Targeted molecular dynamics: A new approach for searching pathways of conformational transitions

J. Schlitter

Institut für Biophysik, Ruhr-Universität Bochum, Bochum, Germany

M. Engels* and P. Krüger

Institut für Biochemie, Rheinisch-Westfälische Technische Hochschule Aachen, Aachen, Germany

Molecular dynamics simulations have proven to be a valuable tool to investigate the dynamic behavior of stable macromolecules at finite temperatures. However, considerable conformational transitions take place during a simulation only accidentally or at exceptionally high temperatures far from the range of experimental conditions. Targeted molecular dynamics (TMD) is a method to induce a conformational change to a known target structure at ordinary temperature by applying a time-dependent, purely geometrical constraint. The transition is enforced independently of the height of energy barriers, while the dynamics of the molecule is only minimally influenced by the constraint. Simulations of decaalanine and insulin show the ability of the method to explore the configurational space for pathways accessible at a given temperature. The transitions studied at insulin comprise unfolding of an α-helical portion and, in the reverse direction, refolding from an extended conformation. A possible application of TMD is the search for energy barriers and stable intermediates from rather local changes up to protein denaturation.

Keywords: conformational transition, MD simulation, insulin

INTRODUCTION

The occurrence of conformational transitions is a common event in chain molecules. In biomolecules, the changes range from local transitions to folding and unfolding of the native conformation. For proteins, conformational changes are known to play an important functional role. They are affected by a number of parameters such as aggregation states, solution condition and even details of the amino acid sequence. For a better understanding of experimental findings, it is necessary to characterize the possible pathways of a transition by a sequence of stable or metastable intermediates and the energy barriers governing the evolution of the process. An increasing amount of experimental information of this kind is available, in particular, for protein folding. (For a recent review, see Schmid.[1])

In spite of extensive experience in computer simulation of macromolecules, the numerical investigation of pathways is still a challenging problem. It is possible, at least in principle, to simulate a conformational change by using (for instance) molecular dynamics (MD) simulation. Unfortunately, the large number of degrees of freedom, the size of the conformational changes usually encountered in "real" systems and the time scale are serious problems for applying familiar methods. Therefore, attempts have been made to tackle the search problem by activated dynamics,[2] by using higher temperatures,[3] or by starting with an initial guess about the pathway between the limiting structures.[4,5] Recently we have developed two new approaches in order to calculate pathways between known structures—one is based on energy minimization, the targeted energy minimization (TEM) method,[6] and the other is based on MD simulation, called *targeted molecular dynamics simulation* (TMD).[7] TMD uses a time-dependent, purely geometrical constraint to find the target structure. It allows a search in the conformational space by constraining the system only minimally. The simulations can be performed at ordinary temperatures; there is no need to activate the transition using high temperatures.

This paper reports on the calculation of pathways using the TMD method applied to different examples. The first example describes a decaalanine helix switching from an alpha to a 3_{10}-helix. This demonstrates a small conformational change of a simple molecule using modeled structures for both the initial and the target conformation. The second example, the T ↔ R transition in insulin,[8,9] represents a more extensive change superimposed by the influence of the

*Present address: MPI für Molekulare Physiologie, Rheinlanddamm 201, D-44139 Dortmund, Germany.
Address reprint requests to Professor Schlitter at the Institut für Biophysik, ND 04, Ruhr-Universität Bochum, D–44780 Bochum, Germany.
Received 1 October 1993; accepted 20 October 1993

surrounding protein matrix. For decaalanine, the transition from alpha to 3_{10} has been calculated and for insulin, the transitions T → R and R → T have been calculated. The numerical results were analyzed with respect to the pathways. For insulin, a detailed discussion of the energy has been given previously.[7]

METHOD

The TMD method[7] was designed as a tool for simulating the conformational transition of a molecule from a given initial conformation, *I*, to an also known target conformation, *F*. On the way from *I* to *F*, each configuration of the molecule is given by a vector, $\mathbf{x} = (x_1 \ldots x_{3N})^T$, containing the 3N Cartesian coordinates of the position vectors $\mathbf{r}_1 \ldots \mathbf{r}_N$ of the individual atoms. (In accordance with the usage in statistical mechanics, configuration here denotes a coordinate vector in the configuration space. The term *conformation* denotes a macromolecular structure including the thermal deviations from its mean configuration.) $\mathbf{x}_I$ and $\mathbf{x}_F$ are the configurations representing the conformations *I* and *F*, respectively. As shown in Figure 1, one may define for each configuration, $\mathbf{x}$, its distance, ρ, to the target configuration.

$$\rho = |\mathbf{x} - \mathbf{x}_F| = [\Sigma(x_i - x_{Fi})^2]^{1/2} \qquad (1)$$

In contrast to conventional MD simulations where the distance, ρ, is a fluctuating quantity that may go to zero if the transition to *F* occurs accidentally, the present method uses ρ as a control parameter in order to force the system to undergo the desired transition. This is achieved by introducing the constraint

$$\phi(\mathbf{x}) \equiv |\mathbf{x} - \mathbf{x}_F|^2 - \rho^2 = 0 \qquad (2)$$

which results in an additional constraint force

$$\mathbf{F}^c \equiv \lambda \, d\phi/dx = 2\lambda \, (\mathbf{x} - \mathbf{x}_F) \qquad (3)$$

λ is a Lagrange parameter that has to be chosen in accordance with Equation (2).

The TMD simulation of a conformational transition is performed as follows:

(1) Set $\rho = \rho_0 \equiv |\mathbf{x}_I - \mathbf{x}_F|$.
(2) Choose initial coordinates $x_i(0) = x_{Ii}$ and appropriate initial velocities.
(3) Solve, numerically, the equations of motion with the additional constraint force, $\mathbf{F}^c$.
(4) After each time step Δt, diminish ρ by $\Delta\rho = (\rho_0 - \rho_f)\,\Delta t/t_s$, where t_s is the total simulation time. At the end of the simulation a final distance ρ_f is reached.

It is easily recognized in Figure 1 that the monotonous decrease of ρ forces the system to find a pathway from $\mathbf{x}_I$ to some final configuration that has a given, small distance, ρ_f, from the target configuration, $\mathbf{x}_F$. However, as only one of 3N coordinates, ρ, is kept constant in each step, the system has maximum freedom to explore the configurational space in accordance with the force field and environmental parameters such as temperature and pressure, which may be set for the simulation. Solvents or other molecules not being subject to a constraint can be included if necessary. It is also possible to keep ρ constant during a whole run in order to study the dynamics of an intermediate conformation.

A particular set of Cartesian coordinates was chosen for the visual presentation of pathways. First, the initial configuration $\mathbf{x}_I$ is used to define the origin of the new coordinate system and an *r*-axis is chosen in such a way that $\mathbf{x}_F$ lies on the positive r-axis at $r = r_F = \rho_0$. The remaining coordinates are denoted by $s_1 \ldots s_{3N-1}$. In Figure 1, a fictitious pathway is displayed as a projection onto the plane spanned by r and one of the s_i.

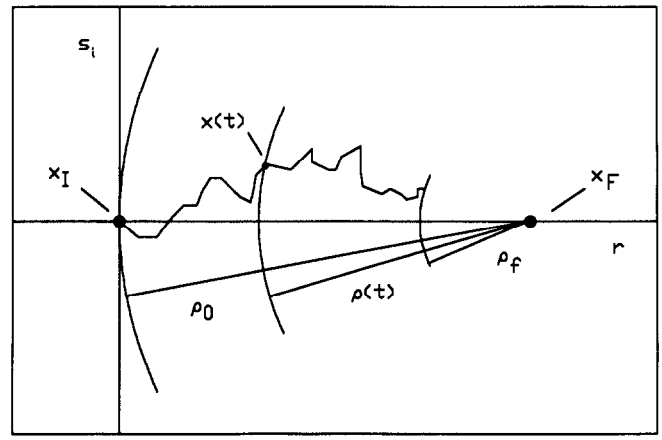


Figure 1. Geometry of the conformational transition from $\mathbf{x}_I$ to $\mathbf{x}_F$. The r-axis is chosen in such a way that the initial configuration defines the origin and the final configuration lies on the positive branch. s_i is one of the residual Cartesian coordinates. A TMD run generates configurations $\mathbf{x}(t)$ lying on hyperspheres with radius $\rho(t)$ centered in $\mathbf{x}_F$. The quasicontinuous contraction of the hypersphere from a radius ρ_o to ρ_f enforces the conformational transition.

Next, a quantity $s \equiv (\Sigma\, s_i^2)^{1/2} \geq 0$ is defined, which obviously is the distance of a configuration from the r-axis. Thus, an appropriate two-dimensional presentation $[r(t), s(t)]$ is obtained that is used in the following plots of calculated pathways. For any intermediate configuration on a pathway, the distances to the initial and the final configurations in configuration space are identical with the corresponding distances measured in the (*r*,*s*) plot, which can be seen by simple vector algebra. We note in passing that the monotonous approach of the pathway to the target point in the (r,s) plot does not imply that any individual coordinate must approach its target value in this way. The use of a shrinking value of ρ solely means that this must hold *on an average* for the Cartesian coordinates, which reflects the enormous freedom the system has in spite of the constraint. For instance, single amino acids may rotate or make way for each other by forward-backward motions.

The simulations presented in the following section were performed with the GROMOS program package.[10] For treating the constraint, an additional subroutine called SHIFT and minor modifications in other routines were included. The acutal implementation of the TMD algorithm also takes into consideration the problem of translational fluctuations, which are possibly induced by the constraint force in Equation (3). As shown previously,[7] the center of mass position remains unaffected when the constraint is reformulated in terms of mass-weighted coordinates and $\mathbf{x}_I$ and $\mathbf{x}_F$ have the same center of mass.

RESULTS AND DISCUSSION

Decaalanine

The small peptide decaalanine was used as a model system for developing TMD. It contains 53 atoms, including charged hydrogens explicitly taken into account (others being included in united atoms). Because helical conformations of small polypeptides are known to be unstable in solution at room temperature, the simulations were performed at temperatures $T \leqq 100$°K *in vacuo*, where a helix is expected to be stable due to internal hydrogen bonds. Different secondary structures were generated by molecular modeling (Insight, Biosym, San Diego, CA) and first investigated by MD runs at 100°K with the GROMOS vacuum force field and weak coupling to a heat bath ($\tau = 0.01$ ps). The covalent bond lengths of the hydrogens were kept constant using SHAKE. The only stable structures were the 3_{10} (at least more than 50 ps) and the α-helix (at least more than 200 ps), which seems to be the stable ground state of decaalanine under these conditions. The equilibrated α-helix was used as the starting configuration and the modeled 3_{10}-helix as the target configuration.

The transition from the α- to the 3_{10}-helix requires rearrangement of the six existing H-bonds and formation of one additional H-bond. The visual control of the process during the TMD run shows a flickering pattern of H-bonds that is gradually shifted to the one of the target structure. In the α-helix, only the first and the two last H-bonds exhibit some instability. During the simulated transition, the instability increases from both ends while the two central H-bonds remain stable over almost one-third of the simulation time. In Figure 2, this is reflected by the energy, which shows an increase and increasing fluctuations. After a short restabilization of the α-helical H-bonds near $\rho = 1$ nm, maximum energy with maximum fluctuations is reached when the central H-bonds change to the new pattern characteristic of the 3_{10}-helix. During the last third of the run, this pattern is stabilized and the energy is about 20 kJ/mol over the value for the α-helix. The energy of activation is 25 kJ/mol or about 30 times the thermal energy, RT, at the temperature given.

The pathway of the transition in Figure 3 shows two remarkable features that are characteristic of a transiton simulated during a TMD run. In the first place, the pathway is far from being linear, but deviates appreciably from the straight line between $\mathbf{x}_I$ and $\mathbf{x}_F$. Apart from fluctuations, it runs first almost linearly from the initial configuration to the one with the highest energy which lies at a distance of 1 nm off the straight line from $\mathbf{x}_I$ to $\mathbf{x}_F$. After making a z-shaped path, it runs again linearly in the direction of the final configuration. Second, the path is rather narrow in its first part and becomes very broad before it contracts again near the end. Both features demonstrate the freedom left to the molecule, although the applied constraint is strong enough to force the conformational transition. The unexpected width of the pathway must be interpreted as a not undesired effect of the simulation at a finite temperature. Instead of generating one smooth track that does not have any statistical meaning, TMD "hatches" a broad track that covers a continuum of possible individual pathways. This interpretation is supported by the observation that, indeed, runs at lower temperatures yield narrower pathways and runs at shorter simulation times yield a bundle of smoother tracks.

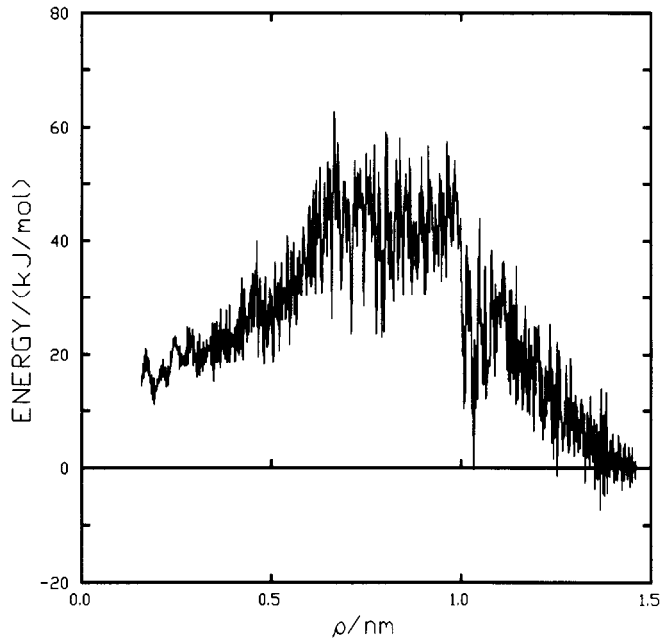


Figure 2. Energy profile of the transition $\alpha \rightarrow 3_{10}$ in decaalanine generated by a TMD run of 100 ps at 100°K.

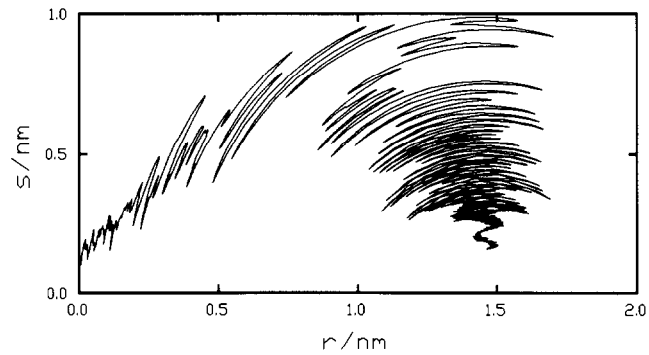
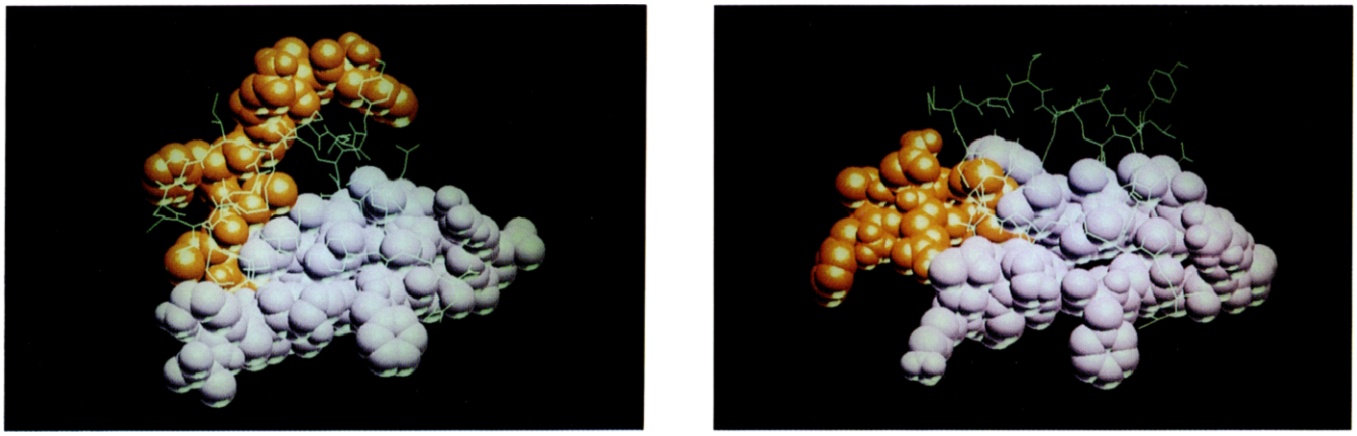


Figure 3. Pathway of the transition $\alpha \rightarrow 3_{10}$ in decaalanine, 10 ps at 100°K. The initial distance, ρ_0, in configurational space is 1.46 nm; the rms distance per atom is 0.18 nm (in mass-weighted coordinates).

Insulin

The T $\leftrightarrow$ R transition in insulin is a suitable example to study a more extensive conformational change. Both limiting structures named T and R are known from X-ray analysis of four crystal modifications of hexameric zinc insulin. The promoter composition found is T_6 in the rhombohedral 2Zn insulin,[11] T_3R_3 in the rhombohedral 4Zn insulin[12] and R_6 in monoclinic insulin as well as in a further rhombohedral form.[13,14] Transformation from T_6 to T_3R_3 is promoted by high concentrations of halide ions and $T_6 \rightarrow R_6$ transformation is promoted by phenolic compounds.[8,9,15–19] In order to reduce the system size, we focus our attention on the transi-



Color Plate 1. Limiting structures T (left) and R (right) of insulin. The eight N-terminal B-chain residues immediately involved in the conformational change are highlighted in orange; the A-chain is represented by a wire model. In comparison with Figure 4, the molecule is shown here from the rear in order to display the interchain contacts.

tion of one promoter. Detailed comparisons of the structures of T and R are given elsewhere.[20,21] The most striking difference is present at the N-terminus of the B-chain. In T, the first eight residues are extended and packed against the protein, while in R they are helical and directed away from the promoter core (Color Plate 1).

Two simulations of 200 ps were performed using the GROMOS vacuum force field—one from T to R, the other from R to T. The energy minimized and superimposed X-ray structures of molecule 2 (Chinese nomenclature, Cutfield, et al.[22]) of rhombohedral porcine 2ZN (T) and 4Zn insulin (R) serve as initial and final structures. SHAKE was applied in the MD simulations to keep the bond lengths constant. The simulations were started at a temperature of 300°K and the velocities were taken from the corresponding Maxwell distribution. The system was coupled to a heat bath with a relaxation time of 0.01 ps.

Figure 4 shows the transitions in both directions using the (r,s) plot. For the transition in the R → T direction, $\mathbf{x}_I$ and $\mathbf{x}_F$ have been reversed. Both pathways are bell shaped in the (r,s) plane. The maximal s-values are visible at $r = 7.5$ nm with more or less the same values for both directions. The value s indicates deviation from the line connecting the initial and final structures. The same value for both curves (forward and backward) does not necessarily mean that the structures are identical. But on the other hand, if s is different, it can be concluded that the structures do not overlap. There are differences between the two curves, which is the first indication that the underlying pathways are not identical.

Additional details of the transition are illustrated in Figure 5. It shows two series of subsequent structures taken from the T → R and R → T simulations. The helix-extended transition of the eight N-terminal residues of the B-chain is simultaneously superimposed by a swivel motion of the whole segment. This motion starts in the T → R direction with the four residues B1 to B4, which in the further course of the simulation seizes all eight N-terminal residues. In the R → T direction, the first event is a bending of the B1-B19 helix at B8, B9 and B10 where it kinks afterward. The whole

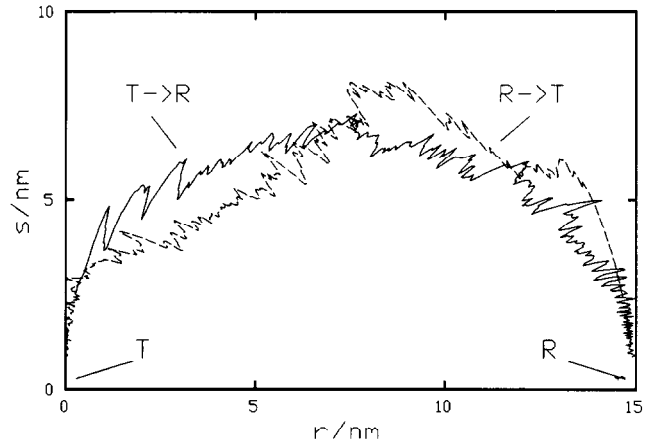


Figure 4. Pathways of the transitions T → R and R → T in monomeric insulin as generated by 200 ps TMD runs at 300°K. The initial distance ρ_0 is 14.9 nm; the rms distance per atom is 0.67 nm (in mass-weighted coordinates). In the R → T transition, the path runs from right to left as $\mathbf{x}_I$ and $\mathbf{x}_F$ have been exchanged.

N-terminal segment swivels in the direction of its final position at which the N-terminal helix is maintained. Thereafter, unwinding starts from B1.

In addition, the conformational change is accompanied by a forward-backward motion of the A_N-helix (A1-A9) and the interhelical segment A10-A12. This motion is caused by the rearrangement of the N-terminal B-chain segment. The transient displacement of the A-chain residues demonstrates that the monotonous decrease of the distance ρ to the target structure does not imply that individual groups or atoms are forced to perform unidirectional motion. The time-dependent constraint leaves the molecule an enormous freedom to approach the target structure in accordance with the dynamics determined by the force field and the temperature.

Although the series of events are different in either direc-

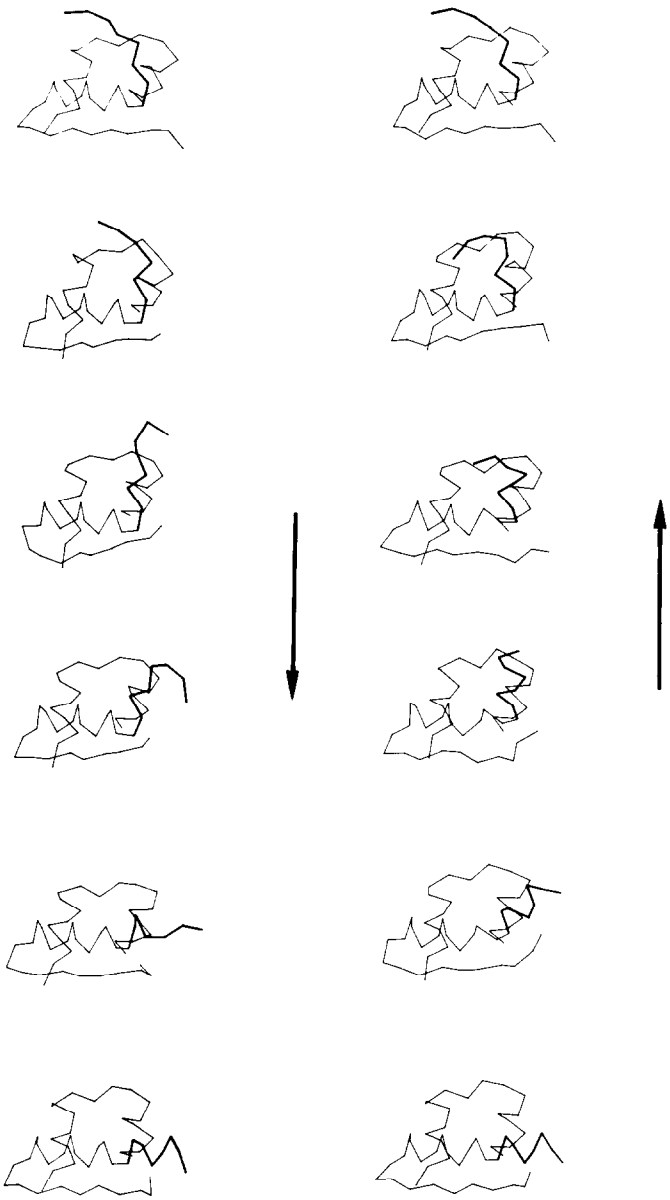


Figure 5. Two series of transient structures taken from the $T \rightarrow R$ and $R \rightarrow T$ transition (top = T, bottom = R). Only the C_α atoms are shown; the eight N-terminal residues of the B-chain are connected by bold lines.

tion, the potential energy profile of pathways are similar, as shown previously.[7] This indicates that more than one pathway can exist for a conformational change that differs only slightly in its energy profile. It seems that mainly two features of the simulation method lead to the different pathways. First, it selects only pathways with a monotonously decreasing distance from the target structure. Second, every MD step onto the next hypershere in the TMD procedure tends to find a configuration that can be reached from the previous one at the minimum cost of energy, thus fulfilling a rather local criterion. Both features contribute to some kind of directedness, which may explain the different pathways found for forward and backward transition.

CONCLUSIONS

The simulation calculations performed on decaalanine and the insulin monomer using TMD have shown that it is possible to generate reasonable pathways for considerable conformational transitions in macromolecules. The method makes use of a minimal geometrical constraint (ensuring that the transition takes place during the simulation), but on the other hand leaves the molecule as much freedom as possible to evolve its dynamics at the temperature given. The first aspect is demonstrated by the large energy barrier of 30 RT that decaalanine crosses when going from the α- to the 3_{10}-helix. It would hardly be possible to observe such a process in a conventional MD simulation, but it can be investigated by TMD because there is no thermal control of the progress made in the coordinate ρ, which may be considered a preliminary reaction coordinate. By way of contrast, a molecule is free to move in all other 3N-1 coordinates and hence to explore the configurational space for a suitable path. The latter is determined by the temperature (and other environmental influences like pressure and solvent not yet included so far) and turns out to be a broad way covering a continuum of possible individual pathways that might be found, for instance, by energy minimization methods. In view of the well-known character of the potential energy surface of a protein consisting of a large number of shallow, local minima separated by flat barriers,[23] this kind of broad temperature-dependent pathway yields a more adequate description of the statistical character of a conformational transition.

The results of the insulin simulations show that essentially different pathways exist, along which the same conformational change can be achieved. In the present example, they were generated for different directions of the transition. It is expected that, at least in sufficiently complex molecules, several pathways can also be found by varying starting coordinates and velocities or by using higher temperatures and longer simulations times. The only restriction imposed to pathways detectable by TMD is the monotonous decrease of ρ, which means that the rms distance of all atoms from their position in the target structure must shrink monotonously during the simulated conformational change. It will be interesting to see in additional applications and in comparison with other methods (like high-temperature MD) if this restriction leaves relevant pathways out of consideration.

Monitoring the energy in the course of a transition is only the first approach to assessing the importance of a particular pathway by its mean energy and the height of barriers (if they can be resolved in the noisy energy profile). The complete characterization of a pathway requires calculation of the *free* energy profile, which is much more difficult to obtain.

After the first experiences with TMD, it is expected that a wide spectrum of applications from local conformational changes up to protein denaturation can be treated in this way. The study of protein denaturation can give an idea of the reverse process—the folding of a protein—which is certainly more difficult to investigate due to the large number of nonproductive pathways characterized by extremely high activation barriers. Apart from the search for pathways providing insight into the underlying mechanisms, energy

barriers and stable intermediates can be found that can be equilibrated for comparison with experimentally accessible structures.

ACKNOWLEDGMENTS

We would like to thank Prof. Dr. A. Wollmer for constant support and encouragement. Thanks are also due to Prof. Dr. G. Dodson for the coordinates of 2Zn and 4Zn insulin. C. Bösinger is thanked for helpful technical assistance. This work was supported by a grant of computing time from the HLRZ Juelich, Germany (HBO 021).

REFERENCES

1 Schmid, F.X. *Curr. Opin. Struct. Biol.* 1992, **2,** 21
2 McCammon, J.A. and Karplus, M. *Biopolymers* 1980, **19,** 1375
3 Tirado-Rives, J. and Jorgensen, W.L. *Biochemistry* 1991, **30,** 3864
4 Elber, R. and Karplus, M. *Chem. Phys. Lett.* 1987, **139,** 375
5 Ech-Cherif El-Kettani, M.A. and Durup, J. *Biopolymers* 1992, **32,** 561
6 Engels, M., Jacoby, E., Krüger, P., Schlitter, J. and Wollmer A. *Protein Engineering* 1992, **5,** 669
7 Schlitter, J., Engels, M., Krüger, P., Jacoby, E. and Wollmer, A. *Mol. Sim.* 1993, **10,** 291
8 Renscheidt, H., Strassburger, W., Glatter, U., Wollmer, A., Dodson, G.G. and Mercola, D.A. *Eur. J. Biochem.* 1984, **142,** 7
9 Wollmer, A., Rannefeld, B., Johansen, B.R., Hejnaes, K.R., Balschmidt, P. and Hansen, F.B. *Biol. Chem. Hoppe-Seyler,* 1987, **368,** 903
10 van Gunsteren, W.F. and Berendsen, H.J.S. Program System GROMOS 87. BIOMOS biomolecular software b.v., Institute of Biophysical Chemistry, University of Gronongen, NL, 1987
11 Baker, E., Blundell, T.L., Cutfield, J.F., Cutfield, S.M., Dodson, E.J., Dodson, G.G., Hodgkin, D.M.C., Hubbard, R.E., Isaacs, N.W., Reynolds, C.D., Sakabe, N. and Vijayan, N.M. *Philos. Trans R. Soc. London* 1988, **B319**, 369
12 Bentley, G., Dodson, E.L., Dodson, G.G., Hodgkin, D.C. and Mercola, D.A. *Nature* 1976, **261,** 166
13 Derewenda, U., Derewenda, Z., Dodson, E.J., Dodson, G.G., Reynolds, C.D., Smith, G.D., Sparks, C. and Swenson, D. *Nature* 1989, **338,** 594
14 Smith, G.D. and Dodson, G.G. *Biopolymers* 1992, **32,** 441
15 Thomas, B. and Wollmer A. *Biol. Chem. Hoppe-Seyler,* 1989, **370,** 1235
16 Roy, M., Brader, M.L., Lee, R.W.-L., Kaarsholm, N.C., Hansen, J.F. and Dunn, M.F. *J. Biol. Chem.* 1989, **264,** 19081
17 Brader, M.L., Kaarsholm, N.L., Lee, R.W.-L. and Dunn, M.F. *Biochemistry* 1991, **30,** 6636
18 Gross, L. and Dunn, M.F. *Biochemistry.* 1992, **31,** 1295
19 Brader, M.L., Borchardt, D. and Dunn, M.F. *Biochemistry* 1992, **31,** 4691
20 Dodson, E.J., Dodson, G.G., Reynolds, C.D. and Vallely, D. In *A comparison between the insulin molecules in 2-zinc and 4-zinc insulin crystals* (D. Brandenburg and A. Wollmer, Eds.) Walter de Gruyter & Co, Berlin (1980) 9–16
21 Chothia, C., Lesk, A.M., Dodson, G.G. and Hodgkin, D.C. *Nature* 1983, **302,** 500
22 Cutfield, J.F., Cutfield, S.M., Dodson, E.J., Dodson, G.G., Reynolds, C.D. and Vallely, D. In *Structural Studies on Molecules of Biological Interest* (G. Dodson, J. Glusker and D. Sayre, Eds.) Clarendon Press, Oxford (1981) 527–546
23 Elber, R. and Karplus, M. *Science* 1987, **235,** 318