Simulation of mutation: Influence of a ''side group'' on global minimum structure and dynamics of a protein model

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

(Received 22 March 1999; accepted 21 May 1999)

The 46-bead, three-color model of a β-barrel-forming protein is modified by the addition of a single side group, represented by a bead which may be hydrophilic or hydrophobic. Molecular dynamics and quenching simulations show how the nature and location of the bead influence both the structure at the global minimum of internal energy and the relaxation processes by which the system finds its minima. The most drastic effects occur with a hydrophobic side group in the middle of a sequence of hydrophobes. © *1999 American Institute of Physics.* [S0021-9606(99)50931-8]

I. INTRODUCTION

That small changes in the sequence of protein's amino acids may drastically affect its biological functions is well known. Much theoretical effort has been expended to understand (or, even better, to predict) the kinds of changes and the points of the polypeptide chain to which these functions are most sensitive. The protein's biological activity is believed to be determined by its geometrical structure at the minimum (or at one of several suitable minima) of its multidimensional potential energy (or free energy) surface. Consequently finding the global-minimum structure of the folded chain and the rate at which the system finds this minimum from an unfolded configuration have been goals of most of those studies.[1–5]

In the present paper we use molecular dynamics simulations and the vehicle of the three-color 46-bead model to study how the addition of a side group to the main chain of 46 amino acids affects the relaxation rates and structure of the lowest minima of the protein potential energy. A single bead attached to one of the beads of the chain represents the side group. We have chosen that protein model for the following reasons. The model, with its tuned sequence of beads and bead–bead interaction parameters, was developed by Skolnick *et al.*[6–9] as a lattice representation and then was extended by Thirumalai *et al.*[10,11] to be a continuum model. It was then shown that this model reaches the β-barrel structure of the polymer guided by the staircase nature of the topography of its potential surface, as structure-seeking systems seem to be.[12,13] Later the model was used to investigate how the topography of the potential surface is linked to the dynamics of its folding.[14] It was then modified for a study of the self-assembly of separated strands into the β-barrel structure.[15] A similar model for a 38-residue chain was the object of a study by Hao and Scheraga.[16–18]

How might we use molecular dynamics to find the effects of mutations on the structures and dynamic behavior of proteins? One standard way is using molecular dynamics (MD) with Stillinger–Weber quenching[19–21] to find the global minimum and low-energy configurations of complicated multidimensional potential surfaces. One usually begins MD runs (constant temperature or energy, or with annealing) from randomly selected configurations and finds, by the quenching procedure, the minima visited along the trajectory of the run. This method does not guarantee, however, that one will find the global minimum. Trajectories of systems of some complexity initiated at rather high temperatures or total energies may well take far too long to enter the basin containing the global minimum, due to the enormous number of their accessible configurations. On the other hand, at low temperatures the system may be trapped at some local minimum and never reach the global minimum.

If we had some idea what the geometry of the global minimum might be, we could use a variety of ''close-to-minimum'' (CTM) structures as starting configurations for MD runs, hoping that they will bring us to the real global minimum faster and with greater probability than runs from randomly chosen initial configurations. In such an approach we should, however, keep in mind that runs starting from some low-energy configuration may very well explore that initial basin only, which may not be the one where the global minimum is located. So if there are several ''suspicious'' CTM configurations, all of them should be checked, especially those separated by high barriers. Surely the closer the assumed CTM structure is to the real global minimum, the better the method works; however we have found (see below) that even when the assumed structure appears not to belong to the global-minimum basin, this method finds the global minimum more effectively and faster than MD runs from unfolded configurations.

We take here, as initial configurations for MD runs, both high-energy unfolded configurations and already folded CTM configurations, which we suspect are close both in energy and geometry to the real global minimum. We have chosen CTM configurations assuming that adding a short one-bead side chain is unlikely to make great changes in the structure and energy of the lowest configurations of the main 46-bead chain. This 46-bead structure is formed by four strands. Two of them, beads 1–9 and 24–32, contain hydro-

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

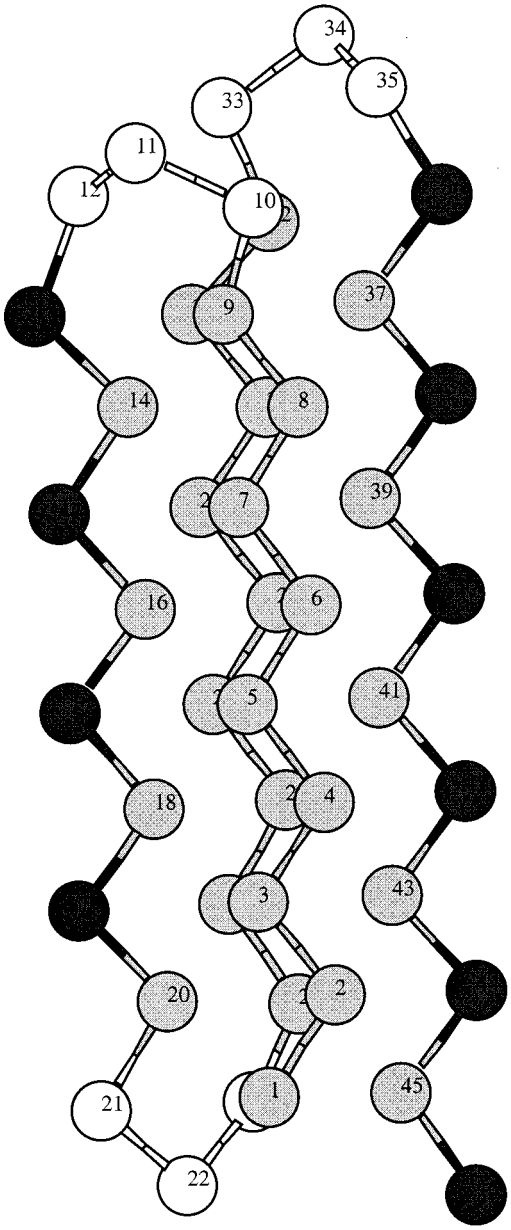

FIG. 1. The β-barrel global minimum structure of 46-bead chain.

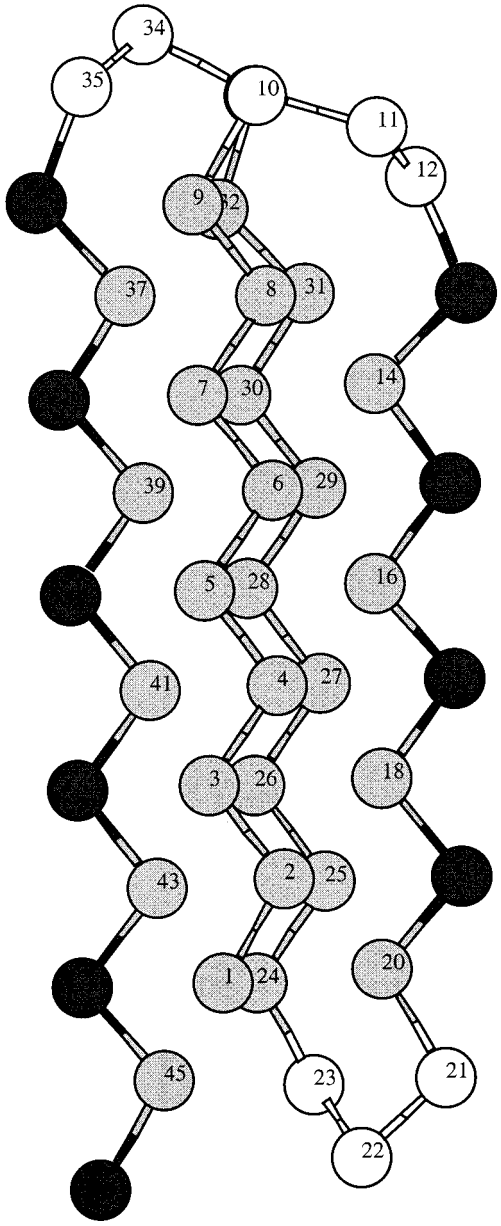

FIG. 2. A structure with *B* strands ordered almost as in Fig. 1, but with the two hydrophobic–hydrophilic (*BL*) strands interchanged.

phobic (*B*) beads only, while the other two (12–20) and (36–46) are constructed of alternating *B* and hydrophilic (*L*) beads. These four strands are connected by three short ''bending junctions'' constructed of neutral (*N*) beads (10–12, 21–23, and 33–35). Hence we have chosen as a CTM configuration the structure whose 46-bead core has the ''pure'' β-barrel form (in the global minimum of the free 46-bead chain, see Fig. 1), and with an extra bead added to some docking site of the main chain at the distance $R0$ which minimizes the local bond energy, and in the geometry which minimizes the local dihedral angle potential. The β-barrel core has, in addition to the structure in Fig. 1, another stable configuration with almost the same ordering of *B* strands, but with the positions of the two *BL* strands interchanged; this is shown in Fig. 2. Because the main contribution to the total potential energy comes from the interaction of the *B* strands, these two configurations of the 46-bead core have similar energies, which implies that one may wish to check related structures to establish their proximity to the real global minimum.

What are the effects of docking the side bead at different sites? The three-color 46-bead model has many ways to add a single-bead side chain: the side chain can be of *B*, *L* or *N* type, and it can be added to any of the 44 beads within the main chain. We are most interested in situations that maximize the effect of the side bead, so we should determine the ''weakest'' (most sensitive) sites in the β-barrel structure. Because the two *B* strands buried inside the folded protein constitute the core of the ordered structure, the protein should be very sensitive to changes (mutations) involving *B* strands. On the other hand, the ''easy-bending'' *N* joints are crucial for the folding, so mutations in their vicinity may also influence greatly the whole picture. For these reasons we

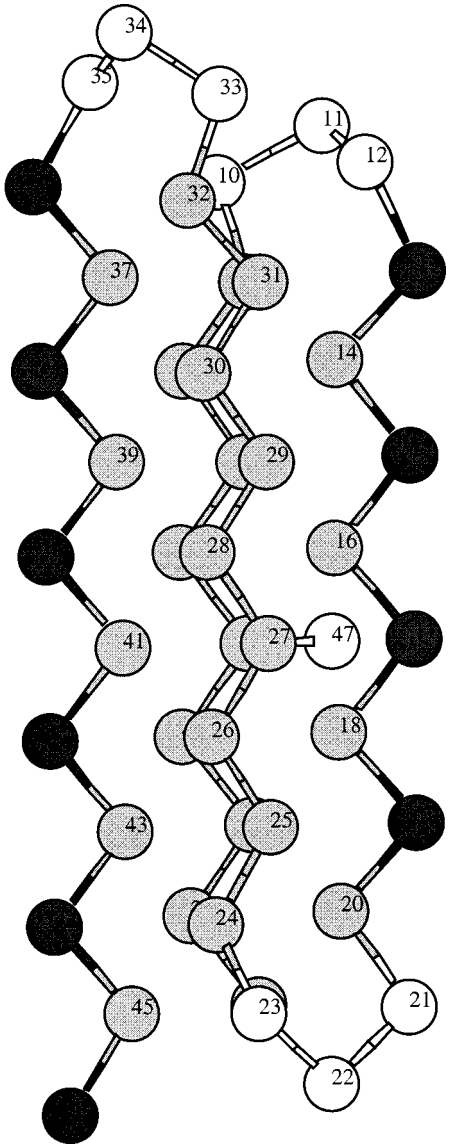

FIG. 3. An assumed close-to-minimum (CTM) structure of the 47-bead protein model with bead 47 attached to the core 46-bead β-barrel structure at site 5.

have concentrated in the present paper on the effects of a side chain attached either to the middle of a *B* strand or close to a bending area.

II. RESULTS

We summarize the results by examining the effects of placement of the side bead, beginning with the placement that has by far the greatest effect on the structures and dynamics of the model. Within each section, we discuss the influence of the side bead on both structure and dynamics.

A. Mutations in the middle of *B* strands

We present the results for a side bead attached at site 5 to the (1–9) *B* strand. Adding the side bead to site 6 or to the (24–32) *B* strand produces essentially the same effects. One of the corresponding CTM configurations is shown in Fig. 3. Having chosen the docking position, we may select any of the three kinds of bead *B*, *L*, or *N*.

Figure 4 shows time histories of the ''quenched energies'' from typical low-temperature MD runs; that is, the sequences are those of the energies of the minima around which the system is oscillating each time the system is interrogated. In these runs, the interrogations occurred every 2000 time steps. The time is a scaled quantity; if masses of all the beads are fixed at 40 D, each time step corresponds to 10^{-14} s. The time per step varies with $m^{1/2}$. The temperature scale is also fixed by the choice of mass and interaction energy. The energy parameters are those of Thirumalai *et al.* With these and a mass of 40, the temperature values given here are in K.

The runs were all initiated from initial CTM structures, and were done with *B*, *L*, and *N* side beads. They illustrate that all three types of augmented structures of the protein model do have stable configurations whose geometries and energies are close to those of the initial structures. Moreover the results show that the protein model finds these configurations by almost barrierless relaxation from those initial CTM structures. However these examples alone do not reveal the most significant consequences of the side bead.

For *B*, *N*, and *L* cases the energies of these minima are -0.51, -0.50, and -0.49 eV, respectively; again, the configurations with *B* and *BL* strands interchanged have very similar energies. The sequence of these energies is simply related to the interbead potentials: attraction for *B* beads, repulsive for *L* (to assure that they stay on the outside of the barrel), and neutral for *N*. At a moderate temperature, the system visits higher-energy configurations, which correspond to the screw-type, rotation–translation structural transformations found previously for the 46-bead model.[12] As Fig. 4 shows, the model with an added *B* bead has a higher probability of occupying an excited configuration than the *L* or *N* mutations, despite the lower temperature of the run; this is a result of a higher density of excited states for the *B* mutation. We discuss this important feature of *B* mutations in some more detail below.

For *L* and *N* cases, neither high-temperature MD runs from CTM configurations nor relaxation from unfolded configurations, as shown in Fig. 5, reveal any minimum with an energy lower than the energy of configurations close to CTM structures. This strongly suggests that the global-minimum structure for *L* and *N* mutations is very close indeed to the initially supposed CTM configuration of Fig. 3, the almost perfect free 46-bead folded core with a side bead at position which minimizes local bond energy and local torsion-angle energy.

With *B* structures the situation is quite different. Although low-temperature runs indicate that there is a stable configuration very close to the initial CTM structure [Fig. 4(a)], some high-temperature runs reveal the existence of configurations with still lower energies, as shown in Fig. 6. Some of the MD runs that start from unfolded structures also find the same lower-energy states; see Fig. 7. We conclude that the structure generated by direct relaxation of the initially supposed CTM configurations corresponds to a local minimum while the global minimum has quite a different structure with the added *B* bead 'buried' inside the *B* strands, as in Fig. 8. Figure 6 illustrates how the transition from a

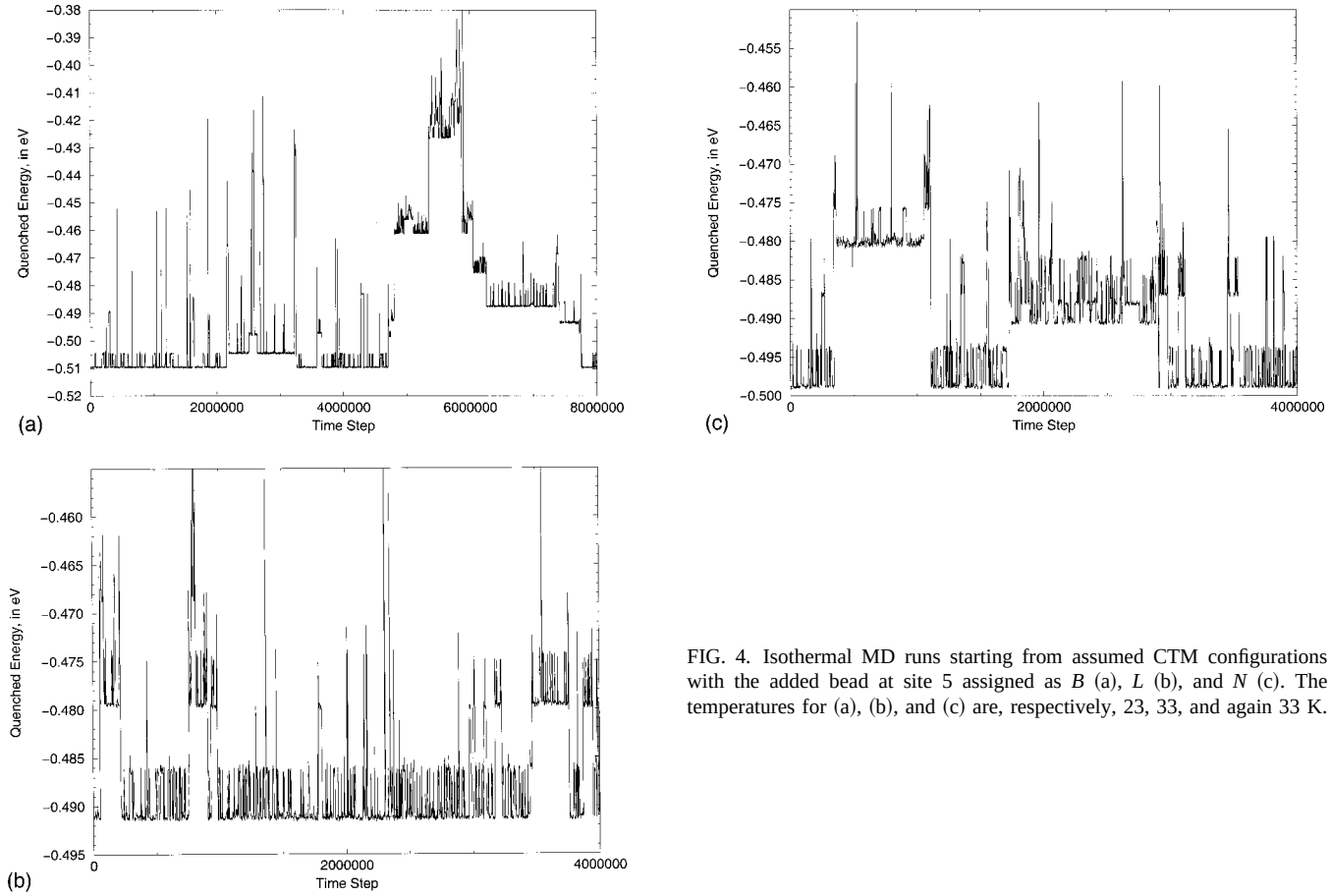

FIG. 4. Isothermal MD runs starting from assumed CTM configurations with the added bead at site 5 assigned as *B* (a), *L* (b), and *N* (c). The temperatures for (a), (b), and (c) are, respectively, 23, 33, and again 33 K.

CTM structure to the global minimum occurs by a rather complicated structural transformation with a rather high barrier.

1. Coexistence phenomena

Many polyatomic systems, clusters for example, may exhibit coexisting phases over ranges of temperature and pressure, due to the small differences between free energies of different phase-like forms. As in all phase transformations that become first order in the large-system limit, one phase-like form gains its stability from its low energy, and another, from its higher density of states and consequent higher entropy. (Such coexistence ''sharpens'' as the number of particles in the cluster increases, so that in systems of macroscopic size, this coexistence is observable only under conditions in which the chemical potentials of the two (or more) forms are equal, the condition giving rise to the traditional first-order phase transition.) We may expect such finite-system behavior also for proteins and many other polymers, in the form of the folding–unfolding transition. When a protein unfolds, we might expect many ''extra'' torsion-angle degrees of freedom to become unfrozen, creating a bulge in the density of states at moderately high energies. However MD runs for the 46-bead model have shown no evidence of such behavior. The reason, we believe, is that the model is ''too well tuned'' insofar as it is designed to have the β-barrel structure as its global minimum; in particular the torsion-angle force constant for *B* and *L* beads has been made rather high, to keep the strands stretched and stiff. This model was designed specifically to imitate solvent contributions in the effective potential. The force constants are, as a result, not particularly close to those of a solvent-free, isolated chain of amino acid residues. Because of the nature of this force field, the angular degrees of freedom become unfrozen only at temperatures much higher than that temperature of the folded–unfolded transition. (This is a well-recognized limitation of the particular model, and we are at work now on a version that will overcome this restriction.)

However the 47-bead protein model, with the addition of its extra side bead in the middle of a *B* chain, loses that precise structural tuning of the β-barrel model. As a result of this modification, the augmented model has ''distorted'' configurations for which the destabilization energy due to changes in the torsion angles is comparable with the stabilization energy accompanying changes in the bead–bead interactions. This circumstance effectively unfreezes the torsion-angle degrees of freedom, and so makes phase coexistence possible. We have indeed found the signature of phase coexistence; it is most evident in *B* mutations, illustrated in the time history of the potential energy of the quenched structures shown in Fig. 9. This phase coexistence occurs in the temperature interval from 39 to 52 K (in the units of the calculation). The excited configurations with energies around -0.50 correspond to structures with totally reconstructed ('melted') cores; see Fig. 10. Each of the ''large-scale'' phase-like forms shows evidence of some sub-

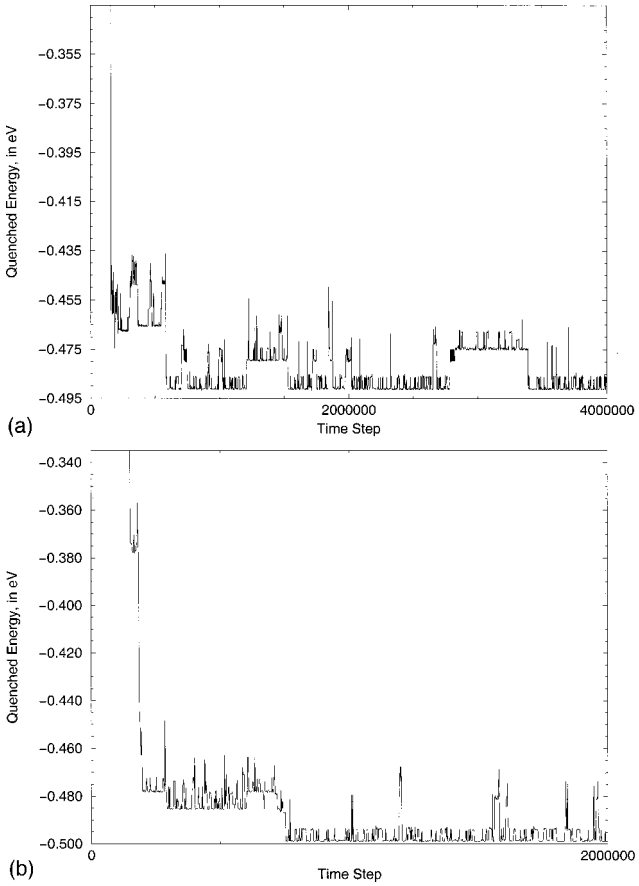

FIG. 5. Typical isothermal MD runs at 33 K for (a) *L* and (b) *N* mutations in the *B*-strand region starting from unfolded configurations and ending at the global minimum.

structure: energies fall between about -0.518 and -0.512 for the global minimum region, and between about -0.495 and -0.487 for the first pronounced set of excited structures.

B. Side bead near the bending region

In another series of simulations, an extra bead (*B, N*, or *L*) has been added to either *B*-site 9 or to *N*-site 10 or 11. The

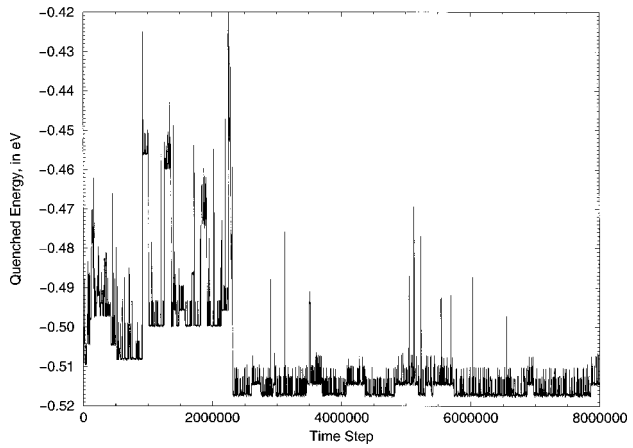

FIG. 6. An isothermal MD run at 33 K for a *B* mutation in the *B* strand area that starts from an assumed CTM configuration but reaches a minimum lower in energy than the CTM configuration, with a high barrier between these low-energy structures.

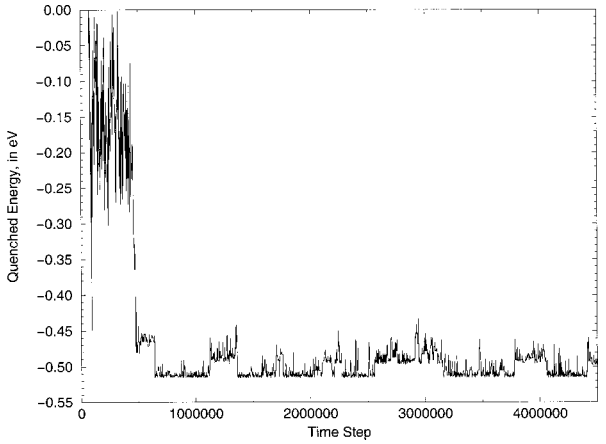

FIG. 7. An isothermal MD run at 40 K for a *B* mutation in the *B* strand area that starts from an unfolded configuration and ends at the global minimum.

results of MD runs starting from both unfolded and folded CTM configurations have shown that the side bead in such locations has considerably less effect on the folded structure of the core 46-bead protein than if it is in the middle of a *B* strand. The reason is that in the bending regions the energy required to change torsion angles is rather small and allows structural rearrangements in the bending area to relieve any stress from the added bead without notable reconstruction of the rest of the structure. When a *B* bead is appended to site 9, in the (1–9) *B* strand, the structural flexibility of that part of the chain shows itself in that the added bead can either participate in coupling between two *B* strands [Fig. 11(a)] or provide an extra nonbonding attraction between the *B* strand and one of *BL* strands [Figs. 11(b) and 11(c)]. Adding a *B* bead to site 9 makes the (1–9) *B* strand effectively longer, providing opportunity for an extended two-step sliding of that strand along another *B* strand, in contrast to the original 46-bead chain in which a one-step screw motion is apparently the lowest-energy path for sliding one strand against the others.

1. Relaxation rates from unfolded configurations

We used both isothermal and annealing (controlled cooling) MD from different unfolded initial configurations to observe the relaxation of the bend-substituted models. As already mentioned, the presence of the extra bead destroys to some extent the well-tuned model of the original 46-bead system adapted particularly for seeking a β-barrel structure. As with the mid-chain substitutions, the side bead increases the density of higher-energy configurations. As a result the pathways to the lowest-energy configurations become more complicated, which becomes apparent during MD runs. Polymers with *L* and *N* side beads relax to low-energy conformations with about the same success rate as the unsubstituted model. However in most runs the polymer with a *B* side bead was trapped in some higher-energy local minimum. If the system is annealed, rather than just allowed to relax isothermally, then all three kinds of substituted cases find their ways to low-lying states with high probability; this is especially striking for the *B*-substituted systems.

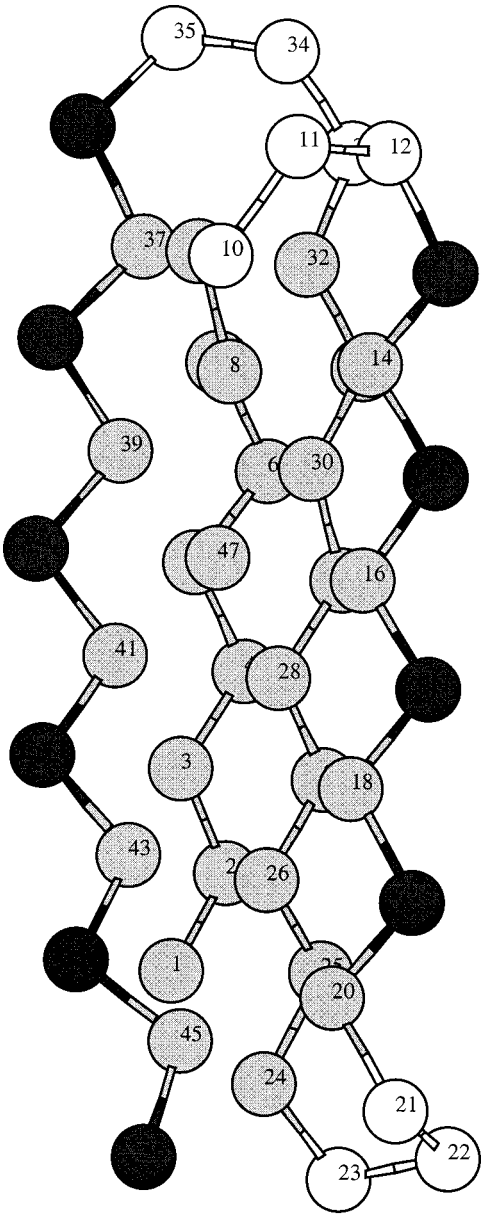

FIG. 8. Geometry of the global minimum structure for *B* mutations in the *B* strand region showing strong rearrangement of CTM structures of Fig. 2. The energy for this structure is -0.518. This figure has the added bead at site 5, but the results are similar for a *B* bead anywhere in the middle part of a *B* strand.

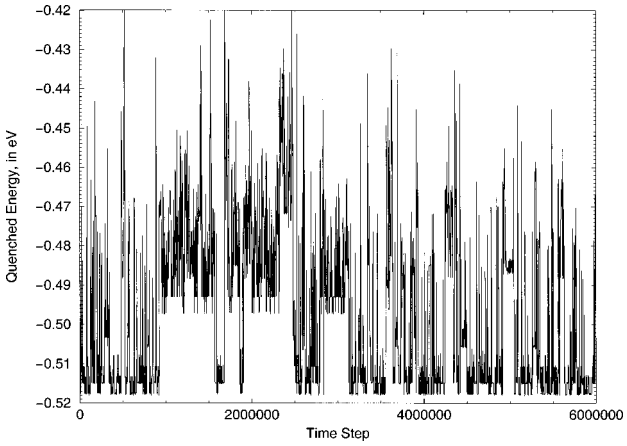

FIG. 9. Time history of the quenched potential energy from an isothermal MD run at 49 K, illustrating the phase coexistence for the model with its extra *B* bead at site 5.

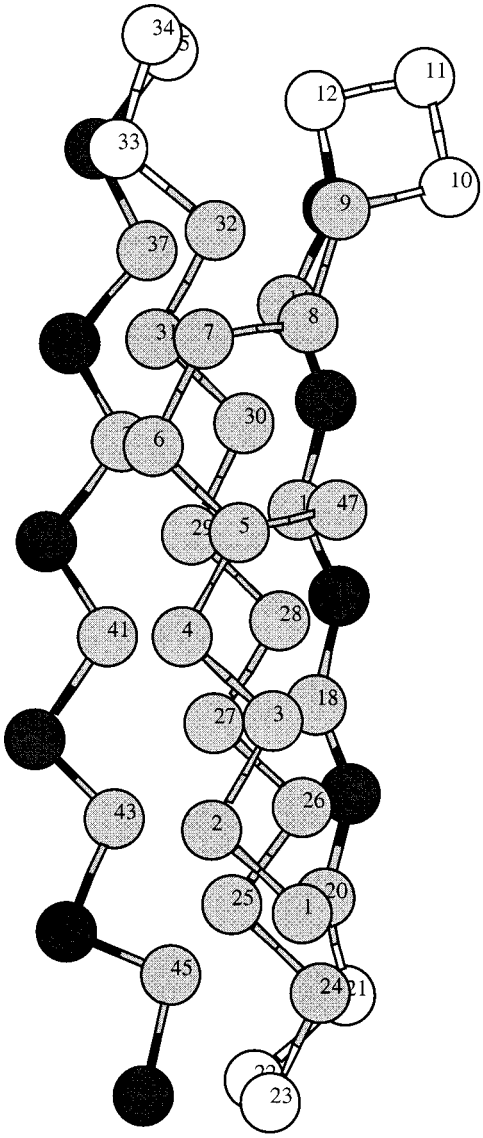

FIG. 10. Structure of excited configurations with energy about -0.49 for an extra *B* bead at site 5.

We have observed that: (1) a *B*-substituted chain folds faster than *L*- and *N*-substituted chains; the more *B* beads the model contains, the stronger the forces are that cause collapse of the chain; (2) whether by annealing or by moderate-temperature isothermal relaxation, protein models with *L* and *N* mutations attain their global minimum structures much more often than do *B*-mutated models. Proteins with mutations in bending areas fold faster than those with an extra bead in the middle of a *B* strand and usually find their global minimum structure readily.

2. Conclusions and discussion

As we have already mentioned, and as Fig. 3 illustrates, attachment of an extra *L* or *N* bead has a rather small impact

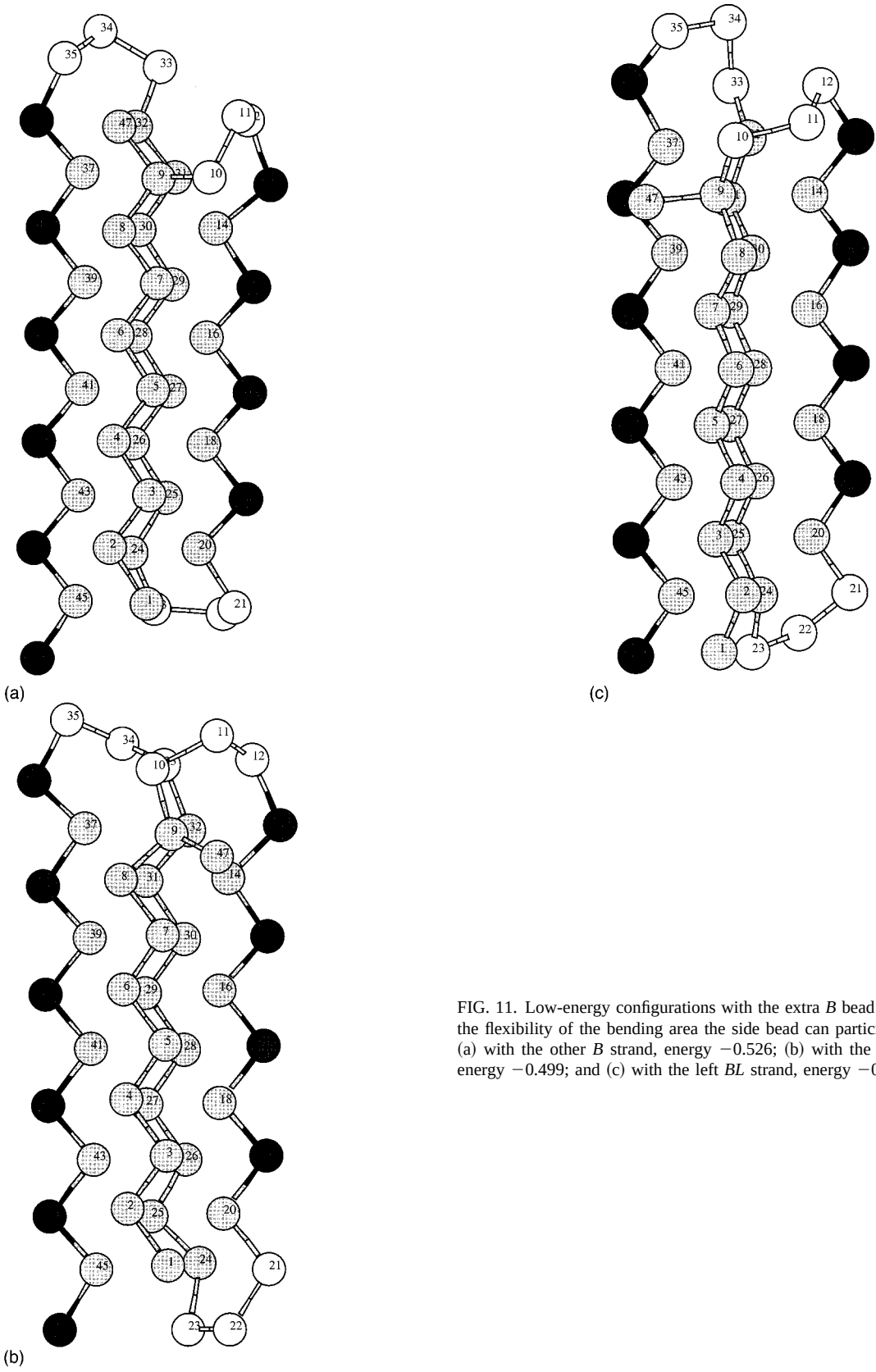

FIG. 11. Low-energy configurations with the extra *B* bead at site 9; due to the flexibility of the bending area the side bead can participate in bonding (a) with the other *B* strand, energy -0.526; (b) with the right *BL* strand, energy -0.499; and (c) with the left *BL* strand, energy -0.515.

upon the folded structures, in the sense that the cores of these structures resemble the global minimum structure of a 46-bead polymer. This is because the low-energy conformations have the hydrophilic *L* bead protruding out of the protein into the solution; were it to ''point inward,'' into the hydrophobic kernel of the β-barrel, it would increase the system's energy considerably. Even a neutral *N* bead pushing into the β-barrel would also increase the energy. Since the interaction of an *N* bead with any of the other beads is weak, the resulting energy change would be determined by the energy loss

due to distortion and breaking of nonbonded attractions of the ideal barrel structure. The excited configurations which contain those distortions have rather high energies compared with those of *B* mutations. That is why *L*- and *N*-mutated structures remain predominantly in their global minimum basins unless the temperatures are rather high (see Fig. 4) and why they do not exhibit phase coexistence as clearly as *B* mutations do.

The addition of an extra hydrophobic bead creates a very different situation. Because *B* beads have strong attractions for each other, the increase in stabilizing energy caused by intrusion of the side bead into the main chain barrel structure can be of the same order as the energy loss due to deformation of that structure. The more *B* beads in the vicinity of the intruding *B* bead, the easier the intrusion is, which is why we observe this phenomenon when the extra *B* bead is added in the middle of a *B* strand, and why such deformed structures are not so prominent when the side bead is attached near or in a bending region. As a result of side-bead penetration into the *B* strand core, the global minimum structure of the system with a *B* bead added to the middle of a *B* strand differs greatly from the pure β barrel. For the same reason, this type of mutation generates a large number of excited configurations with low energies, the circumstance responsible for the realization of phase coexistence and a high probability of trapping in local high-energy minima.

The same considerations explain the general features of the relaxation rates mentioned above. The *B*-mutated systems, with the new bead in the middle of a *B* strand, do not tend to relax to their global minima and are often trapped at higher local minima; proteins with mutations close to (or inside) the bending areas relax much more readily to their global minima and their relaxation does not depend sensitively on the type of added bead.

These results may give some added insight at the molecular level into how some kinds of mutations may produce nonviable effects in proteins. However it must be remembered that this study is based on a simplified, rather unrealistic model, and that the results must be considered suggestive rather than demonstrative.

ACKNOWLEDGMENT

This research was supported by a grant from the National Science Foundation.

[1] C. L. Brooks, M. Petitt, and M. Karplus, *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics* (Wiley, New York, 1988).
[2] J. A. McCammon and S. C. Harvey, *Dynamics of Proteins and Nucleic Acids* (Cambridge University Press, Cambridge, 1988).
[3] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, Proteins: Struct., Funct., Genet. **21**, 167 (1995).
[4] D. Thirumalai and S. A. Woodson, Acc. Chem. Res. **29**, 433 (1996).
[5] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes, Ann. Rev. Phys. Chem. **48**, 539 (1997).
[6] J. Skolnick, A. Kolinski, and R. Yaris, Proc. Natl. Acad. Sci. USA **85**, 5057 (1988).
[7] J. Skolnick, A. Kolinski, and R. Yaris, Biopolymers **28**, 1059 (1989).
[8] A. Sikorski and J. Skolnick, Biopolymers **28**, 1097 (1990).
[9] A. Sikorski and J. Skolnick, J. Mol. Biol. **212**, 819 (1990).
[10] J. D. Honeycutt and D. Thirumalai, Proc. Natl. Acad. Sci. USA **87**, 3526 (1990).
[11] Z. Guo, D. Thirumalai, and J. D. Honeycutt, J. Chem. Phys. **97**, 525 (1992).
[12] R. S. Berry, N. Elmaci, J. P. Rose, and B. Vekhter, Proc. Natl. Acad. Sci. USA **94**, 9520 (1997).
[13] K. D. Ball, R. S. Berry, A. Proykova, R. E. Kunz, and D. J. Wales, Science **271**, 963 (1996).
[14] N. Elmaci and R. S. Berry, J. Chem. Phys. **110**, 10606 (1999).
[15] B. Vekhter and R. S. Berry, J. Chem. Phys. **110**, 2195 (1999).
[16] M.-H. Hao and H. A. Scheraga, J. Phys. Chem. **98**, 4940 (1994).
[17] M.-H. Hao and H. A. Scheraga, J. Phys. Chem. **98**, 9882 (1994).
[18] M.-H. Hao and H. A. Scheraga, Proc. Natl. Acad. Sci. USA **93**, 4984 (1996).
[19] F. H. Stillinger and T. A. Weber, Kinam **3**, 159 (1981).
[20] F. H. Stillinger and T. A. Weber, Phys. Rev. A **25**, 978 (1982).
[21] F. H. Stillinger and T. A. Weber, Phys. Rev. A **28**, 2408 (1983).