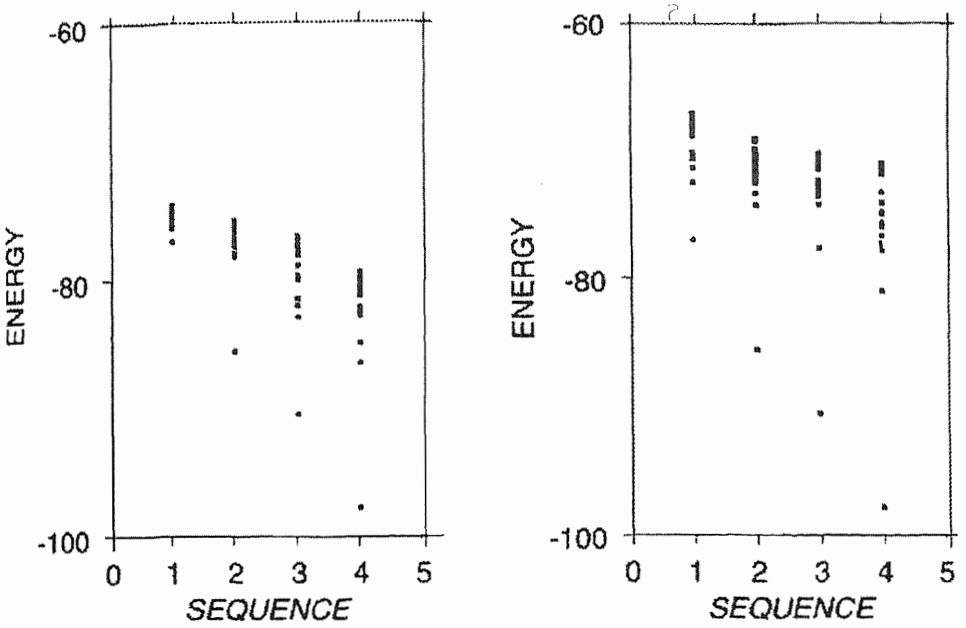
sent study we have designed sequences which fold to the same conformation with varying native state energy. This is achieved as follows. We begin with an SSK[6] type of sequence. Then its native state and the native state interactions are identified. The number of rank-ordered interactions (B_{ij}'s) in the native state are evaluated and other sequences are designed by increasing the rank-ordered B_{ij}'s in the native state and the end member of such a design will have 28 rank-ordered native B_{ij}'s as in our previous case[3]. The native states of all the designed sequences are determined from the energy spectrum and it is ensured to be the same as that of the original chosen random sequence. Further, partially repulsive sequences are designed by introducing repulsive B_{ij}'s in 20% of the non-native interactions in the above sequences. This procedure of rearranging the interaction energies and replacing the non native contacts by repulsive terms gives us sequences with the same native state but with varying degree of stability and conformational entropy.

A set of sequences corresponding to the native state given in Figure 1, were designed as described above. These sequences were studied in detail for their conformational and thermodynamic properties. The results thus obtained are presented in Table 1. As expected, the energy of the native state increases as the number of rank-ordered native contacts increase (from 13 to 28) and remains the same when 20% repulsion is introduced in the non-native contacts. The energy gap between the native and the first excited state, however, changes significantly. The gap increases as the native state energy increases and the partially repulsive case also leads to an increased gap. The number of common contacts (*N*cc) between the native and the first excited state is different in different sequences,

Table 1
Conformational and thermodynamic data for the sequences of native structure A (Figure 1).

Sequence	NRc	E0	Gap	*N*Cc	*Tf*	PC
1	13	–76.997	1.706	9	0.647	0.44
2	16	–85.551	7.316	20	1.875	0.52
3	20	–90.469	7.549	15	2.322	0.55
4	28	–97.822	11.368	22	3.042	0.62
1R*	13	–76.997	2.044	11	1.969	0.44
2R	16	–85.551	11.303	20	2.971	0.52
3R	20	–90.469	12.913	20	3.460	0.55
4R	28	–97.822	16.880	20	4.197	0.62

R* – Indicates that 20% of the non-native contacts are repulsive with the native interation being the same as in the chosen sequence.
*N*Rc - Number of native contacts whose B_{ij} values are from amongst the highest 28 b_{ij} values.
*E*0 - Native state energy.
Gap - The gap denotes the energy difference between the native state and the first excited state among the compact self avoiding (CSA) conformations.
*N*Cc - Number of common contacts between the native and the first excited state.
Tf - Is the temperature at which the probability of native state among CSA conformations is 0.5.
PC - Denotes the Pearson correlation coefficient between the interaction matrix and the native contact map.



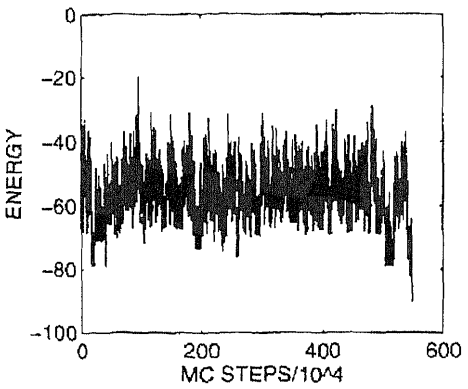
G. 2. Energy spectra of self-avoiding maximally compact conformations for the 4 sequences with different ıtive state energy (Table 1) of the native state A (Figure 1) for the purely attractive case (A) and for partially pulsive case (B). First 20 lowest energies are shown.

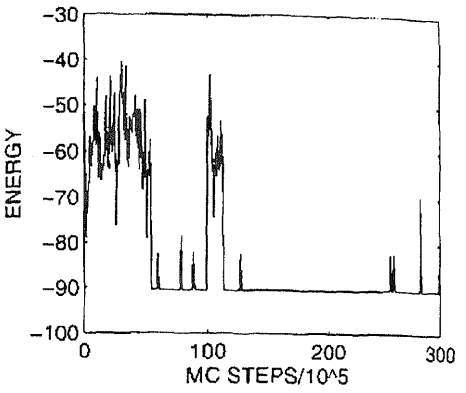
ıdicating different excited state conformations. It is interesting to note that Ncc generally ıcreases as the native state energy increases, which means that the first excited confor- ıation resembles the native conformation more as the stability of native state increases.

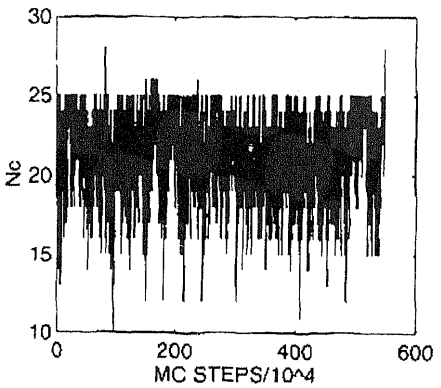
The increase in gap with increased native state energy is a result of change in the sta- ility of the native and the first excited states. The lowest 20 states of the sequences given ı Table 1 are plotted in Figure 2. The effect of increased native state energy with in- reased rank-order native B_{ij} and the reduced conformational space in the partially repul- ive case on the gap value can be seen from this plot.

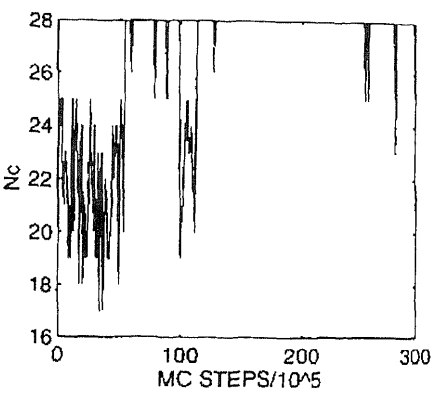
The foldicity studies on the designed sequences are carried out at different tempera- ıres and are presented in Table 2. At temperatures 1.3 and 1.5, the foldicity of all the de- igned sequences except for the initial random sequence is 1. The average number of MC teps require the fold generally decrease with increased native state energy and in partially epulsive cases. A complete trajectory of energy (E), the number of contacts (Nc) and the raction of the native contacts ($Q0$) for sequences 3 and 3R at $T = 1.5$ are presented in igures 3a and 3b, respectively. The sequence 3 with 20 rank-ordered native contacts and R with partially repulsive interactions fold at 5.4 and 0.28 million MC steps, respectively, t $T = 1.5$. The continuation of the simulation shows that they remain in the native state for large fraction of time. The low foldicity of all the sequences at $T = 1.0$ (Table 2) is due ɔ the trapping of the sequence in the local minimum and the low foldicity at higher tem- eratures ($>Tf$) is related to denaturation and the sequences not being able to remain in

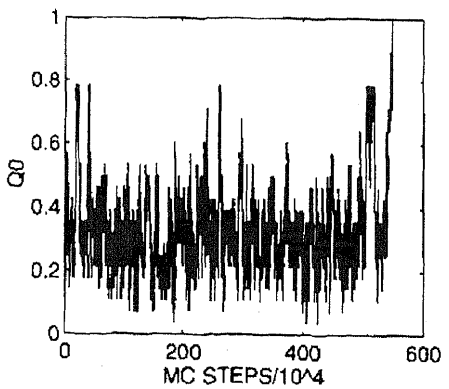
TRAJECTORY OF MC RUN(SEQUENCE 3,T = 1.5)

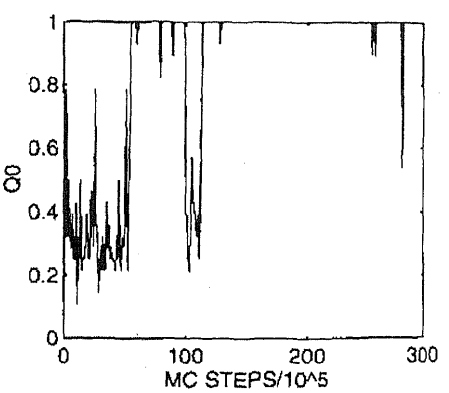




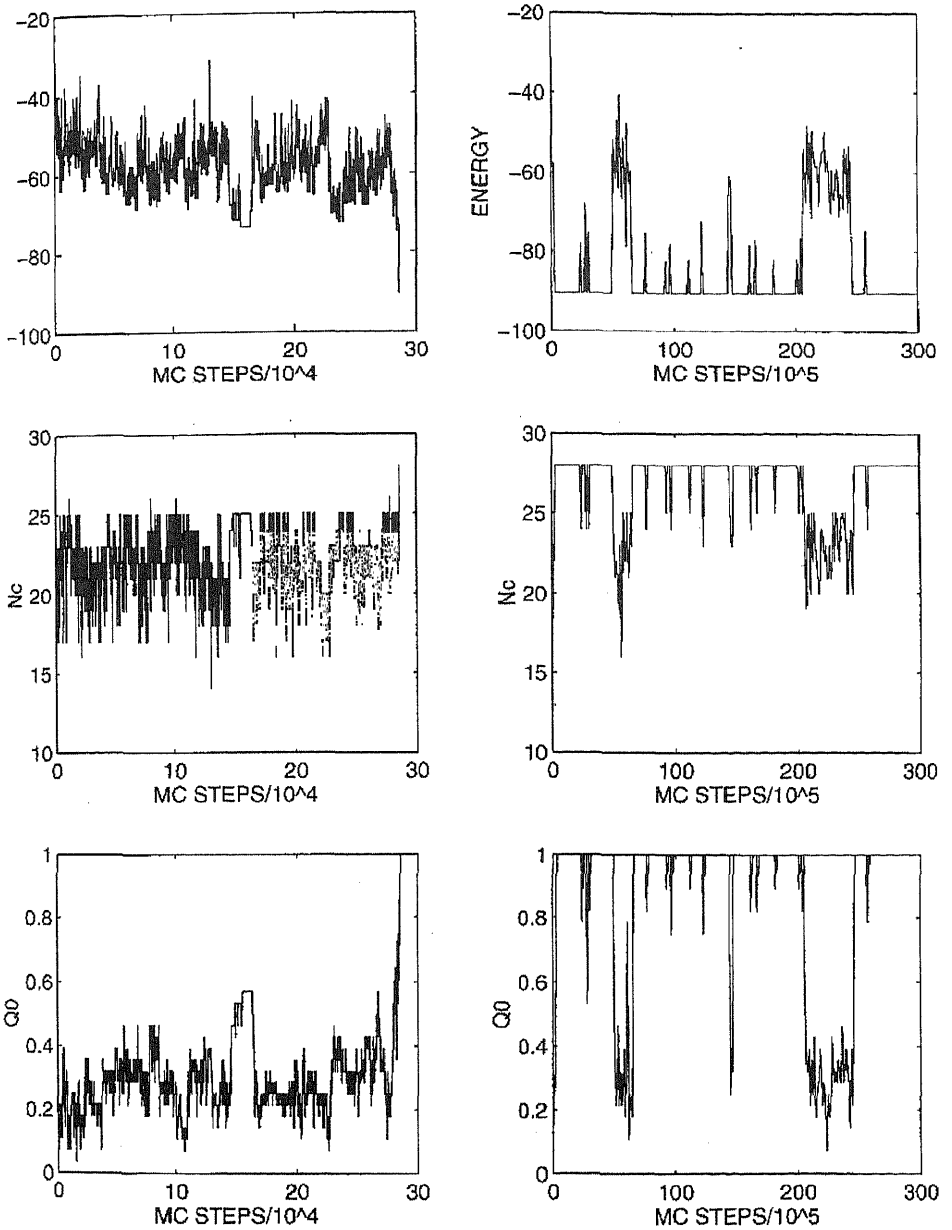








TRAJECTORY OF MC RUN(SEQUENCE 3R,T = 1.5)



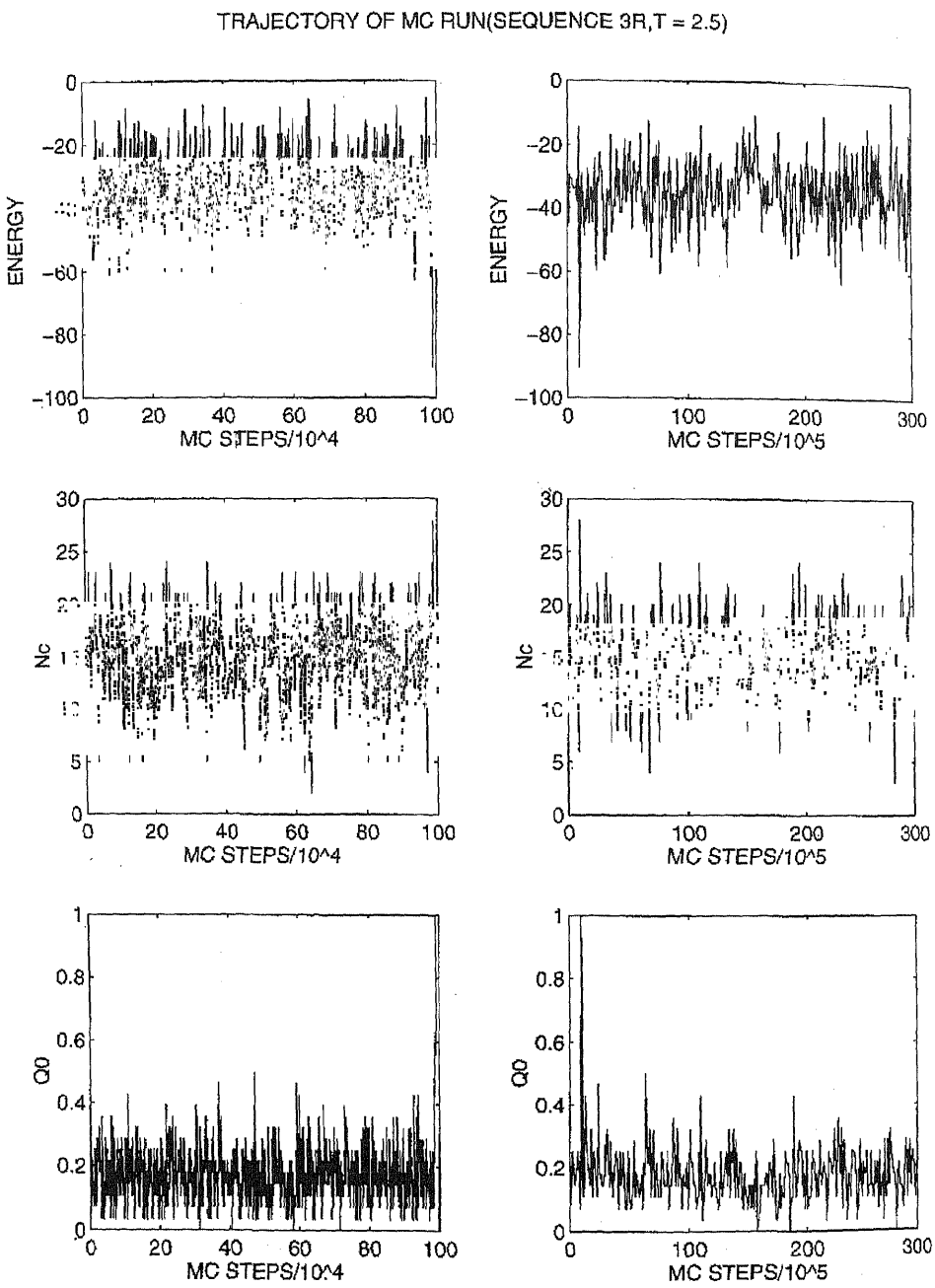


FIG. 3: (a) Typical MC trajectories of the purely attractive sequence 3 (Table 1) at $T = 1.5$. The panels on the left side correspond to the trajectories upto folding to the native state and the trajectories upto 30 million steps

are shown on the right-hand side panels. The energy E is measured in units of T. Nc is the number of contacts and $Q0$ is the fraction of native contacts.(Nc = 28 for a maximally compact state and $Q0$ = 1 for the native state) (b) Trajectories similar to those presented in (a) for the partially repulsive sequence 3R at T = 1.5. (c) Trajectories similar to those presented in (a) for the partially repulsive sequence 3R at T = 2.5.

compact states. This is demonstrated in Figure 3c, where the sequence 3R at T = 2.5, although folds at 0.99 million MC steps, is not stable in that conformation, as shown by E and $Q0$ of extended simulation. A lower average value of Nc also shows that the system does not retain compact conformation at high temperatures.

To summarize, we have worked out the conditions for maximum foldicity of polypeptide chains based on the principle of maximum compatibility. We have found that the process of protein folding is speeded up when the native state energy is increased and by the presence of repulsive groups, which suggests a solution to Levinthal's paradox.

We have also shown that the increase in the native state energy as found in real proteins need not be to the maximum extent and only a certain amount of increase in energy is sufficient to make the sequence fold. The sequences are shown to fold only at temperatures in an optimum range. While our studies have been carried out in a well characterized lattice model, the governing principle ought to be valid more generally and indeed in real proteins. A crucial aspect of our work is that the design of the protein as well as the test of the foldicity have been carried out with the same interaction potentials. The studies have evolutionary implications and implications in devising algorithms for designing proteins and for the prediction of their structure from the sequence information.

Table 2.
Foldicity* of sequences of native state A(From Table 1) as a function of temperature.

T Sequences	1.0	1.3	1.5	2.5	3.5
1	0	0.1 (23.3)	0		
2	0.4 (9.5)	1 (4.76)	1 (5.86)	0	
3	0.6 (13.67)	1 (7.02)	1 (2.07)	0.2 (5.08)	
4	0.4 (12.06)	1 (4.76)	1 (0.74)		0
1R	0.2 (17.27)	1 (7.26)	1 (8.55)		
2R	0.3 (16.23)	1 (3.70)	1 (1.75)	0.6 (6.76)	
3R	0.4 (4.65)	1 (2.45)	1 (0.92)	1 (3.65)	
4R	0.5 (6.82)	0.8 (2.07)	1 (0.76)		0

T – Temperature of the MC simulation.
*-10MC runs are carried out to determine foldicity, the entries in bracket are the average number of MC steps (in million).

3. Cell dynamics of model proteins

The above methodology allows us to design fast folding sequences. In order to understand the mechanism and pathway of folding , it is essential to follow the dynamics of folding. Such studies require the enumeration of the entire conformational space, which is not feasible in a 3-D lattice. Hence, rigorous studies have been undertaken at 2-D level on our designed fast folding sequences. In this framework, the pathway of folding and their dependence on temperature are illustrated via a mapping of the dynamics into motion within a space of maximally compact cells [4].

A 16-residue heteropolymer chain has been considered on a square lattice. The total conformations are 8,02,075 and maximally compact (cell) conformations are 69 for this system[17]. The chain has 49 interaction energy terms (B_{ij}'s), excluding the interactions with two immediate sequential neighbouring residues. B_{ij}'s are represented as $b_{ij} + B0$, $B0 < 0$ leads to compact states. An advantage of such a model is that, all possible conformations and parameters such as T_f can be exactly enumerated. Indeed, the equilibrium phase diagram for such a system showing transitions between an extended coil state, a disorganized globule state and an organized globule state has been constructed by Dinner *et al.*[18].

A fast-folding sequence (R) was designed (Figure 4) according to the method described in the previous section. The glass transition temperature (T_g) was operationally defined as the value of the temperature at which the median folding time is 300,000 MC steps. Several hundred independent runs were carried out to determine the median folding time. T_f and T_g as a function of $B0$ for the chosen sequence are given Figure 5. The ratio of T_f/T_g as a function of $B0$ (inset in Figure 5) is high when $B0$ is close to zero. Tf is determined from

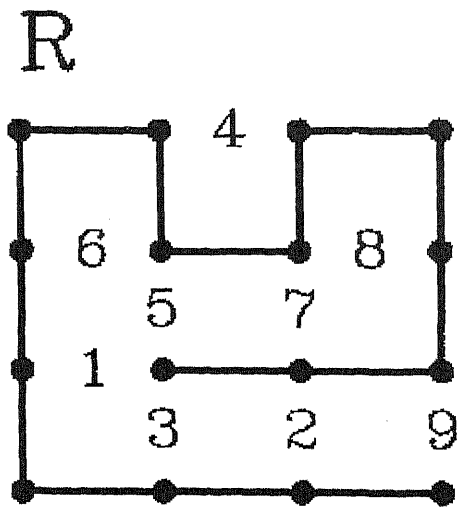


FIG. 4. The designed fast-folding 16-residue sequence (R) on a 4×4 square lattice. The numbers 1–9 represent the rank-ordered interaction energies (B_{ij}).

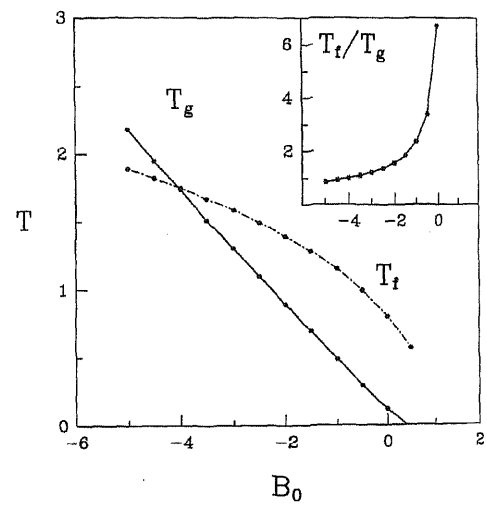


FIG. 5. A plot of T_f and T_g of the designed sequence as a function of $B0$. The T_f/T_g ratio as a function of $B0$ for this sequence is given in the inset.

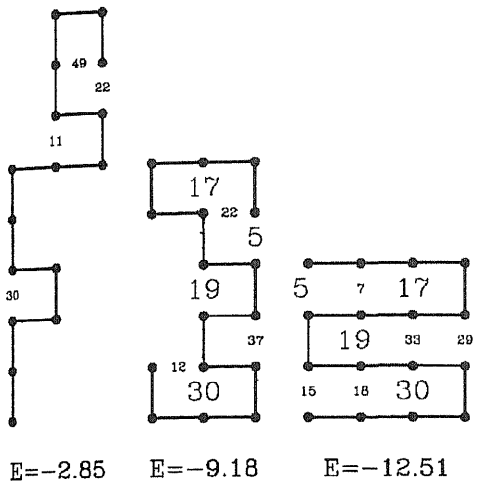


FIG. 6. Mapping a random conformation to a compact cell. The random conformation (on the left) is quenched to a local minimum (middle conformation) and mapped onto the compact cell (on the right). The bold numbers represent the common contacts between the local minimum and the cell.

the energy spectrum of the conformations and T_g is obtained dynamically via MC simulations[19].

The pathways of folding and their dependence on temperature are monitored via a mapping of the dynamics into the motion within the space of maximally compact cells. The method adopted derives inspiration from the pioneering work of Stillinger and Weber[20] in the context of super cooled liquids and that of Cieplak and Jaeckle's work[21] in the context of spin glasses. Each conformation of the heteropolymer is dynamically mapped onto one of the 69 cells. This is done by steepest descent quenching , followed by determining the cell with the largest energetic overlap through common contact. An example of such a mapping is given in Figure 6. In this way, the true dynamics is represented in a coarse grained manner as motion within the space of the 69 cells.

The cell to cell connections for the sequence R are shown in Figure 7 at temperatures 2, 0.8 and 0.3, which are above T_f, between T_f and T_g and below T_g respectively. The data is based on 200 starting configurations with the MC runs proceeding until folding and refer to the folding stage of the evolution. The column and rows denote the cell number arranged according to the energy with 1 corresponding to the native cell. The diagonal entries show the average occupancy of the cells during the runs. These entries when summed over all 69 cells, are normalized to 1000. For the off-diagonal terms, only successful transitions to a different cell are considered. They show the interlinking between different cells. The sum of all off diagonal entries is normalized to 1000 and the most significant entries, those bigger than or equal to 10 are displayed. The matrix is found to be approximately symmetric. The situation corresponding to the best folding is illustrated by the ma-

R

T=2.0

	1	7	14	19	36
1	293	13	15	21	18
7	13	60			
14	16		23		
19	21			32	
36	18				30

T=0.8

	1	2	3	4	6	7	8	17	20
1	378	21	24	20	29	13	60	22	10
2	21	67		24					
3	24		24						
4	20	24		27					
6	29				10				
7	13					13			
8	60						20		
17	22							8	
20	10								3

T=0.3

	1	3	7	10	19	20	23	29	43
1	544		34			20			
3		16					68		
7	34		64			67			
10				63			11	74	
19					2				28
20	20		67			3			
23		68		11			14	10	
29				74			10	18	
43					28				8

FIG. 7. A segment of 10000 Monte Carlo steps in the cell space at three temperatures (0.5, 0.8 and 2.0), one below T_g, the second one between T_g and T_f and the third one above T_f, for the fast-folding sequence R. The native cell has significantly less occupancy at $T = 0.3$ with fewer dynamical contacts with the other cells than at $T = 0.8$ showing the lack of equilibration, a breakdown of ergodicity, and the difficulty of rapid folding at the lower temperature. Less occupancy of native cell and a small number of other cells with dynamical contacts are also the features at $T = 2.0$.

S

T=2.0

	1	5	8	13	22	23	40
1	206	15	16	24	12	10	10
5	15	64					
8	17		63				
13	24			36			
22	12				37		
23	10					41	
40	10						18

T=0.8

	1	4	5	6	7	9	13
1	274		17			11	16
4		66			24		
5	17		58	44		13	
6			43	28			
7		24			22		
9	11		13			95	14
13	16					14	30

T=0.3

	1	2	4	5	6	9	11	13	14	21
1	117			25		13		14		
2		34	12							
4		12	16							
5	25			86	85					
6				85	68					17
9	13					96				
11							12		187	
13	14							40		
14							187		24	
21					17					11

FIG. 8. A plot similar to that of Figure 7 for a typical sequence.

trix calculated for $T = 0.8$–the low energy cells are directly connected to the native cell and the corresponding transition rates are high. At $T = 2$, the most significant transitions are from higher lying cells directly to the native cell but most of the other cells are also present in the dynamics. Below T_g, on the other hand, the pathways to folding are primarily through intermediate stages, starting from cell of high energy:; in addition there is a

significant weight in transition between cells which do not have any substantial links to the native cell. A similar cell dynamics for a typical sequence (S)[18], at various temperatures are presented in Figure 8. The sequence is a bad folder and the cell dynamics is characterized by a predominance of the hierarchical steps, and/or high energy jumps, at any temperature.

The cell-to-cell connectivity is observed to depend primarily on the value of the temperature relative to T_g and T_f. Our description based on 69 cells suggest that at least in 2-D systems, the pathways to folding are through a direct linking of a significant portion of viable cells to the native cell. This cluster of well connected cells can be considered to be a funnel for folding[22,23].

A related study to ours on folding pathways in a 3-D lattice model has been carried out by Leopold *et al*[22]. Their principal finding was that the good folder has a folding funnel through which kinetic access to the native state was facilitated. In contrast, in our 2-D study, we map each conformation of the heteropolymer to one of the maximally compact cells and study the dynamics of evolution in this coarsed grained cell space. Our studies suggest that good folders have a large T_f/T_g ratio, the cell motion is qualitatively different at different temperatures, and the folding pathways in cell space involve a substantial number of direct links to the native cell. Our study provides a systematic method of investigating the dynamics of folding.

Acknowledgment

We thank Supercomputer Education Research Center of Indian Institute of Science, for computational facilities and S.Vinayasree for the help in figures and manuscript preparation.

References:

1. CREIGHTON, T. E. ed. Protein Folding (Freeman,New York), 1992.

2. LEVINTHAL, C. In Mossbauer Spectroscopy in Bological Systems, eds. De brunner, P., Tsibris, J .C. M. and Munch, E. (Univ.Illinois Press, Urbana), 1969, pp.22–24.

3. SHRIVASTAVA, I.,VISHVESHWARA, S., CIEPLAK, M., MARITAN, A. and BANAVAR, J. R. *Proc. Nat. Acad. Sci.* USA, **92**, 9206–9209, (1995).

4. CIEPLAK, M., VISHVESHWARA, S., AND BANAVAR, J. R., PRL, 1996, **77**, 3681–3684.

5. MIYAZAWA, S AND JERNIGAN, R. L. Macromolecules, 1985, **18**, 534–552.

6. SALI, A., SHAKHNOVICH, E. AND KARPLUS, M. Nature (London) 1994, **369**, 248–251.

7. SALI, A., SHAKHNOVICH, E. AND KARPLUS, M. *J. Mol. Biol.*, 1994, **235**, 1614–1636.

8. GRIBSKOV, M. AND DEVEREUX, J., EDS Sequence Analysis Primer Stockton, New York, Appendix III, 1991, p. 229.

9. BRYNGELSON, J. D., ONUCHIC, J. N., SOCCI, N. D. AND WOLYNES, P. G. *Proteins Struct. Funct. Genet.*, 1995, **21**, 167–195.

10. BRYNGELSON, J. D. and WOLYNES, P. G. *J. Phys. Chem.*, 1989, **93**, 6902–6915.

11. BRYNGELSON, J. D. AND WOLYNES, P. G. *Proc. Natl. Acad. Sci.* USA, **1987**, 84, 7524–7528.

12. GOLDSTEIN, R. A., LUTHEY-SCHULTEN, Z. A. AND WOLYNES, P. G. *Proc. Natl. Acad. Sci* USA, 1992, **89**, 4918–4922.

13. GOLDSTEIN, R. A., LUTHEY-SCHULTEN, Z. A. AND WOLYNES, P. G. Proc. Natl. Acad. Sci. USA, 1992, **89**, 9029–9033.

14. GO, N. *Annu. Rev. Biophys. Bioeng.*, 1983, **12**, 183–210.

15. CIEPLAK, M., MARITAN, A. AND BANAVAR, J. R. *Phys.Rev. Lett.*, 1994, 72, 2320–2323.

16. NEWMAN, C. M. AND STEIN, D. L. *Phys. Rev. Lett.*, 1994, 72, 2286–2289.

17. LAU, K. F. AND DILL, K. A. *Macromolecules*, 1989, **22**, 3986–3997.

18. DINNER, A., SALI, A., KARPLUS, M. AND Shakhnovich, E. *J. Chem.Phys.*, 1994, **101**, 1444–1451.

19. SOCCI, N. D. AND ONUCHIC, J. N. *J. Chem. Phys.*, 1994, 101, 1519–1528.

20a. STILLINGER, F. H. AND WEBER, T. A. *Phys. Rev. A*, 1983, **28**, 2408–2416.

20b. STILLINGER, F. H. AND WEBER, T. A. *Science*, 1984, 225, 983–989.

21. CIEPLAK, M. AND JAECKLE, J. *Z. Phys.*, 1987, **66**,.325–332.

22. LEOPOLD, P. E., MONTAL, M. AND ONUCHIC, J. N. *Proc. Natl. Acad. Sci.*, USA, 1992, **89**, 8721–8725.

23. ONUCHIC, J. N., WOLYNES, P. G., LUTHEY-SCHULTEN, Z. AND SOCCI, N. D. *Proc. Natl. Acad. Sci.* USA, 1995, **92**, 3626–3630.