Eur. Phys. J. B **14**, 787–792 (2000)

The European Physical Journal B
EDP Sciences

Simple models of proteins with repulsive non-native contacts

M.S. Li[a] and M. Cieplak

Institute of Physics, Polish Academy of Sciences, Al. Lotnikow 32/46, 02-668 Warsaw, Poland

Received 10 May 1999 and Received in final form 13 October 1999

Abstract. The Go model is extended to the case when the non-native contact energies may be either attractive or repulsive. The folding temperature is found to increase with the energy of non-native contacts. The repulsive non-native contact energies may lead to folding at $T = 0$ for some unusual two-dimensional sequences and to reduction in complexity of disconnectivity graphs for local energy minima.

PACS. 87.10.+e General theory and mathematical aspects – 87.15.By Structure and bonding

Functionally useful proteins are sequences of amino acids that fold rapidly under appropriate conditions (temperature range, acidity of the water solution etc.) into their native states commonly assumed to be their ground state configurations [1]. The dynamics of folding is akin to motion in a rugged free energy landscape [2] and it crucially depends on two factors: the interactions between the amino acids and the target conformations. It should be noted that the repulsive non-native contact interaction are expected to improve the foldability because they restrict the size of the relevant phase space. This effect has been demonstrated for a Gaussian model [3], a designed model [4], and the so called HP+ model [5]. In this paper, we focus on a simpler model which would allow to see the role of non-native contact energies in the folding process in an easy way. We choose a model which is an extension of the so-called Go model proposed by Go and Abe [6].

Specifically, we consider the standard two-dimensional lattice model of 16 monomers. Its Hamiltonian is as follows

$$H = \sum_{i<j} \alpha_{ij} \Delta_{ij}, \tag{1}$$

where $\Delta_{ij} = 1$ if monomers i and j are in contact and $\Delta_{ij} = 0$ otherwise (monomers i and j are considered to be in contact if they are separated by one lattice bond and $|i - j| \neq 1$). The quantity $\alpha_{ij} = -1$ if monomers i and j are in contact in the native conformation and $\alpha_{ij} = \alpha$ otherwise. The sequence is thus defined by the native conformation. We allow α to be attractive ($\alpha < 0$) or repulsive ($\alpha > 0$). In this paper only maximally compact native states are considered. In what follows α is assumed to be greater than -1.0 so that the target native structure is the ground state. In the original Go model [6] $\alpha = 0$. It should be noted that using model (1) one can monitor the effect of non-native contact energies by varying only one parameter α. Furthermore, the two-dimensional model is simple enough to study the effect of the target conformations on the folding dynamics. [5]. The advantage of our generalized Go model (1) compared, say, to the HP+ [5], is that it allows one to vary the strength, and sign, of the non-native contacts relative to the native ones.

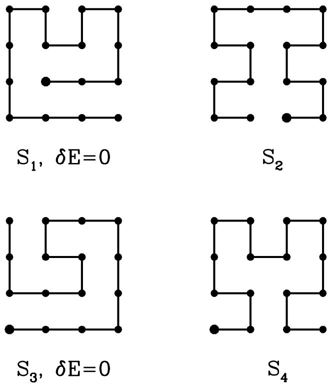

Fig. 1. Four sequences studied in this paper. S$_1$ and S$_3$ have $\delta E = 0$.

In reference [7] one observes that some 13-monomer chains of the HP model can fold even at $T = 0$. One of our purposes is to examine if such exotic phenomenon takes place in other models.

We focus on 4 target conformations shown in Figure 1. There are 37 compact conformations (for the two-dimensional 16-monomer chain one has 69 compact 4×4 conformations but for the Go-like model only 38 of them remain different due to the end-to-end reversal symmetry and one is not accessible kinetically with respect to the standard dynamics envolving the single and double monomer moves [1]) which may act as the native conformations. Among these S$_1$ and S$_4$ shown in Figure 1 are the two fastest folders at $T \neq 0$, whereas S$_2$ and S$_3$,

---

[a] e-mail: `masil@thp.uni-koeln.de`

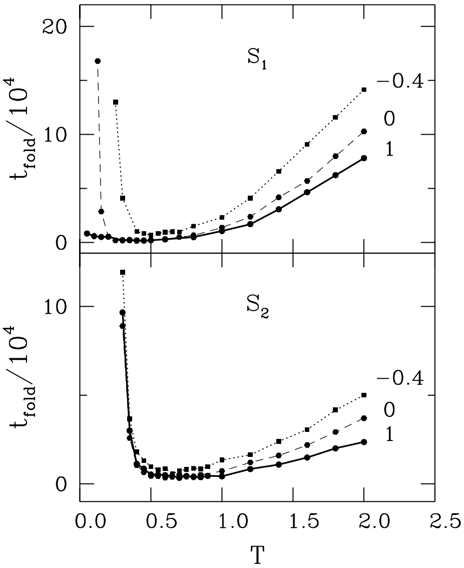

Fig. 2. Temperature dependence of $t_{\rm fold}$ for S_1 and S_2 and for 3 selected values of α, $\alpha = -0.4, 0$ and 1. The results are based on 2–4 batches, each containing 200 trajectories from initial random conformations.

also shown in Figure 1, have intermediate folding properties for positive values of α. It should be noted that the folding times of all of these sequences at $T \neq 0$ have the same order of magnitude.

We have shown that the effect of the repulsive non-native contact is so dramatic that for $\alpha > 0.1$ sequences S_1 and S_3 can fold with a finite $t_{\rm fold}$ even at $T = 0$. So this exotic phenomenon is not restricted to the HP model [7]. Here, we study how this arises as a function of α and show how do folding characteristics, *i.e.* characteristic temperatures and folding times, depend on α. The ability of S_1 and S_3 to fold at $T = 0$ may be partly understood by the fact that for these sequences the repulsive non-native interactions reduce the number of local minima by two orders of magnitude compared to the case of $\alpha < 0$. Furthermore, repulsive interactions dramatically affect partitioning of the phase space into regions associated with the local energy minima. We demonstrate this by using the disconnectivity graph technique [8] and show, in particular, that connectivities to the folding funnel become simplified significantly.

We have found that the folding temperature, $T_{\rm f}$, increases with α. This result agrees with that of Camacho [9] for an effectively zero-dimensional model [10]. We obtain it, however, not only by the numerical calculations for the lattice model but also by the analytical argument. The minimum folding time, $t_{\rm min}$, defined at the temperature where the folding is fastest has found to decrease with α and it gets saturated for $\alpha \to \infty$.

The folding dynamics of a chain is studied by a Monte Carlo procedure that satisfies the detailed balance condition [11], and was motivated by the studies presented in references [7,12]. The dynamics allows for single and two-monomer (crankshaft) moves. For each conformation of the chain one has A possible moves and the maximum value of A, $A_{\rm max}$, is equal to $A_{\rm max} = N + 2$. In our 16-monomer case $A_{\rm max} = 18$. For a conformation with A possible moves, probability to attempt any move is taken to be $A/A_{\rm max}$ and probability not to do any attempt is $1 - A/A_{\rm max}$ [7,12]. In addition, probability to do a single move is reduced by the factor of 0.2 and to do the double move by 0.8 [12,7]. The attempts are rejected or accepted as in the standard Metropolis method. The folding time is equal to the total number of Monte Carlo attemps divided by $A_{\rm max}$.

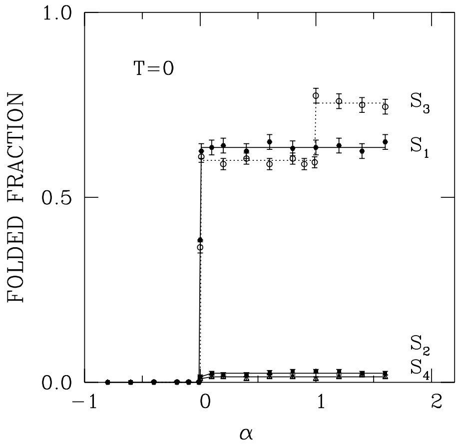

Fig. 3. Dependence of the folded fraction on α at $T = 0$ for 4 sequences shown in Figure 1. The results are averaged over 4–6 simulations, each corresponding to 200 trajectories.

We have carried out the Monte Carlo simulations to determine the dependence of the folding time, $t_{\rm fold}$, on T and α. $t_{\rm fold}$ is defined as the median first passage time. The results for S_1 and S_2 are shown in Figure 2. For each temperature, $t_{\rm fold}$ defined as the median first passage time, is obtained based on 200 independent runs starting from random configurations. The results are averaged over 2-4 batches, of 200 trajectories each.

For sequence S_2, the standard U-shape [13] for the temperature dependence of $t_{\rm fold}$ is observed for all values of α. In other words, no qualitative change occurs if the non-native contact energies change from attractive to repulsive. In the case of sequence S_1, however, for $\alpha = 1$ the standard U-shape disappears suggesting that the glass transition temperature $T_{\rm g}$ which is operationally defined as the value of the temperature at which the median time is equal to some cut-off value (usually this cut-off value is chosen to be 300 000 Monte Carlo steps for the two-dimensional 16-monomer chain [13]) becomes zero. In the standard scenario, at low temperatures the system may get trapped in some local minima and the folding process becomes extremely slow. The folding time is then governed by the Arrhenius law, $t_{\rm fold} \sim \exp(\delta E/T)$ [11,14,15], where δE is the energy barrier energy (at $T = 0$ one has $t_{\rm fold} \to \infty$). Thus, the absence of the U-shape dependence suggests that the energy barrier $\delta E = 0$.

In order to know whether δE is exactly zero or not one has to study the folding at $T = 0$. At zero temperature the system gets trapped in some local minimum or in the native state. If the fraction of the trajectories from

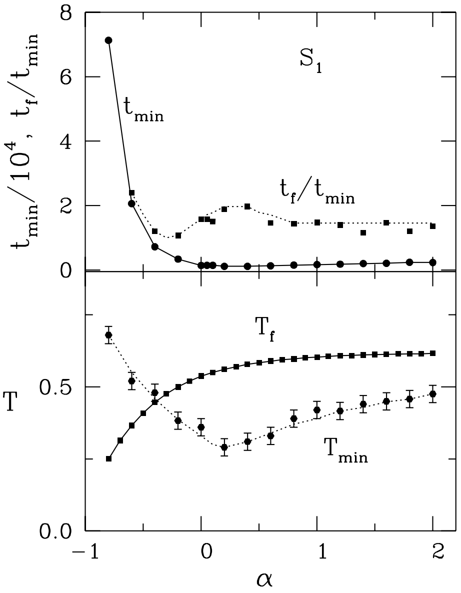

Fig. 4. Dependence of T_{f}, T_{min}, t_{min} (solid line) and $t_{\mathrm{f}}/t_{\mathrm{min}}$ (dotted line) for sequence S_1 on α. The results are averaged over 4 batches, each containing 200 trajectories. For $\alpha = -0.8$, S_1 does not fold at $T_{\mathrm{f}} = 0.25$ within 3×10^6 Monte Carlo steps and $t_{\mathrm{f}}/t_{\mathrm{min}} > 42$.

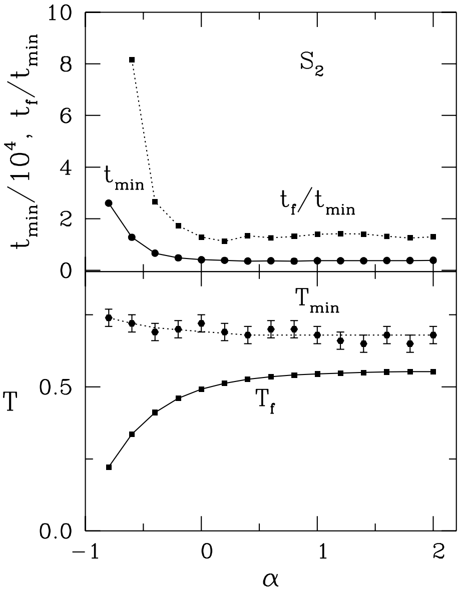

Fig. 5. The same as in Figure 4 but for sequence S_2. For $\alpha = -0.8$, S_2 does not fold at $T_{\mathrm{f}} = 0.22$ within 3×10^6 Monte Carlo steps and $t_{\mathrm{f}}/t_{\mathrm{min}} > 115$.

random conformations fold into the native state is bigger than 50%, then the chain is said to be folded and $\delta E = 0$.

The fraction of folded trajectories is shown in Figure 3 for S_1, S_2, S_3 and S_4. For S_1 and S_3 this fraction becomes bigger than 50% for $\alpha > \alpha_{\mathrm{c}}$, where $\alpha_{\mathrm{c}} \approx 0.1$. Sequences S_2 and S_4 have $\delta E \neq 0$ for any value of α. It is interesting to mention that the folding rates of S_4 and S_1 are comparable at $T \neq 0$ but the foldability of S_4 becomes much worse at $T = 0$. Thus the geometry of the native targets has a dramatic effect on the folding at $T = 0$. Among the 37 maximally compact 4×4 structures it is only S_1 and S_3 that do not obey the Arrhenius law at low temperatures. It should be noted that the folding at $T = 0$ strongly depends on the kind of dynamics assigned to the system. If one uses the dynamics with only single monomer moves or the so-called slithering snake dynamics [16,17] then S_1 and S_3 can not fold at $T = 0$ for any value of α.

We study the dependence of T_{f} on α for two typical sequences S_1 and S_2 shown in Figure 1. The total number of conformations of the 16-monomer chain is only 802 075 [18,19] and is amenable to exact enumeration. This allows for an exact evaluation of the equilibrium parameters such as the folding temperature T_{f}. The latter is defined as a temperature at which the probability of occupancy of the native state is 1/2. The results for sequences S_1 and S_2 are shown in Figures 4 and 5. Clearly, T_{f} increases with α but this dependence gets weaker for larger values of α.

The increase of T_{f} with α may be understood in the following simple way. Let $P_{\varGamma_0}(\alpha)$ be the probability of occupying the native state of conformation $\varGamma_0$. Then

$$P_{\varGamma_0}(\alpha) = \frac{\exp^{-\beta E_{\varGamma_0}}}{\sum_{\varGamma} \exp^{-\beta \alpha n - \beta E'_{\varGamma}}}, \tag{2}$$

where E_0 is the energy of the native state, n is the number of non-native contacts and $E'_{\varGamma}$ is the part of energy in conformation $\varGamma$ which corresponds to the native contacts. Then

$$\frac{\partial \ln(P_{\varGamma_0})}{\partial \alpha} = \beta \langle n \rangle_T \geq 0, \tag{3}$$

where $\langle n \rangle_T$ is the average number of non-native bonds at temperature T. So the probability of being in the native state cannot decrease with α. T_{f}, therefore, should increase with increasing α and then become α-independent.

It should be noted that in Camacho's model [9] T_{f} was found to increase with α ($\alpha < 0$) linearly. Our results presented in Figures 4 and 5 show that the region of α where one can observe the linear dependence is rather narrow. Such region becomes much wider, for example, in the case of the 27-monomer chain in three dimensions (the results are not shown). Overall, the results for T_{f}, shown in Figures 4 and 5, demonstrate that the non-native contact repulsive energies improve both the thermodynamic stability and dynamical characteristics of folding.

Figures 4 and 5 also show the dependence of the minimal folding time, t_{min}, on α. The dependence seems to saturate at large values of α. Such saturation has been also seen in the zero-dimensional model of Camacho [9]. Thus, the repulsive non-native contact energies make the polypeptide chain to fold faster which is similar to what has been observed in references [3,5]. We now focus on the foldability of S_1 and S_2 using different criteria. One of

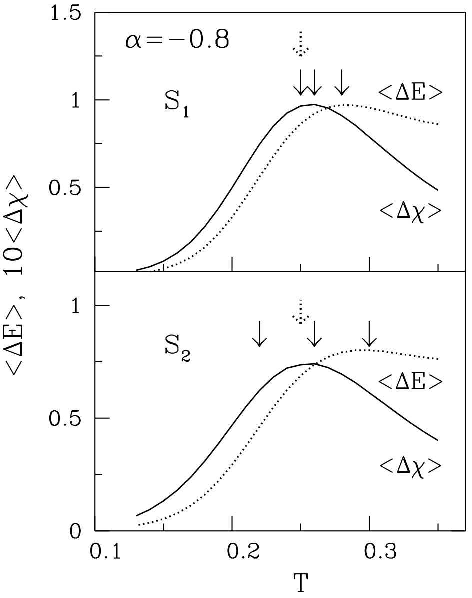

Fig. 6. Temperature dependence of $\langle \Delta E\rangle$ and $\langle \Delta\chi\rangle$ for S_1 and S_2 and $\gamma = -0.8$. The solid arrows indicate the positions of $T_{\rm f}$, T_χ and T_θ from the left to the right respectively. The dotted arrow corresponds to the maximum of C_v. Clearly, T_θ and T'_θ do not coincide.

them is based on the ratio of $T_{\rm f}$ and $T_{\rm min}$: if $T_{\rm f}$ is larger or comparable with $T_{\rm min}$ then a sequence should be a good folder [13]. The other criterion is based on the Thirumalai-Camacho ratio σ or σ',

$$\sigma = \frac{|T_\chi - T_\theta|}{T_\theta},\ \sigma' = \frac{|T_\chi - T'_\theta|}{T'_\theta}, \tag{4}$$

where T_χ, T_θ and T'_θ correspond to the maximum of the structural susceptibility $\langle \Delta\chi\rangle$, the energy fluctuation $\langle \Delta E\rangle$ and the specific heat C_v, respectively [15,20,21]. We have

$$\begin{aligned}\langle \Delta\chi\rangle &= \langle\chi^2\rangle - \langle\chi\rangle^2,\\ \chi &= 1 - \frac{1}{M^2-3M+2}\sum_{i\neq j, j\pm 1}\delta(r_{ij} - r_{ij}^{\rm N}),\\ \langle \Delta E\rangle &= \langle E^2\rangle - \langle E\rangle^2,\\ C_v &= \frac{\langle \Delta E\rangle}{T^2},\end{aligned} \tag{5}$$

where M is a number of beads, r_{ij} is a distance between i-th and j-th monomers and the superscript N denotes the native state [15]. It should be noted that in some cases σ and σ' are different. The sequence should be a good folder if $\sigma \leq \sigma_{\rm c}$ (or $\sigma' \leq \sigma'_{\rm c}$). In reference [22], it is argued that $\sigma'_{\rm c}$ should be of order =0.4. For $\alpha \geq -0.6$ sequence S_1 is a good folder by both criteria because $T_{\rm f}$ and $T_{\rm min}$ are comparable and the peaks in the T-dependencies of the structural susceptibility, specific heat coincide and energy fluctuation. For $\alpha = 0$, the same positioning of the peaks is, in fact, a general feature of the Go models because the proximity to the native state and occurrence of rapid changes in an average energy as a function of T are both controlled by establishment of the same native contacts. The situation becomes more complicated for $\alpha = -0.8$. In this case we have $\sigma \approx 0.07$ and $\sigma' \approx 0.04$ (see Fig. 6) which suggests that S_1 should be a good folder by the Thirumalai-Camacho criterion. On the other hand, a comparison of $T_{\rm f}$ to $T_{\rm min}$ suggests that S_1 remains a bad folder. In order to gain some insights into this discrepancy, we focus on the kinetics directly and calculate the ratio of $t_{\rm f}/t_{\rm min}$, where $t_{\rm f}$ and $t_{\rm min}$ are folding times at $T_{\rm f}$ and $T_{\rm min}$ respectively. We propose that a sequence should be considered to be a good folder if this ratio is small. The results for $t_{\rm f}/t_{\rm min}$ are shown in Figure 4 for S_1. For $\alpha \geq -0.6$ the ratio $t_{\rm f}/t_{\rm min}$ is around 2 and S_1 is then a good folder. This agrees with the conclusion drawn from the first two criteria. For $\alpha = -0.8$ sequence S_1 could not fold at $T_{\rm f} = 0.25$ within 3×10^6 Monte Carlo steps and we have $t_{\rm f}/t_{\rm min} > 42$. This large value of $t_{\rm f}/t_{\rm min}$ indicates that S_1 is a bad folder for $\alpha = -0.8$ which does not agree with the thermodynamic criterion but it does agree with the criterion based on the relationship between $T_{\rm f}$ and $T_{\rm min}$. It should be noted, however, that all of these criteria have some quantitative uncertainties and are thus naturally somewhat qualitative. We consider the kinetic criterion to carry the biggest weight since it directly relates to the folding process.

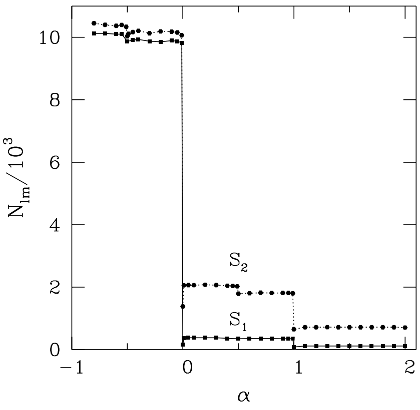

Fig. 7. Dependence of $N_{\rm lm}$ on α for S_1 and S_2.

For $\alpha \geq -0.4$ sequence S_2 is a good folder by all of the three criteria discussed above. For $\alpha = -0.6$ and -0.8 we have $\sigma \approx 0.08$ ($\sigma' \approx 0.03$) and $\sigma \approx 0.13$ ($\sigma' \approx 0.04$) (see Fig. 6), respectively. The corresponding values of $t_{\rm f}/t_{\rm min}$ is ≈ 8.2 and > 115. The small value of σ in the thermodynamic criterion suggests good foldability of S_2 for these values of α. However, the kinetic criteria, based on the ratios of $t_{\rm f}/t_{\rm min}$ and $T_{\rm f}/T_{\rm min}$ indicate otherwise.

In order to understand why does the repulsion improve the folding times so much we study the dependence of number of local minima, $N_{\rm lm}$, on α. $N_{\rm lm}$ depends on the type of the dynamic moves that are allowed (see, for example, Ref. [17]). Our results presented below are obtained by the same set of moves as in the study of folding kinetics. The results for S_1 and S_2 are shown in Figure 7. $N_{\rm lm}$ of S_1 is found to be smaller than for S_2. Clearly, the number of local minima strongly depends on α and for $\alpha > 0$

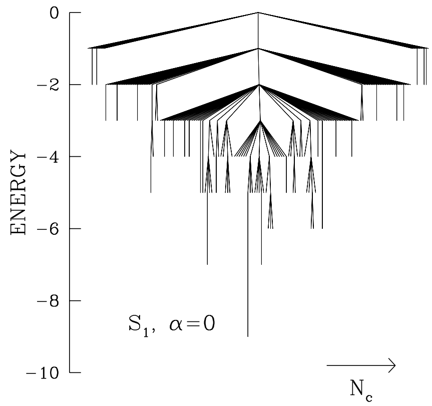

Fig. 8. The disconnectivity graph for S_1 and $\alpha = 0$. N_c is a symbolic notation for the index labelling the local energy minima.

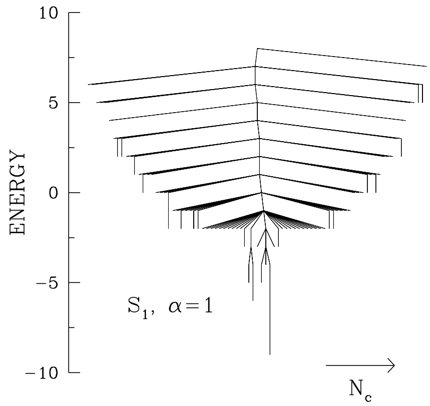

Fig. 9. The same as in Figure 7 but for $\alpha = 1$.

it becomes by 2 orders of magnitude smaller than that for $\alpha < 0$. Thus the positive values of α make the energy landscape less rugged and the folding dynamics get faster.

In order to get more insight into the nature of the energy landscapes in the models studied here, we use the disconnectivity graph technique which maps the potential energy surface onto the set of local minima [8,23]. The disconnectivity graph representation of the phase space involves checking what local energy minima are connected by optimal pathways (strings of moves that are allowed kinetically) that do not exceed a given total threshold energy. Such optimal pathways have been discussed in reference [15]. For each value of this energy the minima are divided into disconnected sets of mutually accessible minima separated by barriers. The local minima which share the lowest energy threshold are joined at a node by lines and are called a basin corresponding to the threshold. The procedure of construction is stopped when one gets only one basin for all of the minima. For S_1 the number of local minima is equal to 152 and 81 for $\alpha = 0$ and $\alpha = 1$, respectively. The construction of the graphs, therefore, may be done exactly. The disconnectivity graphs obtained for S_1 and $\alpha = 0$ and $\alpha = 1$ are shown in Figures 8 and 9, respectively. For both values of α sequence S_1 is a good folder and consequently, the structure corresponding to a folding funnel is clearly visible. However, the funnel for the repulsive case of $\alpha = 1$ has a significantly less complex pattern than for $\alpha = 0$. Moreover, the repulsive non-native interactions reduce the number of the local energy minima that have links to the native state below any predetermined energy threshold. Thus, the comparison of the disconnectivity graphs also shows that the repulsion may facilitate the folding substantially.

We now address the question of what happens with trapped local minima when α changes from negative to positive values. Figure 10 shows the histogram of the local minima for $\alpha = -0.1$ and 0.1 for sequence S_1 and S_2. Interestingly, for S_1 none of the local minima obtained for $\alpha = -0.1$ appears for $\alpha = 0.1$. The situation changes dramatically for S_2 for which 19 local minima are common for both $\alpha = -0.1$ and $\alpha = 0.1$. Our results suggest that for the sequence with $\delta E = 0$ the local minima in which the chain is trapped at the negative values of α are effectively avoided if α is changed to positive values.

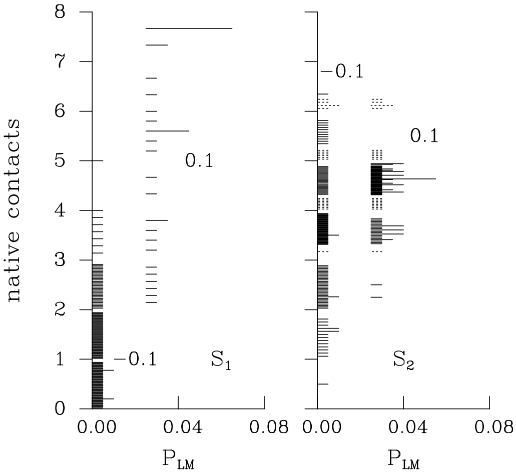

Fig. 10. The density of the (V-shape) local minima, $P_{\rm lm}$, in which the system gets trapped at $T = 0$ is plotted *versus* the number of native contacts for S_1 and S_2. We choose $\alpha = 0.1$ and $\alpha = -0.1$. The results are obtained for the batch of 200 trajectories. In the case of S_1 the local minima are different for $\alpha = -0.1$ and $\alpha = 0.1$. For S_2 there are 19 common local minima which are marked by dotted lines.

In conclusion, we state that for the simple extended Go model $T_{\rm f}$ ($t_{\rm min}$) increases (decreases) with the non-native contact energy and it gets saturated for large values of α. The complexity of the disconnectivity trees becomes reduced on making α more and more repulsive so that some two-dimensional sequences may even lose the Arrhenius like behavior of $t_{\rm fold}$ at low temperatures. It would be interesting to determine whether there are any three-dimensional Go-like sequences that fold even at $T = 0$. Another important question is what kind of an effective parameter α, or its conceptual equivalent, characterizes real proteins. Note that in real proteins interactions between aminoacids are more varied energetically than in the Go models studied here. Nevertheless, the Go models may be useful in the studies of folding of real proteins. In particular, Zhou and Karplus [24] have recently performed molecular dynamics studies of an off-lattice three-helix protein with the Go-like couplings. They have demonstrated that the mechanisms of folding change as a function of the strength of the non-native contacts relative to the native ones.

We thank J.R. Banavar, P. Garstecki and T.X. Hoang for many useful discussions. This work was supported by KBN (grant No. 2P03B-025-13).

References

1. K.A. Dill, S. Bromberg, S. Yue, K. Fiebig, K.M. Yee, D.P. Thomas, H.S. Chan, Protein Sci. **4**, 561 (1995).
2. J.D. Bryngelson, J.N. Onuchic, N.D. Socci, P.G. Wolynes, Proteins **21**, 167 (1995).
3. I. Shrivastava, S. Vishveshwara, M. Cieplak, A. Maritan, J.R. Banavar, Proc. Natl. Acad. Sci. USA **92**, 9206 (1995).
4. A.M. Gutin, V.I. Abkevich, E.I. Shakhnovich, Biochemistry **34**, 3066 (1995).
5. H.S. Chan, K.A. Dill, Proteins: Struct. Funct. Genet. **33**, 2 (1998).
6. N. Go, H. Abe, Biopolymers **20**, 1013 (1981).
7. H.S. Chan, K.A. Dill, J. Chem. Phys. **100**, 9238 (1994).
8. O.M. Becker, M. Karplus, J. Chem. Phys. **106**, 1495 (1997).
9. C.J. Camacho, Phys. Rev. Lett. **77**, 2324 (1996).
10. A. Gutin, V. Abkevich, E. Shakhnovich, Phys. Rev. Lett. **80**, 208 (1998).
11. M. Cieplak, M. Henkel, J. Karbowski, J.R. Banavar, Phys. Rev. Lett. **80**, 3654 (1998); M. Cieplak, M. Henkel, J.R. Banavar, Cond. Matt. Phys. (Ukraine) **2**, 369 (1999).
12. H.S. Chan, K.A. Dill, J. Chem. Phys. **99**, 2116 (1994).
13. N.D. Socci, J.N. Onuchic, J. Chem. Phys. **101**, 1519 (1994); M. Cieplak, J.R. Banavar, Folding and Design **2**, 235 (1997).
14. A.M. Gutin, V.I. Abkevich, E.I. Shakhnovich, Phys. Rev. Lett. **77**, 5433 (1996).
15. C.J. Camacho, D. Thirumalai, Proc. Natl. Acad. Sci. USA **90**, 6369 (1993).
16. F.T. Wall, F. Mandell, J. Chem. Phys. **63**, 4592.
17. T.X. Hoang, M. Cieplak, J. Chem. Phys. **109**, 9192 (1998).
18. K.F. Lau, K.A. Dill, Macromolecules **22**, 3986 (1989).
19. A. Dinner, A. Sali, M. Karplus, E. Shakhnovich, J. Chem. Phys. **101**, 1444 (1994).
20. C.J. Camacho, D. Thirumalai, Europhys. Lett. **35**, 627 (1996).
21. D. Thirumalai, J. Phys. I France **5**, 1457 (1995).
22. T. Veitshans, D. Klimov, Thirumalai, Fold. Des. **2**, 1 (1997).
23. D.J. Wales, M.A. Miller, T.R. Walsh, Nature **394**, 758 (1998).
24. Y. Zhou, M. Karplus, Nature **401**, 400 (1999).