Unfolding and Refolding of the Native Structure of Bovine Pancreatic Trypsin Inhibitor Studied by Computer Simulations†

M.-H. Hao,‡ M. R. Pincus,§ S. Rackovsky,‖ and H. A. Scheraga*,‡

Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853-1301, Department of Pathology, State University of New York, Health Science Center, 750 East Adams Street, Syracuse, New York 13210, and Department of Biophysics, School of Medicine and Dentistry, University of Rochester, Rochester, New York 14642

Received April 23, 1993; Revised Manuscript Received June 23, 1993®

ABSTRACT: A new procedure for studying the folding and unfolding of proteins, with an application to bovine pancreatic trypsin inhibitor (BPTI), is reported. The unfolding and refolding of the native structure of the protein are characterized by the dimensions of the protein, expressed in terms of the three principal radii of the structure considered as an ellipsoid. A dynamic equation, describing the variations of the principal radii on the unfolding path, and a numerical procedure to solve this equation are proposed. Expanded and distorted conformations are refolded to the native structure by a dimensional-constraint energy minimization procedure. A unique and reproducible unfolding pathway for an intermediate of BPTI lacking the [30,51] disulfide bond is obtained. The resulting unfolded conformations are extended; they contain near-native local structure, but their longest principal radii are more than 2.5 times greater than that of the native structure. The most interesting finding is that the majority of expanded conformations, generated under various conditions, can be refolded closely to the native structure, as measured by the correct overall chain fold, by the rms deviations from the native structure of only 1.9–3.1 Å, and by the energy differences of about 10 kcal/mol from the native structure. Introduction of the [30,51] disulfide bond at this stage, followed by minimization, improves the closeness of the refolded structures to the native structure, reducing the rms deviations to 0.9–2.0 Å. The unique refolding of these expanded structures over such a large conformational space implies that the folding is strongly dictated by the interactions in the amino acid sequence of BPTI. The simulations indicate that, under conditions that favor a compact structure as mimicked by the volume constraints in our algorithm, the expanded conformations have a strong tendency to move toward the native structure; therefore, they probably would be favorable folding intermediates. The results presented here support a general model for protein folding, i.e., progressive formation of partially folded structural units, followed by collapse to the compact native structure. The general applicability of the procedure is also discussed.

The protein folding problem involves several questions, *viz.*, (1) Why does a protein fold to a unique native structure rather than to other low-energy conformations? (2) What kind of interactions stabilize the native structure? (3) By what pathway does a protein find its native structure? Since proteins with different amino acid sequences usually adopt different native structures, presumably some common rules govern the acquisition of such different structures. Therefore, insights into the folding process of a particular well-characterized structure are relevant to the general problem of protein folding.

Early efforts to calculate the native structures of proteins focused on local and nonlocal *pairwise* interactions of atoms in polypeptides (Burgess & Scheraga, 1975; Meirovitch & Scheraga, 1981; Levitt, 1983). More recently, simulations based on simplified protein models (Chan & Dill, 1989a; Covell & Jernigan, 1990; Gregoret & Cohen, 1991; Skolnick & Kolinski, 1991; Hao *et al.*, 1992; Liwo *et al.*, 1993) have indicated that certain *global* features of the native structures of proteins, such as their limited size and compact packing, could also contribute to the formation of the native structure. For example, Monte Carlo simulations with a simple lattice chain (Chan & Dill, 1989a) suggested that the compactness of a structure could induce a significant amount of regular local structure, analogous to that found in native proteins. The compact folding patterns of native proteins are largely the result of hydrophobic interactions, which are important forces in the formation of the native structure (Kauzmann, 1959; Tanford, 1979). While the hydrophobically induced global features of protein native structures have been clearly demonstrated in simulations with simplified protein models, it is of great interest to investigate what concrete roles the global factors play in dictating the folding pathways of proteins at a detailed molecular level, because the biological functions of proteins critically depend on their precise atomic structures.

Conventional molecular dynamics (MD) and Monte Carlo simulations on protein folding and/or unfolding require an excessive amount of computing time. For example, a 550-ps molecular dynamics simulation on the unfolding of bovine pancreatic trypsin inhibitor (BPTI) in a solvent (Daggett & Levitt, 1992), the longest MD simulation carried out on a protein in solvent to date, was still unable to unfold the protein at room temperature. At higher temperatures (about 500 K), the native structure of BPTI could be partially unfolded

† This work was supported by Grant No. GM-14312 from the National Institutes of Health, Grant No. DMB 90-15815 from the National Science Foundation, Grant No. CA-42500 from the National Cancer Institute (to M.R.P.), and Grant No. N00014-91-J-1943 from the Office of Naval Research (to S.R.). Support was also received from the National Foundation for Cancer Research.

* Author to whom correspondence should be addressed.
‡ Cornell University.
§ State University of New York Health Science Center.
‖ University of Rochester.

® Abstract published in *Advance ACS Abstracts,* September 1, 1993.

0006-2960/93/0432-9614$04.00/0

in a shorter time (Dagett & Levitt, 1992). In order to treat the unfolding/folding problem of proteins without resorting to excessively high temperatures, it is helpful to exploit possible simplifications in the simulation procedures.

This article deals with the issues raised above, *viz.*, simplifying the treatment of the protein folding problem, correlating the global features of a protein structure with the unfolding and folding of the native structure, and studying a well-characterized protein structure. To begin with, we consider the unfolding of a protein by swelling the native structure with solvent. This leads to an expansion of the protein structure. Thus, we propose a dynamic equation to describe the expansion of the principal components of the radius of gyration of the protein structure. We solve this dynamic equation by a numerical method. The purpose of this procedure is to simplify the simulation of protein unfolding with a reasonably detailed description of the molecular structure of the protein. Because a basic characteristic of a folding intermediate is that it can be folded to the native structure, after unfolding, we carry out refolding of the deformed protein structure. Toward this goal, we use an energy minimization procedure with a simple dimensional-constraint function to mimic the hydrophobic forces that force the protein structure to be compact. In the present problem, we find that our refolding procedure is sufficient to fold the expanded conformations of BPTI with correct local structures to the native structure. In particular, the procedure can overcome some of the energy barriers in compactly packed protein structures.

The approach in which the native structure is first unfolded and then the deformed conformations are refolded, in attacking the protein folding problem, has some computational advantages. First, as Daggett and Levitt have pointed out (1992), an unfolding simulation can proceed from a well-defined starting structure, *i.e.*, the crystal structure of the protein. Second, if protein folding and unfolding reactions obey microscopic reversibility, then the conformations unfolded from the native structure would correspond to those that are on the path in (the reverse process) folding to the native structure; thus, by studying such unfolded conformations, one may avoid many random and presumably nonproductive conformations while still being able to extract the essential information about protein folding.

Our study is focused on the small globular protein, BPTI. Extensive experimental (Creighton, 1977; Creighton & Goldenberg, 1984; States *et al.*, 1984; Weissman & Kim, 1991; Eigenbrot *et al.*, 1990) and theoretical (Levitt & Warshel, 1975; Burgess & Scheraga, 1975; Meirovitch & Scheraga, 1981; Levitt, 1983; Vásquez & Scheraga, 1988a,b) studies on the structure and folding of this protein have been carried out. As a result, there is a wealth of information about the stability and kinetics of folding of this protein. Of particular relevance to this study is the experimental characterization of the intermediates of BPTI with one or more broken disulfide bonds. An examination of the crystal structure of BPTI suggests that the native structure with three disulfide bonds has little freedom to unfold. Indeed, there is apparently no experimental information about the unfolding and refolding of disulfide bond-intact BPTI. When the disulfide bonds are broken, the structure can form a number of intermediates which offer a good opportunity for examining the folding pathway of this protein (Creighton, 1990; Weissman & Kim, 1991; Kosen *et al.*, 1992). Some intermediates themselves can form stable and native-like structures (States *et al.*, 1984; Creighton, 1990; Eigenbrot *et al.*, 1990; Weissman & Kim, 1991). We therefore use this molecule as the subject of our study, in the hope that insights from this protein will be relevant to other proteins as well.

In the Methods section, we describe the dynamic method for unfolding the native structure of a protein and the energy minimization procedure with dimensional constraints for folding the resulting expanded intermediates back to the native structure. In the Results section, we first report the results of simulated unfolding, *e.g.*, the unfolding paths and the resulting conformations under various simulation conditions, and then describe how the expanded conformations obtained from the dynamic procedure are refolded to the native structure, the folding trajectories, and the closeness of the refolded structure to the native one. In the Discussion section, we describe the insights gained from these simulations on the structure and folding of native BPTI. We compare the present work with previous studies and draw correlations between the theoretical results and experimental observations. We also discuss the general applicability of the present approach and its limitations.

METHODS

A. Unfolding of the Native Structure of a Protein. A primary indicator for the unfolding of the native structure of a protein is the expansion of its volume, presumably due to the loss of close-packing interactions in the structure. For example, early hydrodynamic data (Scheraga & Mandelkern, 1953) indicated that protein denaturation is accompanied by an increase in the volume of the molecule. It has been observed experimentally (Hawley & Mitchell, 1975) that high pressures dramatically slow the reaction rates of the reversible folding and unfolding of proteins. On the basis of these observations, Creighton (1990) pointed out that the conformations of proteins must pass through an expanded-volume state in reaching, or unfolding from, the native structure. It is therefore reasonable to assume that unfolding of the native structure of a protein is associated with an expansion in its volume or dimensions. Our approach is based on this hypothesis.

Let us consider an isolated protein molecule suspended in a solvent. When the solvent or environmental conditions are changed appropriately, the native structure becomes fluid and unstable, and it gradually unfolds. Imagine that the unfolding process is observed at low resolution, such that one cannot distinguish individual atoms but only the gross shape of the protein. Then, the unfolding process can be described as an evolution of the shape of a continuous protein body with time. In order to reduce the number of degrees of freedom required for characterizing the unfolding process, we assume that the time interval of observation is sufficiently long, so that the deformation of the solid protein body along each unfolding coordinate (defined below) becomes uniform. Then, the variables needed for characterizing the process are only the deformation rates along the unfolding coordinates. To obtain a rough estimate of the appropriate time scale, we note that the transverse and longitudinal speeds of elastic waves of native protein structures are about 2.2×10^5 and 3.3×10^5 cm s^{-1}, respectively (Gō, 1980). If the linear dimensions of a protein structure are 50 Å, then in a time of several nanoseconds or longer the deformation rate of a protein along each deformation coordinate will become uniform, due to the translation of these vibration waves in each of these directions. The unfolding processes of many protein structures are completed in times longer than nanoseconds; therefore, a dynamic procedure with the approximation of a uniform rate of deformation should be valid for describing the unfolding of a protein.

We now formulate the equation of motion for protein unfolding under the approximations stated above. To simplify

our treatment further, we neglect the shearing components of deformation. It can then be shown (Appendix A) that a natural set of coordinates for describing protein unfolding consists of the three principal radii of the protein structure, because in this set, the apparent kinetic energy of the protein has the following simple form:

$$T = \frac{M}{2}(\dot{r}_a^2 + \dot{r}_b^2 + \dot{r}_c^2) \quad (1)$$

where $r_a < r_b < r_c$ are the three principal radii, the dots over each variable represent derivatives with respect to time, and the number M is approximately equal to the mass of the protein. For convenience, we introduce a vector X:

$$\mathbf{X} \equiv \{r_a, r_b, r_c\}^T \quad (2)$$

with the superscript T standing for the transpose of the vector. Then, when there is no external force acting on the protein, the dynamic equation describing the variation of X (see Appendix A) can be written as

$$[M]\frac{d^2\mathbf{X}}{dt^2} = -\frac{dU(\mathbf{X})}{d\mathbf{X}} \quad (3)$$

where $[M]$ is a diagonal mass matrix with all its diagonal elements equal to the value of M in eq 1 and $U(\mathbf{X})$ is the potential energy of the protein molecule due to interatomic interactions. Limitations of the conditions under which eq 3 is valid are discussed in Appendix A.

The potential energy $U(\mathbf{X})$ in eq 3 is calculated here with the ECEPP/3 algorithm (Némethy *et al.*, 1992), which is the latest version of the Empirical Conformational Energy Program for Peptides (Momany *et al.*, 1975; Némethy *et al.*, 1983). We do not consider explicitly the contributions of the solvent–protein interactions, which can either stabilize or destabilize the native structure. Such interactions here are roughly attributed to the momentum in protein unfolding or refolding processes. It is also possible, in the same theoretical framework, to set up a more general expression for the unfolding process by a Langevin-type equation of motion (McCammon & Harvey, 1987), with an additional damping term and a random force term, to mimic the hydrodynamic and thermal effects on the motion of a protein. However, for simplicity, we confine ourselves here to the simple Newton-type equation of motion, which produces a unique trajectory of motion under a given set of parameters.

Equation 3 must be solved numerically because the potential energy has a highly nonlinear form and the variables X are themselves functions of the coordinates of the atoms of the protein. We use the backward-Euler scheme to solve the differential equation (Dahlquist & Bjorck, 1974; Peskin & Schlick, 1989; Schlick & Olson, 1992). In this scheme, a differential equation of the type $dy/dt = f(y)$ is discretized as $(y^{n+1} - y^n)/\Delta t = f(y^{n+1})$, where Δt is the time increment and y^n is the difference equation approximation to the functional value of y at the nth time interval Δt. This procedure is different from the usual "forward" molecular dynamics algorithms (McCammon & Harvey, 1987), in which the same differential equation is discretized as $(y^{n+1} - y^n)/\Delta t = f(y^n)$. Some of the reasons for using the backward procedure are self-explanatory in the following algorithm; others will be discussed later. Now, by the backward-Euler formalism, eq 3 can be expressed as

$$[M]\left(\frac{\mathbf{V}^{n+1} - \mathbf{V}^n}{\Delta t}\right) = -\frac{dU(\mathbf{X}^{n+1})}{d\mathbf{X}} \quad (4)$$

and

$$\mathbf{V}^{n+1} = (\mathbf{X}^{n+1} - \mathbf{X}^n)/\Delta t \quad (5)$$

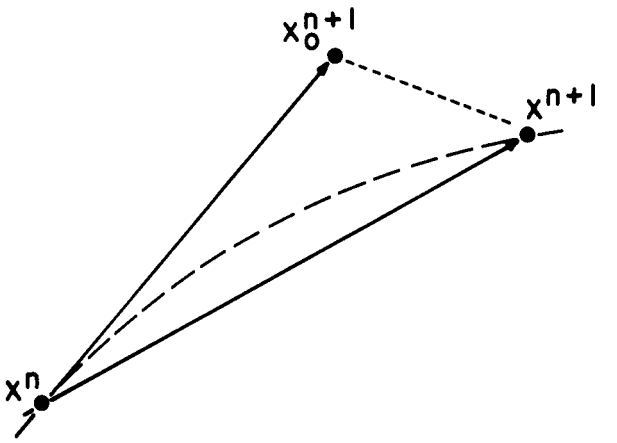

FIGURE 1: Illustration of the "backward" procedure for calculating the dynamic trajectory of protein unfolding. $\mathbf{X}^n$ is the starting conformation, $\mathbf{X}_0^{n+1}$ is the projected conformation according to eq 7, and $\mathbf{X}^{n+1}$ is the actual new conformation determined by minimization of $\Phi(\mathbf{X})$ of eq 9. The unfolding trajectory is indicated by the curve with large dashes, and the velocity at $\mathbf{X}^n$ is tangent to the unfolding trajectory.

The vector, V, is the velocity of the variable, X. The momentum, P, of unfolding is related to V by $\mathbf{P} = [M]\mathbf{V}$. In the simulation of an unfolding trajectory, when the unfolding time has reached $n\Delta t$ the function values of $\mathbf{X}^n$ and $\mathbf{V}^n$ are known, and the problem is only to solve for $\mathbf{X}^{n+1}$ and $\mathbf{V}^{n+1}$ for the next time increment. This is done as follows.

By inserting eq 5 into eq 4, the variable $\mathbf{V}^{n+1}$ can be eliminated from eq 4. After rearrangement, eq 4 can be written as

$$[M](\mathbf{X}^{n+1} - \mathbf{X}_0^{n+1}) + (\Delta t)^2\frac{dU(\mathbf{X}^{n+1})}{d\mathbf{X}} = 0 \quad (6)$$

where $\mathbf{X}_0^{n+1}$ is defined as

$$\mathbf{X}_0^{n+1} = \mathbf{X}^n + \Delta t\mathbf{V}^n \quad (7)$$

$\mathbf{X}_0^{n+1}$ can formally be viewed as an expectation value for $\mathbf{X}^{n+1}$ predicted from $\mathbf{X}^n$ and $\mathbf{V}^n$. Now, the only unknown in eq 6 is $\mathbf{X}^{n+1}$. Since eq 6 is a highly nonlinear function, $\mathbf{X}^{n+1}$ cannot be calculated directly from the equation. It is therefore solved by a minimization method as proposed by Schlick *et al.* (Peskin & Schlick, 1989; Schlick & Olson, 1992). Explicitly, eq 6, with $\mathbf{X}^{n+1}$ being denoted as the variable X, can be expressed as

$$d\Phi(\mathbf{X})/d\mathbf{X} = 0 \quad (8)$$

with the function $\Phi(\mathbf{X})$ defined as

$$\Phi(\mathbf{X}) = \frac{1}{2}(\mathbf{X} - \mathbf{X}_0^{n+1})^T[M](\mathbf{X} - \mathbf{X}_0^{n+1}) + (\Delta t)^2U(\mathbf{X}) \quad (9)$$

The value of X that satisfies eq 8 is the solution for $\mathbf{X}^{n+1}$ in eq 6. However, finding the X which satisfies eq 8 is equivalent to finding the value of X that minimizes (or more generally, extremizes) the function $\Phi(\mathbf{X})$ of eq 9. Thus, one can solve for $\mathbf{X}^{n+1}$ by minimizing $\Phi(\mathbf{X})$ of eq 9 with a nonlinear minimization algorithm, starting from the known values of $\mathbf{X}^n$ and $\mathbf{V}^n$. The geometric relationships among $\mathbf{X}^n$, $\mathbf{X}_0^{n+1}$, and $\mathbf{X}^{n+1}$ are illustrated schematically in Figure 1, where the curve passing through the points $\mathbf{X}^n$ and $\mathbf{X}^{n+1}$ represents the actual unfolding trajectory.

The algorithm that solves the dynamics of eq 3 consists of the following steps: (1) Starting from the native structure, the initial value is denoted as $\mathbf{X}^0$. (2) The first projected position is taken as

$$\mathbf{X}_0^1 = \mathbf{X}^0 + \mathbf{V}^0\Delta t \quad (10)$$

where $\mathbf{V}^0$ is the initial velocity and Δt is the basic time increment of the simulation in one iteration step. (3) The function $\Phi(\mathbf{X})$ is minimized at fixed Δt:

$$\min\left[\frac{1}{2}(\mathbf{X}-\mathbf{X}_0^1)^T[M](\mathbf{X}-\mathbf{X}_0^1)+(\Delta t)^2U(\mathbf{X})\right] \quad (11)$$

The starting position of minimization is $\mathbf{X}^0$ (see Figure 1). (4) The value of $\mathbf{X}$ that minimizes Φ is the new position on the unfolding path, denoted as $\mathbf{X}^1$. The new velocity is then calculated as

$$\mathbf{V}^1=(\mathbf{X}^1-\mathbf{X}^0)/\Delta t \quad (12)$$

(5) Set $\mathbf{X}^1$ to $\mathbf{X}^0$ and $\mathbf{V}^1$ to $\mathbf{V}^0$. Make the next projected position

$$\mathbf{X}_0^1=\mathbf{X}^0+\mathbf{V}^0\frac{\Delta L}{\|\mathbf{V}^0\|} \quad (13)$$

where ΔL is a preset basic distance increment in one iteration step. (6) Go to step 3.

Several remarks are necessary to clarify the meaning of the above algorithm. First, eq 10 starts the unfolding process by providing an initial unfolding velocity, $\mathbf{V}^0$. The first term on the right-hand side of eq 9 or eq 11 is related, but is not equal, to the kinetic energy (which is defined by eq 1) through its dependence on molecular mass and velocity (see eq 13); this term will hereafter be called the *kinetic factor*. The native structure of a protein does not unfold automatically. It is the kinetic factor that initiates the unfolding and drives the conformation from the current position to the next position. At each iteration step, the old kinetic factor makes a prediction, and the actual direction and distance of the movement are determined by both the potential energy and the kinetic energy contributions through minimization of the total function $\Phi(\mathbf{X})$ of eq 9 or 11. The outcome of this minimization determines the velocity (see eq 12) for the next step of the motion. In this way, we find an unfolding path for the native structure of the protein.

Second, the projection step of eq 13 in the algorithm is a modification of eq 7. If the velocity resulting from the previous step of motion is denoted as $\mathbf{V}_o^0$, and the velocity used in making the projection for the next step of motion is $\mathbf{V}^0$, then by comparing eq 7 and eq 13 and making use of eq 12, one finds that

$$\mathbf{V}^0=\mathbf{V}_o^0\left(\frac{\Delta L}{\|\mathbf{X}^1-\mathbf{X}^0\|}\right) \quad (14)$$

In other words, in each step, the velocity is rescaled by the ratio of the fixed distance increment to the actual distance moved in the previous step. We use this procedure to maintain the unfolding simulation at a constant rate of progress; the procedure will not take a very large step in a downhill motion, nor will it slow down too dramatically in an uphill situation.

The use of the backward-Euler procedure is essential in the present approach because of the implicit relationships between the variables $\mathbf{X}$ and the coordinates of all of the atoms in the protein. The resulting conformations must finally be expressed in terms of all of the atomic coordinates, which are correlated to the variables $\mathbf{X}$ through the approximation of the present approach, *i.e.*, the uniform rate of deformation of the protein structure in the directions of three principal axes. However, direct calculation of the atomic coordinates according to a uniform rate of deformation can lead to high-energy conformations when the moving step size is large. Using the minimization procedure, the lowest energy conformation at each given $\mathbf{X}$ on the unfolding path can be determined naturally.

In the implementation of this algorithm in our present problem, we write the function to be minimized, *i.e.*, $\Phi(\mathbf{X})$ of eq 11, in the following form:

$$\min[\Phi(\mathbf{X})]=\min\{U(\mathbf{X})+W[(r_a-r_{a0}^1)^2+(r_b-r_{b0}^1)^2+(r_c-r_{c0}^1)^2]\} \quad (15)$$

where $W\equiv M/[2(\Delta t)^2]$ is a weight for the kinetic factor with respect to the potential energy and $(r_{a0}^1,r_{b0}^1,r_{c0}^1)$ are the components of $\mathbf{X}_0^1$ determined by eq 13, which are the projected principal radii for the new conformation. This is possible because the coordinates of $\mathbf{X}$ which minimize the function $\Phi(\mathbf{X})$ at each step of iteration depend only on the ratio of M to $(\Delta t)^2$ under the given potential function and moving step size. Since each new conformation is always determined by the minimization of $\Phi(\mathbf{X})$ of eq 15, *i.e.*, the potential energy of the protein plus a constraint term (the kinetic factor), the new conformation always stays on a low-energy path in the unfolding, no matter how large the basic distance increment has been set. This ensures the stability of the algorithm. A further condition for the stability of the unfolding trajectory is that the weight, W, should not be so large that the kinetic factor term dominates the conformational change. In our simulations, we choose the smallest possible value for W (see below for details), and the above condition is satisfied.

A final remark concerns the temperature of the above dynamic procedure. It should be stressed that the expansion of the protein can be caused by either solvent (pH and detergents) or temperature changes or by a combination of both. The dynamic equation describes only the net expansion of the protein structure, while the real unfolding could also involve random fluctuations. Therefore, the expansion of the protein is expected to be diffusion-like, and the rate of expansion, $\dot{D}$ (see Appendix A), should be related to a frictional coefficient, f, of the protein structure through an Einstein-type relation: $\dot{D}=k_BT/f$, where T is the temperature. The frictional coefficient of a protein structure may be obtained from short-time, rigorous molecular dynamics simulations on protein structures. If the frictional coefficient of a protein structure and the temperature are known, one can determine the appropriate step lengths in the dynamic procedure. We have not attempted to determine the frictional coefficient of protein structures. Since the purpose of this work is only to unfold the protein, we simply use a trial-and-error process to determine the simulation steps. In this aspect, the present dynamic procedure differs from the conventional molecular dynamics method.

B. Minimization of Expanded Protein Structures with Dimensional Constraints To Determine the Refolding Pathway. In principle, the unfolding algorithm described above can also be used to fold an expanded conformation of a protein into a compact and native-like structure, provided that the initial velocity is taken in the direction of decreasing principal radii. The problem with this approach is that, unlike the unfolding process, folding to the native structure has a precise end point. The introduction of a kinetic energy term can change the path of motion, and the resulting structures will usually exhibit large deviations from the native structure. Energy minimization is therefore the method of choice in order to fold a structure more closely to the native structure. Energy minimization by itself, however, has another drawback: even if the overall motion of the conformation is in the direction of decreasing energy, the structure can still be trapped by any minor and shallow local minimum between the starting conformation and the lowest energy conformation. It therefore becomes a question of whether the difference between the

minimized structure and the native one is due to minimization or due to errors of the folding path. To overcome this problem, we have used the following procedure.

The dimensional parameters, *i.e.*, the three principal radii, of the native structure of the protein are used as constraint functions in minimizing the conformation of the protein to the native structure. The function to be minimized has the following form:

$$F = U + K[(r_a - \rho_a)^2 + (r_b - \rho_b)^2 + (r_c - \rho_c)^2] \quad (16)$$

where U is again the ECEPP/3 potential energy of the protein, K is the weight for the constraint term, r_a, r_b, and r_c are the principal radii of the structure for which the function F is being minimized, and ρ_a, ρ_b, and ρ_c are the principal radii of the crystal or other experimental structure of the protein. It should be noted that the above function is quite similar in form to eq 15 in the dynamic procedure. However, these two functions have been constructed on the basis of different considerations and are used differently. The constraint function in eq 16 is intended to modify the potential energy surface of the protein. Since the minimum of the simple quadratic function is easily reached by minimization, the constraint function could help to drive the conformation out of local energy minima of nonnative conformations, so that the minimized structure can be closer to the native structure. In the dynamic program, the new direction in each iteration is determined by the motion itself (*i.e.*, the velocity is determined by eq 12). In the constraint–function approach, on the other hand, the moving direction is always pointed toward the native structure during minimization, and the final minimized structure can approach the native structure more accurately.

Some physical justification can also be suggested for the use of dimensional constraints in minimizing the energy and structure of a protein. The effects of compact volume on the native structures of proteins have been indicated by a number of studies (Chan & Dill, 1989a,b; Covell & Jernigan, 1990; Gregoret & Cohen, 1991; Hao *et al.* 1992). The contributions of dimensional constraint for locating the native structure of a protein come from two sources. First, the origin of the compactness of the native structure of a protein is the hydrophobic interaction. By requiring that the minimized structure also must have the correct dimensions of the native structure, one implicitly forces the structure to adopt the most favorable conformation for hydrophobic interactions. Second, compactness induces packing effects. This is best illustrated by Flory's remark (1972) that the compact conformation which accommodates a long polypeptide chain in a very limited space, and at the same time juxtaposes their bulky side chains in a manner providing for favorable tertiary interactions, would be a rarity. The native structure of a protein must be one of those rare conformations. Then, by requiring that the energetically minimized structure also satisfies the dimensional constraints, one can significantly increase the chance of finding the native structure which, according to the conventional view, is likely to be the conformation of lowest energy.

By analogy with the stepwise dynamic program, the minimization with a dimensional constraint is also carried out in a number of steps. Given the dimensional parameters of the starting conformation, say r_a^0, r_b^0, and r_c^0, and if the principal radii of the crystal structure of the protein, ρ_a, ρ_b, and ρ_c, are known, the difference vector between the two conformations, *i.e.*, $(r_a^0 - \rho_a, r_b^0 - \rho_b, r_c^0 - \rho_c)$, is divided into a number of segments. In each step of minimization, the conformation is forced only to an intermediate compactness between that of the starting structure and that of the most compact native structure. The dimensions of the intermediate constraints are gradually decreased until they reach that of the native structure. This gradual decrease of dimensions is designed to avoid abrupt conformational changes in the extended structure driven by the constraint function. An abrupt change could produce some undesirable conformations. Also, at the beginning of minimization, when the conformation is very extended, the weight, K, for the dimensional-constraint function is given a small value. When the conformation becomes more compact, the weight is increased. This procedure reflects the intended objectives of the dimensional constraint: when the conformation is expanded, the dimensional-constraint function is used for driving the conformation to more compact structures, and this requires only a small force. When the density of the structure is close to that of the native structure, the dimensional-constraint function is used for enforcing proper packing of the polypeptide chain in the confined space, which requires chain rearrangement and hence a larger force. The actual value of K and the number of intermediate compactness constraints are chosen by trial-and-error. In the present calculations, at the beginning of minimization when the conformation is expanded, the starting value of the weight K is set to 1 kcal/mol/Å^2, which was found to be sufficient to drive the structure to a more compact form. With each increase in the compactness of structure, the weight is increased by a factor of 1.5. The procedure for calculating the principal radii and their derivatives is described in Appendix B.

The minimization algorithm used in this study is the secant unconstrained minimization solver (SUMSL) (Gay, 1983). The algorithm has been tested in this laboratory and found to be very efficient for minimizing the complicated potential functions of polypeptides (ECEPP/3) with torsional angles as variables (Scheraga, 1992). The minimization is carried out with single precision on a Stardent Series 3000 computer. The covergence criteria for minimization are set to 0.005 kcal/mol for the function value and 0.1 kcal/mol/deg for the gradient of the function. Minimization of the energy of a single conformation of BPTI (886 atoms, 322 variable torsional angles) with the ECEPP/3 program takes about 20 min on the above machine. A typical unfolding simulation usually requires 40–60 minimizations to reach the maximally expanded structure of BPTI, but in a folding simulation, only about 10–15 minimization steps are needed. The other computations in the procedure, *e.g.*, calculating the principal radii, evaluating the velocities, calculating the constraint functions, etc., involve comparatively negligible time. The crystal structure of BPTI used here is 5PTI from the Brookhaven Protein Data Bank (Wlodawer *et al.*, 1984).

RESULTS

A. Unfolding of the Native Structure of BPTI. BPTI is a small globular protein with 58 residues. In the native form, there are three disulfide bonds in the protein formed by residues 5 and 55, 14 and 38, and 30 and 51, denoted as [5,55], [14,38], and [30,51], respectively. In the crystal structure of BPTI, there is an antiparallel β-sheet formed by residues 16–24 and 28–35 and a short α-helix formed by residues 47–55. The starting structure of our unfolding simulation is the standard geometry version of the ECEPP/2-minimized crystal structure of BPTI (Vásquez & Scheraga, 1988a). The ECEPP/2-minimized structure has an rms deviation for all C^α atoms of 1.1 Å from the original crystal structure. This structure was first minimized with ECEPP/3. The resulting structure has an rms deviation of 1.0 Å from the crystal structure. Since the minimized crystal structure represents

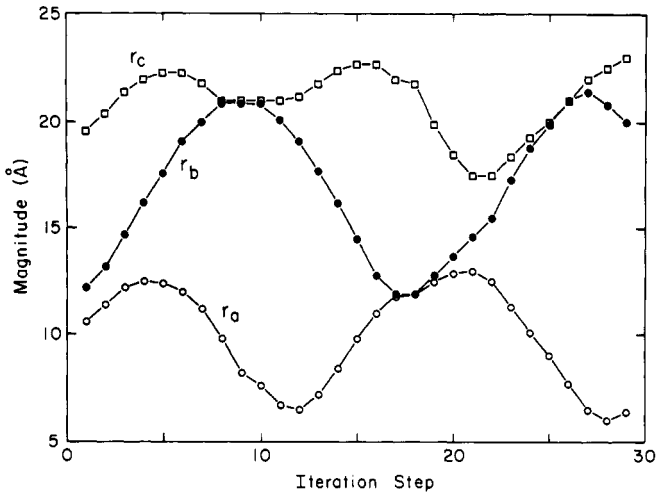

FIGURE 2: Principal radii, r_a, r_b, and r_c, of the ellipsoidal structure of BPTI with its disulfide bonds intact as a function of iteration step obtained from the dynamic simulation. The initial velocity increases uniformly in the directions of all three principal axes, the basic distance increment in one iteration step is 2 Å, and the weight of the kinetic factor is 12 kcal/mol/Å^2.

the minimum-energy structure under the potential used, the difference of all other minimized structures with respect to the native one, measured by rms deviation, is calculated with respect to the standard geometry version of the ECEPP/3-minimized crystal structure of BPTI. The three principal radii ρ_a, ρ_b, and ρ_c of this reference structure are 10.6, 12.3, and 19.6 Å, respectively.

We first examined the fluctuations and unfolding of the native structure of BPTI with all three disulfide bonds intact. It was found that the native structure cannot be unfolded completely by the dynamic procedure if reasonable initial velocities are imposed. In a typical example, from the simulation with an initial unfolding velocity consisting of a uniform expansion in all three principal directions, a step increment of 2 Å, and a weight W of 12 kcal/mol/Å^2, the trajectories of motion of the three principal radii, r_a, r_b, and r_c, have the form shown in Figure 2. Instead of unfolding being driven by the kinetic energy, the native structure actually undergoes an oscillation, with large fluctuations in the principal radii. As can be seen from Figure 2, the three principal radii of the structure increase and decrease alternately and periodically, even though they are not exactly in phase. The motion of the structure is found to be confined within a certain region around the native structure. This result is not totally surprising. There appears to be no obvious way for the native structure of BPTI to unfold completely when all three disulfide bonds are intact. The simulation results indicate that the native structure of BPTI is a very stable one, which is difficult to unfold by expansion of the structure. The native structure can, however, eventually be unfolded with an arbitrarily very large initial kinetic energy. Such a situation, however, is not relevant to a realistic unfolding process and is therefore not considered here.

It is clear from the above results that unfolding of BPTI can be realized only when one or more of its disulfide bonds is broken. Among the three native intermediates of BPTI in which only one disulfide bond is broken, the one with the [14,38] S–S bond broken also exhibits a similar oscillation around the native state (results not shown here) when simulated with the dynamic program, and this intermediate is therefore omitted from further consideration. The intermediate with the [5,55] disulfide bond broken, when unfolded by the dynamic program, leads to a conformation with two separated structural units: in one, the β-sheet (from residues 16–35) and the α-helix (from residues 39–53) of the native BPTI structure stack together, and in the other there is a single strand consisting of residues 1–14. The unfolding of the intermediate with the [5,55] S–S bond broken primarily involves a flipping around of the single strand. It is interesting to note that the closely packed α-helix and β-sheet of this intermediate obtained by the present procedure are similar to the experimental structure of the same two fragments of BPTI, connected together by the [30,51] disulfide bond, proposed by Oas and Kim (1988). Because the folding and unfolding of the single strand do not follow a unique path, for simplicity we also omit this intermediate from further consideration in the present study. The most interesting unfolding path found among the intermediates of BPTI is the one with the [30,51] disulfide bond broken. The unfolding as well as refolding of this intermediate shows a unique and reproducible path over a large conformational region. The following investigation is therefore focused on this intermediate of BPTI.

Table I: Properties of Energy-Minimized Structures of BPTI

structure[a]	energy (kcal/mol)	r_a (Å)	r_b (Å)	r_c (Å)	rms (Å)
1	−480.5	10.6	12.3	19.6	0.0
2	−478.8	10.7	12.2	19.6	0.2
3a	−350.3	6.1	12.4	48.9	20.5
4a	−354.4	9.1	10.2	49.3	20.8
5a	−360.4	9.7	10.0	48.9	20.1
6a	−349.1	11.1	14.6	48.8	19.4
7a	−400.8	7.3	15.1	39.9	16.2
3b	−479.4	10.6	13.0	19.9	2.6
4b	−464.6	10.8	12.7	19.7	3.0
5b	−468.1	10.7	13.0	19.9	3.1
6b	−478.6	10.7	12.5	19.7	2.0
7b	−477.2	10.6	12.6	19.6	1.9
3c	−464.5	10.9	12.5	19.6	1.7
4c	−470.3	10.7	12.4	19.6	2.2
5c	−481.1	10.7	12.6	19.6	2.0
6c	−486.5	10.9	12.4	19.5	1.1
7c	−470.9	10.8	12.5	19.4	0.9

[a] Structure 1 is the standard geometry version of the ECEPP/3-minimized crystal structure of BPTI. Structure 2 is the minimized structure from structure 1 with the [30,51] disulfide bond broken. Structures 3a–7a are the unfolded conformations obtained by the dynamic procedure, and the simulation conditions for each of the structures are described in the main text. Structures 3b–7b are the structures refolded by energy minimization with the dimensional constraints of the crystal structure, starting from the corresponding structures 3a–7a. Structures 3c–7c are the energy-minimized structures starting from 3b–7b with the constraints of both the [30,51] disulfide bond and the dimensions of the crystal structure. The rms deviations of all the structures are calculated with respect to structure 1 for all C^α atoms.

In breaking the disulfide bonds in the above simulations, the S atoms involved in the disulfide bond(s) are first replaced by SH groups. The energies of the resulting structures are then minimized, first under the constraints of the dimensions of the native structue and then without any constraint. The perturbations of the energy-minimized native structure caused by breaking a single disulfide bond are very small: the dimensions of the structure in terms of the principal radii are almost unchanged (see Table I, structures 1 and 2), the rms deviation between the two structures is only 0.19 Å, and even the ECEPP/3 energies of the structures with and without the [30,51] disulfide bond are not very different. The ECEPP/3 energy of the minimized structure of the intact BPTI structure is −480.5 kcal/mol, and that with the [30,51] disulfide bond broken is −478.8 kcal/mol. The energies of these two conformations, however, are not exactly comparable because their chemical structures are not identical.

The unfolding path produced by the dynamic procedure is affected by three parameters: the initial velocity of unfolding, the coefficient or weight of the kinetic factor term, and the

moving step size in each iteration. The effects of each of these parameters are now discussed separately.

The initial velocity determines the initial direction of unfolding. Its effects can be appreciated from the hypothetical situation that, if there were no influence of potential energy, the unfolding motion would simply proceed in the direction of the initial velocity. Since we are interested in unfolding, we assume that the direction of the initial velocity is toward the expansion of the structure. Under realistic conditions, the initial velocity will possibly have a certain distribution of values. For simplicity, we consider the following four representative cases: (1) the components of the initial velocity are equal in the directions of all three principal axes, *i.e.*, an isotropic expansion of the structure; (2) the component in the direction of r_c, the longest principal radius, (3) the component in the direction of r_b, the second longest principal radius, and (4) the component in the direction of r_a, the shortest principal radius, are each 2 times larger than the other two components, respectively. The last three cases correspond to the situation where unfolding occurs predominantly in the direction of one principal axis.

The trajectories of the principal radii r_b and r_c produced by the dynamic procedure for cases 1–3, respectively, are plotted in Figure 3a, in terms of variation of r_b versus r_c. In these simulations, the coefficient of the kinetic factor is set to 10 kcal/mol/Å^2, the step size in each iteration is 2 Å, and the magnitude of the initial velocity is 1.5 Å per basic time unit in the simulation. Curves 1–3 correspond to cases 1–3, respectively. Figure 3b shows the variation in the conformational energy of the protein as a function of the longest principal radius, r_c, for the same three cases. The principal radius, r_c, is found to increase monotonically with the iteration steps in the simulation and is an indicator of the extent of unfolding of the structure. r_c is therefore taken as the variable in terms of which the other conformational properties are expressed in the above (and in the following) figures. In case 4, it was found that, in most of the unfolding trajectory, the potential energy was nearly 150 kcal/mol higher than those in the other three cases, and if the energy of case 4 were plotted in Figure 3b, the curve would be far off scale. On the basis of energy considerations, it appears that case 4 is an unlikely unfolding path, and this case is therefore not considered any further.

Unlike the oscillation pattern found in Figure 2, the unfolding process shown in Figure 3a indicates that one of the principal radii (here r_c) increases monotonically with iteration steps, and the structure is gradually opened up. The unfolding trajectories of cases 1 and 2 (curves 1 and 2) have basically similar features, and the energy profiles of the two paths shown in Figure 3b are also comparable. The most interesting unfolding path is that for case 3, in which the component of the initial velocity in the r_b direction is 2 times larger than the other two components. For case 3, a loop appears at the beginning of the unfolding trajectory in which r_b first increases and then falls back close to the initial value; in other words, the structure moves cyclically when perturbed with this initial velocity. However, by the end of the cycle, the structure finds the possible direction of unfolding, and thereafter it proceeds to unfold along a path that is basically similar to that in cases 1 and 2. The energy profile of the third path, however, is about 30 kcal/mol higher than that of the other paths, as shown in Figure 3b. To summarize for all four cases considered, it may be conjectured that the most likely unfolding path for the intermediate of BPTI lacking the [30,51] disulfide bond is initialized in the direction in which the longest principal radius increases most rapidly.

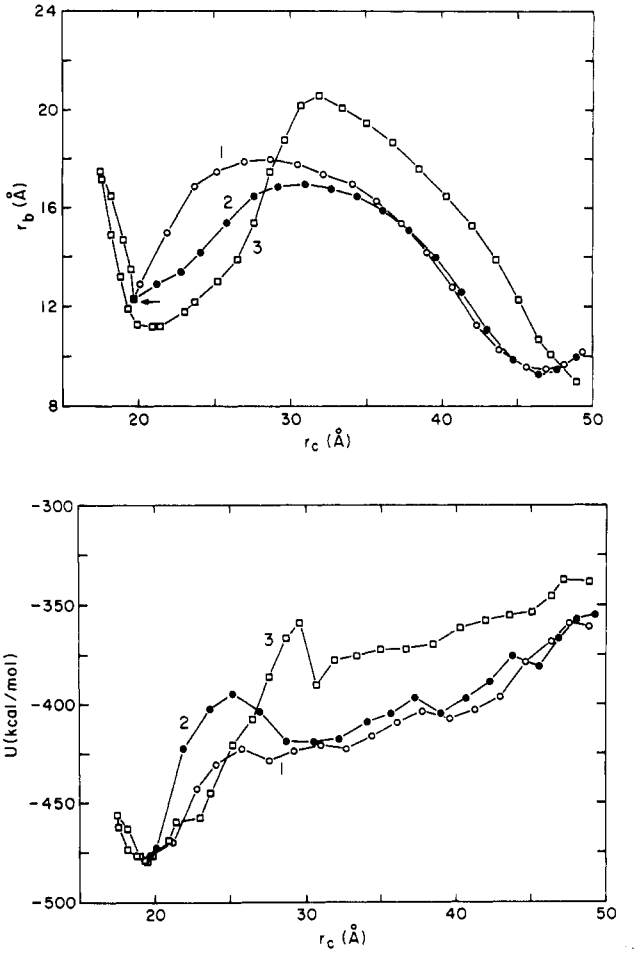

FIGURE 3: Comparison of the unfolding trajectories under different initial velocities for the intermediate of BPTI lacking the [30,51] disulfide bond. Curves 1, 2, and 3 correspond to simulations with the components of initial velocities that increase equally in all directions, are 2 times larger in the r_c direction, and are 2 times larger in the r_b direction, respectively. The starting points of all three trajectories are the single common point at the lower left end of the trajectories shown by the arrow. The extent of unfolding is measured by the increase of r_c, the longest principal radius. a (top) and b (bottom) are plots of the principal radius r_b and the conformational energy U, respectively, as a function of r_c. The coefficient of the kinetic factor is 10 kcal/mol/Å^2, the distance increment in one basic time unit is 2 Å, and the magnitude of the initial velocity is 1.5 Å per basic time unit.

The coefficient, W, of the kinetic factor is correlated with the moving step size. In order for unfolding to proceed, the parameter W must be increased when the moving step size is decreased and vice versa. The reason for this correlation is as follows: the unfolding trajectory is determined by the kinetic and potential energies of the molecule. Since the potential energy surface is the same for a given protein, in order for two trajectories to be comparable, their kinetic factors, or magnitudes of velocity, should be identical. With the same velocity, the longer the distance a structure moves in a step, the longer the time it will take in the move. Since the coefficient W is proportional to $(\Delta t)^{-2}$ (see eq 15), an increase of Δt corresponds to a decrease in the coefficient W. The value of W in each simulation is determined by trial-and-error: we start with the lowest possible value for W, corresponding to the smallest kinetic energy required for unfolding. If unfolding cannot proceed, the coefficient W is increased cumulatively by a factor of 1.5.

Figure 4a shows the unfolding trajectories as the principal radius r_b versus r_c, corresponding to moving step sizes of 1, 2, and 3 Å in each iteration, respectively. Figure 4b shows the energy profiles in unfolding as a function of r_c. The initial

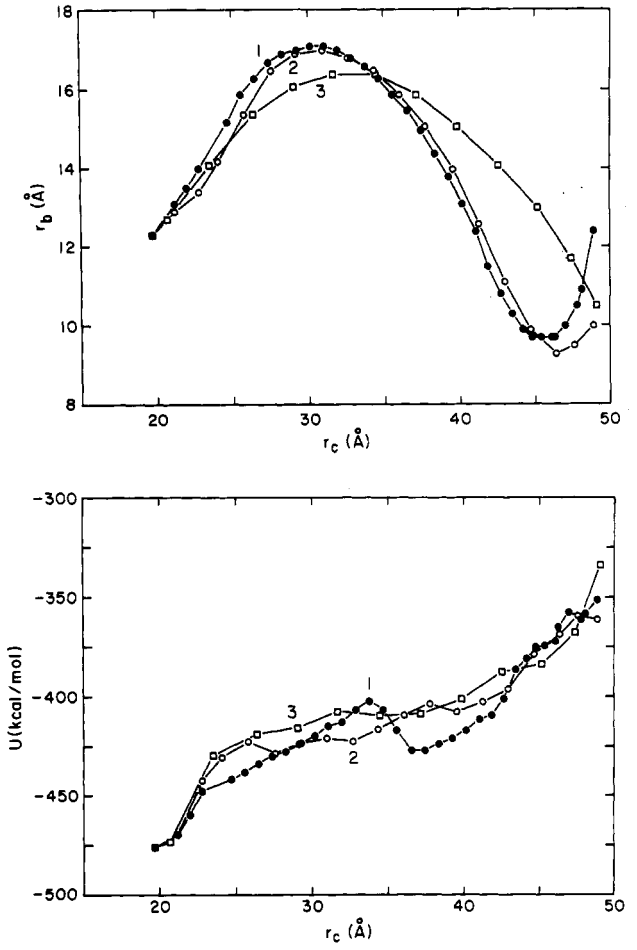

FIGURE 4: Comparison of the unfolding trajectories under different distance increments for the intermediate of BPTI lacking the [30,51] disulfide bond. Curves 1, 2, and 3 correspond to simulations with distance increments of 1, 2, and 3 Å in each iteration step, respectively. The extent of unfolding is measured as an increasing function of r_c. a (top) and b (bottom) are plots of the principal radius r_b and the conformational energy U, respectively, as a function of r_c. The coefficients of the kinetic factor are 15, 10, and 6.7 kcal/mol/Å^2 for the distance increments of 1, 2, and 3 Å, respectively. The component of the initial velocity in the direction of r_c is twice that in the r_a and r_b directions in all three simulations.

velocity of unfolding is the same as that in case 2 described above. The factor W is 15, 10, and 6.7 kcal/mol/Å^2 for the simulations with step sizes 1, 2, and 3 Å, respectively. The starting points of unfolding in Figure 4a,b are the common point of the three curves at the lower left end. From Figure 4a, it can be seen that the unfolding trajectories for step sizes 1 and 2 Å are basically identical, but the trajectory for step size 3 Å is flatter than the other two curves. A possible reason for the leveling of the third trajectory in Figure 4a is that the contribution of the kinetic factor for this simulation is larger than those for the other two, because even though its W factor is smaller, its step size is larger and the kinetic factor is proportional to the square of the step size length (see eqs 9 and 13). When the kinetic factor is larger, the unfolding trajectory is influenced to a greater extent by the kinetic energy, and as a result, the fastest increasing dimension changes more rapidly. In practice, with a larger kinetic energy contribution, the minimization coverges more rapidly and unfolding proceeds faster.

From the energy trajectories shown in Figures 3b and 4b, it can be seen that the energy profiles of unfolding obtained under various simulation conditions have some common features: the most energetically costly part is the beginning of unfolding, in which the energy increases very rapidly. Once the structure is initially opened up, the energy of the structure changes very slowly (nearly constant) over a significantly large conformational region, in which the longest principal radius increases from 24 to 44 Å. The conformational energy, however, increases rapidly again after the structure is expanded beyond $r_c = 45$ Å. In all of the above calculations, except case 4 described earlier, the unfolding simulations were terminated when the conformational energy of the protein increased beyond −320 kcal/mol or when the longest principal radius of the structure was larger than 50 Å. We found that this is approximately the limit beyond which the deformed structure cannot be refolded to the native structure by our folding procedure.

To illustrate how the native structure of BPTI unfolds, Figure 5 shows a series of intermediate conformations during unfolding simulated by the dynamic procedure. These conformations correspond to curve 1 shown in Figure 4a. The molecular conformations at a given value of r_c are very similar for the three trajectories shown in Figure 4a, even though the trajectories of their dimensional parameters are somewhat different. Figure 6a,b shows the contour distance maps for the C^α atoms of the ECEPP/3-minimized crystal structure and the expanded structure shown in Figure 5f, respectively. By comparing Figure 6b with 6a, it can be seen that, in the expanded conformation, only the contacts within the short α-helix and within the β-sheet of the native structure exist; other contacts in the native structure of BPTI, *e.g.*, the contacts between residues 5–8 and 22–26 and between residues 20–30 and 43–54, are lost in the expanded conformation. It should be noted that the contact patterns within the β-sheet loop region (residues 13–23 to 25–35) of the expanded conformation are also different from those of the native structure, reflecting the fact that the extended loop is untwisted (see also the stereo structure of Figure 5f).

Figure 7a,b shows the changes in the backbone torsional angles of all of the expanded conformations with respect to their values in the native BPTI. Two observations can be made from these data. First, while the average rms deviation of all of these torsional angles with respect to their native values is about 27°, the distribution of the torsional angle changes along the sequence of BPTI is not uniform, ranging from near zero to over 100°. In all but one (the one marked by stars) of the expanded conformations, numerous torsional angles (ranging in number from 10 to 20 in a structure) occur in different rotamer (*i.e.*, gauche and trans) states from the native structure. Thus, some residues have to jump from one rotamer state to another in unfolding or folding. Second, even though the global conformational changes in these expanded structures are remarkably similar, the changes of individual torsional angles vary greatly from one simulation to another. This indicates that unfolding, as well as folding, can be carried out over a range of paths in the torsional angle space. The unfolding simulations seem to cover a representative subset of the possible backbone torsional angle changes in BPTI unfolding.

The changes of the side-chain torsional angles in the simulations are found to be quite small, with an average deviation of 12° from the native values. This is caused by the following simplified aspects of the procedure. The algorithm simulates the expansion of the protein structure, and because changes in side-chain conformation make only small contributions to the overall dimensional changes, there is little momentum to force the side chains to cross energy barriers as the structure is being unfolded. Considering that the variations of the torsional angles in protein unfolding would be rather random, to fix the side chains in their stable rotamer states could represent one possible choice for these side chains.

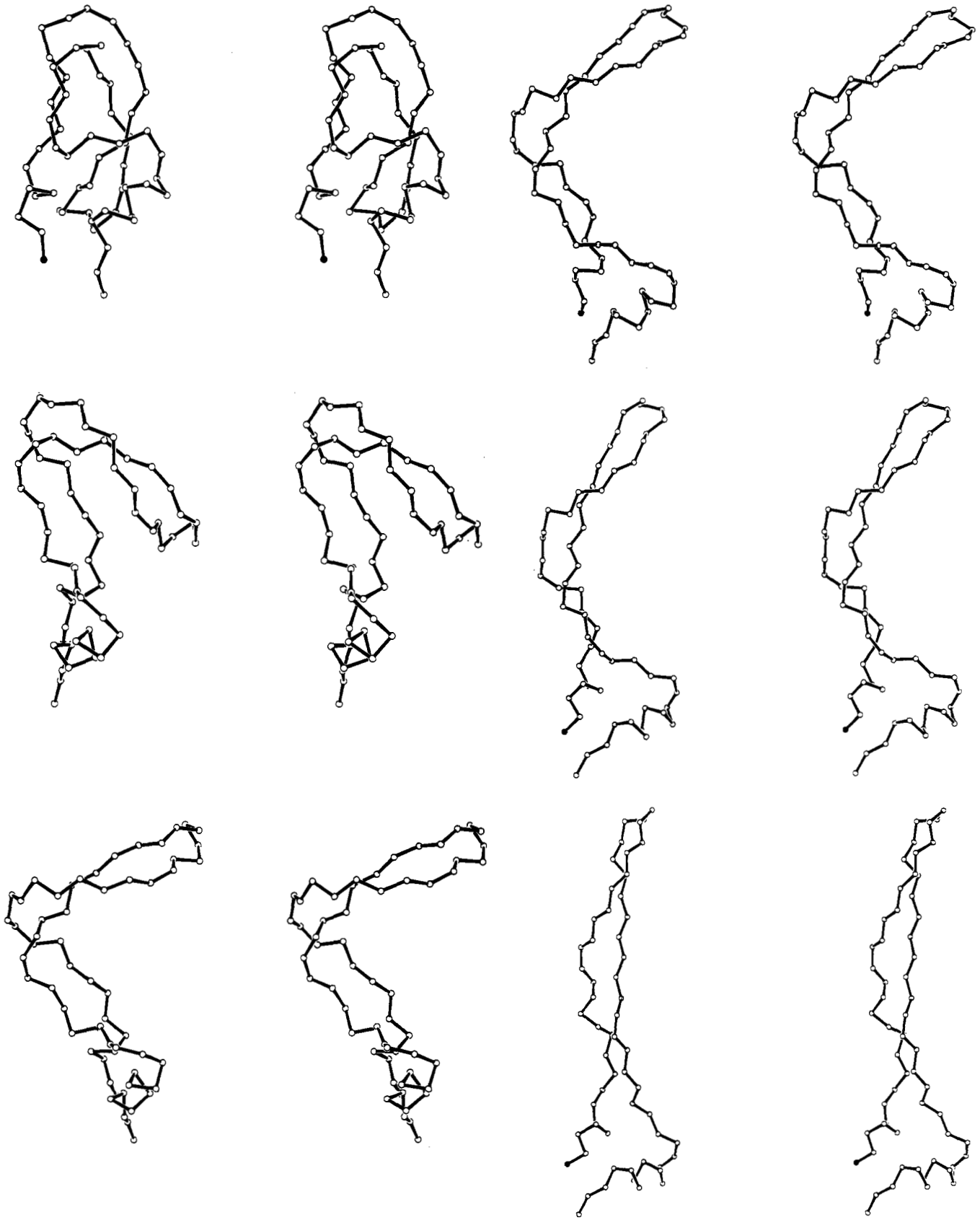

FIGURE 5: Stereo diagrams illustrating the unfolding of BPTI lacking the [30,51] disulfide bond from a dynamic simulation. The conformations of BPTI are represented by the C^{α} trajectory of the protein. The unfolding path corresponds to the trajectory of curve 1 in Figure 4a. a–f are the conformations obtained at iteration steps of 0, 5, 10, 16, 22, and 33 in the simulation, respectively. a–c are top to bottom, respectively, on the left, and d–f are top to bottom, respectively on the right.

Therefore, this simplification should not alter the global conformational changes in protein unfolding.

To summarize all of the above results, we may conclude that the most favorable unfolding path for the BPTI intermediate without the [30,51] disulfide bond is the opening up of the structure, in which the loop formed by residues 15–37 turns away from the rest of the structure. This unfolding may be described simplistically as a rotation of the loop around

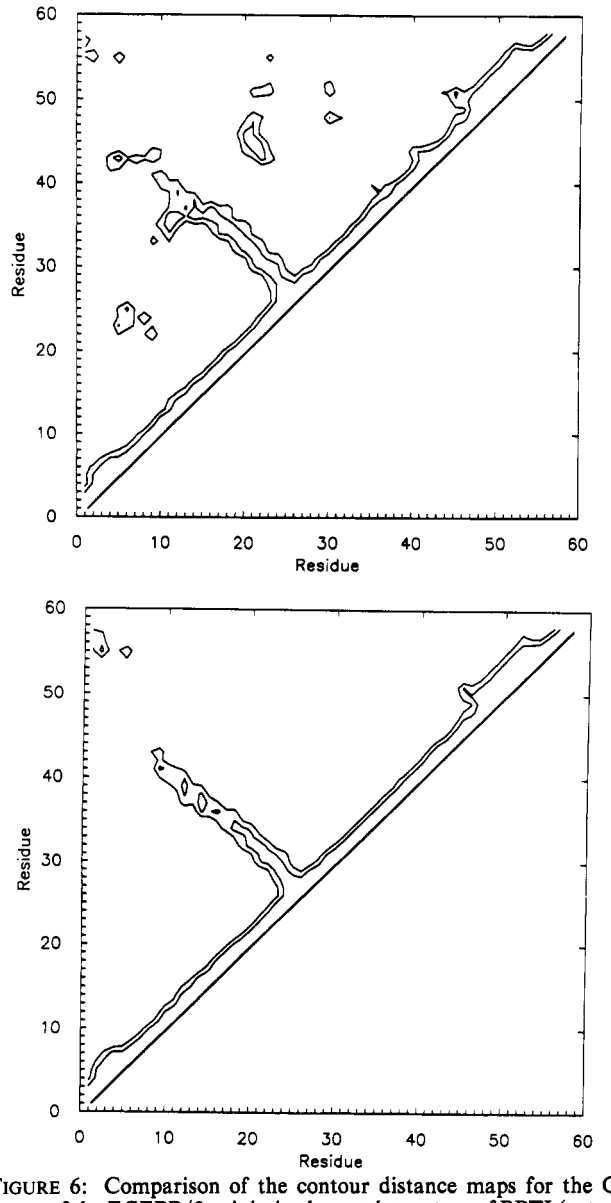

FIGURE 6: Comparison of the contour distance maps for the C^{α} atoms of the ECEPP/3-minimized crystal structure of BPTI (a, top) and the expanded conformation (b, bottom). The expanded structure corresponds to Figure 5f. The inner and outer contours of the two maps represent distances of 6 and 8 Å, respectively, between C^{α} atoms of different residues.

an axis which passes roughly through residues 14 and 38, accompanied by some untwisting of this loop. An examination of a wire model of BPTI (results not shown here) indicated that this mechanism appears to be the easiest way for BPTI to unfold, on the basis of considerations of stereochemical constraints in the native structure. An earlier, simpler procedure that we developed, in which the native structure of BPTI was expanded in the direction of its longest principal axis with a minimization procedure (unpublished work), also yielded qualitatively similar deformed structures. The unfolded conformations resulting from the dynamic procedure maintain the essential secondary structures of native BPTI, *e.g.*, the short α-helix, the β-sheet with a right-handed twist, and the β-turn. It is of interest to point out that Chou *et. al.* (1985) found that the twisted β-sheet present in residues 18–35 of native BPTI could be obtained by energy minimization of the untwisted form of the *isolated* fragment. On this basis, Chou *et al.* (1985) conjectured that the twisted conformation of residues 18–35 may form early in the folding process. Also, Nakazawa *et al.* (1992) have reported a similar twisted β-sheet structure for the 16–36 fragment of BPTI on the basis of Monte Carlo simulated annealing simulations; these authors further stated that their 2D NMR experiments also support the existence of the loop conformation formed by this fragment. Taken together, these facts suggest that the loop conformation that we obtained is a stable structure, which can be a favorable intermediate for BPTI in its folding to the native structure.

B. Refolding of the Expanded Conformations to the Native Structure. We next carried out refolding of the expanded conformations generated by the dynamic procedure. These refolding simulations show that the unfolded conformations obtained above are indeed favorable intermediates, which can be folded into the native structure by minimization of their energies. The unfolded conformations, obtained from the dynamic program at the end of each unfolding run, consisted of the following five: (1) one generated with a step size of 1 Å and an initial velocity that was isotropic in all three directions (the trajectory of curve 1 in Figure 4a,b; this conformation is denoted as 3a in Table I); (2) and (3) two conformations generated with a step size of 2 Å and with initial velocities that were isotropic, and largest in the r_c direction, respectively (these correspond to the trajectories of curves 1 and 2, respectively, in Figure 3a,b; conformations 2 and 3 are denoted as 4a and 5a, respectively, in Table I); (4) and (5) two conformations generated with a step size of 3 Å and initial velocities that were largest in the r_c direction and isotropic in all three directions, respectively (conformations 4 and 5 are denoted as 6a and 7a, respectively, in Table I).

In folding each of the above expanded conformations, we first minimized the energy of the conformation with the ECEPP/3 potential only, corresponding to a weight (K) equal to zero. After the structures reached local minima, which were still expanded, the dimensional constraints were added to the potential function in further minimizations, and as described in the Methods section, the weight K was introduced with an empirical value of 1 kcal/mol/Å^2. The constrained principal radii for the structures were decreased in 5–10 steps until the radii reached the values of the crystal structure of BPTI. After each step, the weight K was increased by a factor of 1.5. The longer the distance between the initial and final structures (measured by their principal radii), the more steps were usually taken. When the dimensional parameters of the structures approached those of the crystal structure (it should be noted that the principal radii of the final minimized structures were rarely exactly equal to those of the crystal structure), the dimensional contraints were kept at the principal radii of the crystal structure, and the weight was increased several more times until an increase in the weight caused no further change in the minimized structure. The structural properties of the final minimized structures depend on the number of steps taken in reducing the dimensional constraints to those of the crystal structure. From trial calculations, however, we found that this dependence is not very great; the qualitative properties of the minimized structures with 5 up to 15 steps of compression are basically similar. No systematic serach was made in optimizing the minimization schedules (*i.e.*, the parameter K and the number of steps). The results reported below are the best ones of a few, and sometimes just one, trial minimizations.

In Table I, the energy and structural parameters of the minimized structures are listed as 3b–7b, starting from the corresponding expanded conformations of 3a–7a. During these minimizations, the [30,51] disulfide bond was kept open. It can be seen from Table I that the minimized structures are reasonably close to the native structure of BPTI in terms of dimensional parameters, energies, and rms deviations. Structures 3a–6a, the most expanded conformations generated in

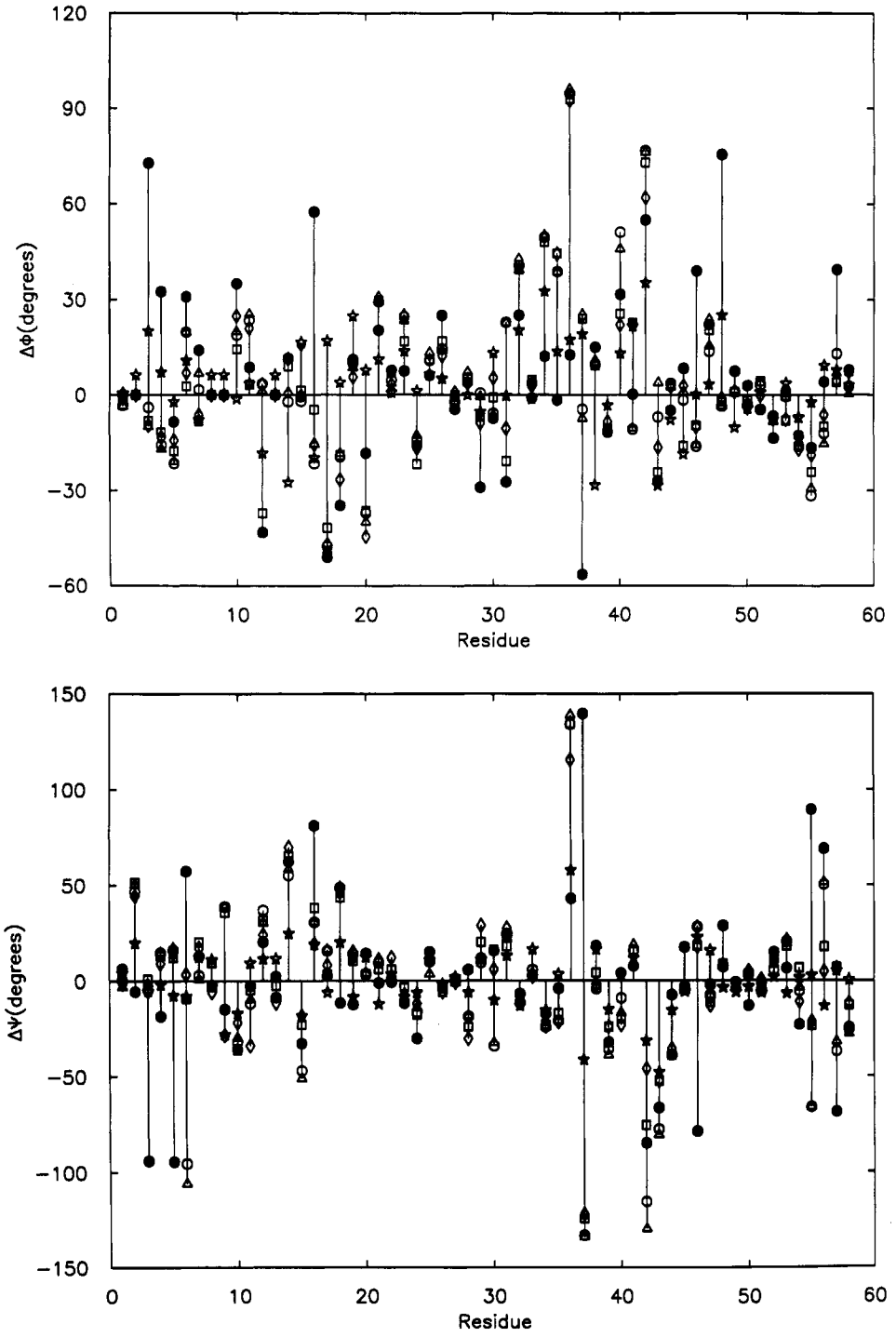

FIGURE 7: Differences ($\Delta\phi$ (a, top) and $\Delta\psi$ (b, bottom)) between the backbone dihedral angles of the expanded conformations and their values in native BPTI. The squares, open circles, triangles, stars, and diamonds are for the expanded structures 3a, 4a, 5a, 6a, and 7a, respectively, in Table I. The filled circles are for the most expanded structure on the trajectory of curve 3 in Figure 3, which failed to refold to the native structure.

separate dynamic runs, all have been refolded successfully to correct compact structures with the same chain threading as in the native structure, with rms deviations of 2.0–3.1 Å from the native structure for all C^{α} atoms and energies within 10 kcal/mol of that of the minimized crystal structure without the [30,51] disulfide bond. For the unfolding simulation that led to structure 7a and other expanded conformations, only structure 7a and those with $r_c < 40$ Å could be refolded correctly to the native structure. The rms deviations calculated for all atoms (backbone and side chains) between the minimized structures and the reference structure (not shown here) are nearly consistently 0.5 Å higher than those for only the C^{α} atoms. The averaged rms derivation for the side-chain torsional angles of the five refolded structures with respect to the native structure is 10°. In general, when the most expanded conformation can be refolded to the native structure with the minimization procedure, those on the unfolding path with less expanded conformations can also be folded to the native conformation.

To illustrate how close the refolded structure is to the native structure, Figure 8 shows an overlap drawing of the standard geometry version of the ECEPP/3-minimized crystal structure of BPTI (the thick chain) and the refolded structure of 6c in Table I (the thin chain). Structure 6c was generated from 6b by minimization with the [30,51] disulfide bond closed. It can be seen that the refolded structure agrees with the original

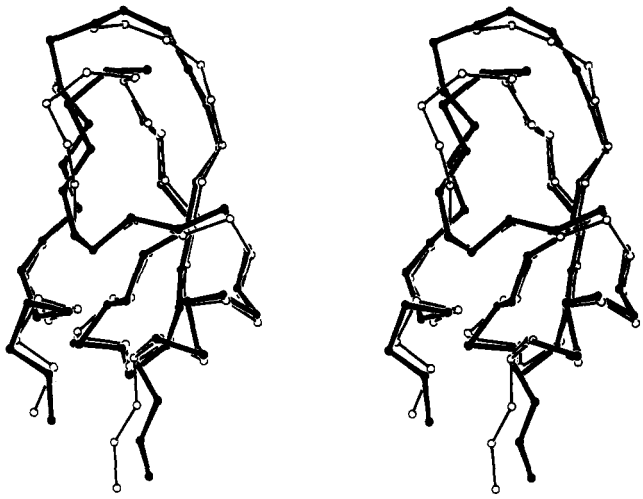

FIGURE 8: Comparison of the standard geometry version of the ECEPP/3-minimized crystal structure of BPTI (thick chain) and the refolded BPTI structure (thin chain). The latter corresponds to structure 6c in Table I, which was generated by minimizing the energy of structure 6b of Table I with both constraints of a closed [30,51] disulfide bond and the dimensional parameters of the crystal structure of BPTI.

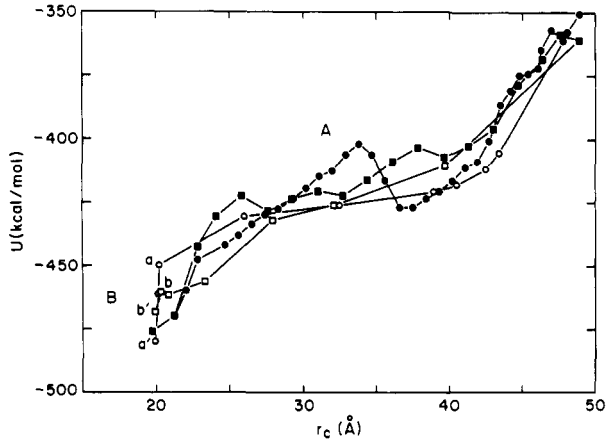

FIGURE 9: Comparison of the energy trajectories of unfolding and folding of BPTI as a function of the principal radius r_c. The curves are unfolding and folding trajectories. The curves with solid circles and squares correspond to curves 1 shown in Figure 3b and Figure 4b, respectively, and the curves with open circles and squares are the folding trajectories of the most expanded conformations on the above two unfolding paths, respectively. The letters a to a′ and b to b′ highlight energy drops in the folding trajectories without much change in the dimensions of the protein. The letters A and B indicate roughly two folding stages, in which the energy changes smoothly and abruptly, respectively, with the dimensions of the BPTI structure.

native structure remarkably well. Considering the fact that the global conformations of the starting structures are so different from that of the native structure (the longest dimensions are about 2.5 times that of the native structure), and that a minimization process can fold these structures so closely to the native structure, it may reasonably be speculated that there must be a strong internal tendency in the expanded conformations to fold into the native conformation.

For the five expanded conformations (3a–7a) listed in Table I, we discuss below more details about the refolding of structure 3a (and structure 4a in some cases); the deviation of this structure from the crystal structure at the final stage of minimization is intermediate between those of the other minimized structures. Figure 9 shows the energy trajectories of structures 3a and 4a of Table I as a function of r_c during their refolding by dimensional-constrained energy minimization. r_c is an indicator of the extent of folding. For comparison, the energy trajectories of unfolding in these two cases also shown in the figure: the curves in the figure with solid circles and squares are the trajectories of unfolding from the dynamic simulation, while the curves with open circles and squares are those of folding by energy minimization with the dimensional constraint. The trajectories with solid and open circles are for structure 3a, while the trajectories with solid and open squares correspond to structure 4a. Each circle or square represents an iteration step. It can be seen that the folding trajectory of structure 3a is quite similar to that of 4a, even though they start from different conformations and have different minimization schedules. More interestingly, the folding trajectories in both simulations are also close to those of unfolding, despite the difference in simulation procedures: one was done by the dynamic algorithm and the other by minimization with a dimensional constraint. These results indicate that both folding and unfolding processes proceed along similar paths on the potential surface. In general, it is observed that the energy profile in the folding process is slightly lower than that of unfolding, due to the fact that the former is carried out by energy minimization, which always seeks local energy minima, while the latter is carried out by the dynamic method, which is influenced by kinetic energy.

Even though the energy trajectories in Figure 9 have only one deep minimum at the position of the native conformation, in actual minimizations the conformations are found to be trapped at local energy minima at each step of iteration, *i.e.*, the positions of the symbols (squares and circles) in the figure. Without the dimensional constraint, the conformations could not be moved from these positions by energy minimization alone. But by decreasing the dimensional parameters of the constraint function, the structure can be moved to more compact conformations with lower energies by minimizing the energy and the constraint function. The contribution of the dimensional constraint is further manifested at the lower end of the energy trajectory in Figure 9, where the compactness of the minimized structure is only slightly different from that of the crystal structure. By increasing the weight of the dimensional-constraint function, the energies of the minimized structures show a significant drop at those positions, as indicated from point a to a′ on the open circle curve and from point b to b′ on the open square curve, while the dimensions of the structure change only slightly.

Figure 10 shows a series of conformations of BPTI at different stages of minimization, corresponding to the curve with open circles in Figure 9. The starting conformation for the minimization is the one shown in Figure 5f. A comparison of the two series of conformations in Figures 5 and 10 shows that, while the gross conformational changes in unfolding and refolding have some similarity, differences exist in the details of the structural changes. In particular, the compact conformation of Figure 10d, resulting directly from the folding of more expanded conformations, has dimensions that are very close to those of the native structure, but the arrangements of its polypeptide chain are quite different from those of the native structure and the structure has an rms deviation of 4.2 Å from the native. With an additional increase in the weight of the dimensional constraint, rearrangement of the structure takes place but without much change in the dimensions. The resulting structure, shown in Figure 10e, is much closer to that of the native structure, with an rms deviation of 2.6 Å.

The energy profiles of BPTI folding, as shown in Figure 9, may be divided roughly into two phases. In the first phase, denoted as A, the conformation is relatively expanded, and the energy decrease in the structure is associated primarily

FIGURE 10: Stereo diagrams illustrating the folding of BPTI by energy minimization with dimensional constraint. The conformations are represented by the C^{α} chain of BPTI. The corresponding energy trajectory is the curve with open circles in Figure 9. The starting conformation in the minimization is the one shown in Figure 5f. a–f correspond to the minimized structures with principal radii (r_a,r_b,r_c) equal to the following: (a) 11.4, 12.8, 40.5; (b) 11.2, 14.8, 32.5; (c) 11.8, 14.7, 25.9; (d) 10.8, 13.8, 20.1; (e) 10.6, 13.0, 19.9 Å without closure of the [30,51] disulfide bond; (f) 11.1, 12.5, 19.6 Å after the [30,51] disulfide bond was closed. a–c are top to bottom, respectively, on the left, and d–f are top to bottom, respectively, on the right.

with the decrease in the dimensions of the structure. In the second phase, denoted as B, the structure has reached a relatively compact state, and a further drop in energy takes place through a conformational rearrangement while the dimensions of the structure change only slightly.

In the above minimizations, the disulfide bond between residues 30 and 51 is not closed. Clearly, there is no obvious physical reason to introduce the disulfide constraint into a folding process when the two cysteine residues are distant in space. However, after the structures had been folded to the compact native-like structures without the [30,51] disulfide bond constraint, introduction of the [30,51] disulfide bond constraint and further minimization improved the agreement between the minimized structures and the native structure significantly. Entries 3c–7c in Table I show the structural and energetic parameters of the structures resulting from minimizing the energies of the corresponding structures 3b–7b with the constraints of the closed [30,51] disulfide bond and dimensions of the native structure. In all of these cases, the introduction of the [30,51] disulfide bond reduced the

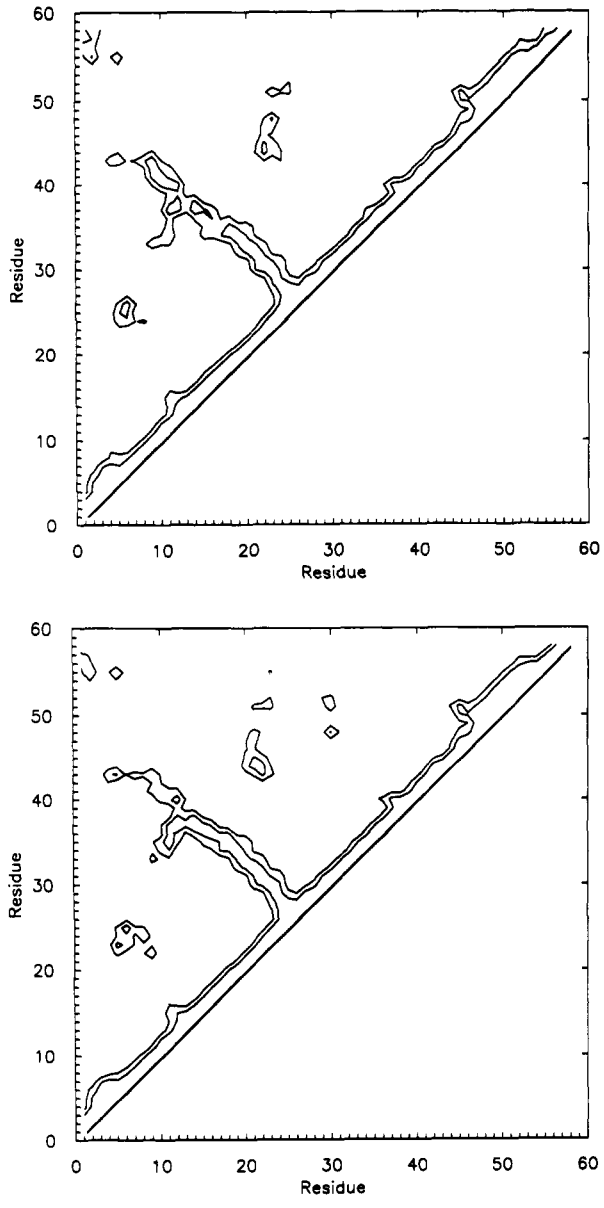

FIGURE 11: Comparison of the contour distance maps for the C^{α} atoms of the refolded structures without (a, top) and with (b, bottom) the [30,51] disulfide bond constraint. a and b correspond to the structures shown in Figure 10e,f, respectively. The inner and outer contours of the two maps represent distances of 6 and 8 Å, respectively.

rms deviations between the minimized structures and the native structure. The refolded structure shown in Figure 8 was obtained by minimizing the energy of structure 6b of Table I with the constraints of both the closed [30,51] disulfide bond and the dimensions of the crystal structure.

Figure 10f shows the structure obtained by energy minimization with the constraints of both the close [30,51] disulfide bond and the dimensions of the native structure, starting from the minimized structure without the [30,51] disulfide bond constraint shown in Figure 10e. In this case, the rms deviation of the minimized structure from the native structure decreased from 2.6 to 1.7 Å. The structural differences between Figure 10e,f are further illustrated by their C^{α} atom contour distance maps shown in Figure 11a,b, respectively. It can be seen that the distance map of Figure 11b for the minimized structure with the closed [30,51] disulfide bond constraint is much closer to that of the crystal structure (shown in Figure 6a) than the distance map of the structure lacking this disulfide bond (Figure 11a). Incidentally, in this case the energy of the minimized structure with the closed [30,51] disulfide bond is higher by about 15 kcal/mol than that without this constraint. But, in the other cases shown in Table I, the introduction of the [30,51] disulfide bond decreased the rms deviations and usually decreased the energy as well. The constraint of the [30,51] disulfide bond apparently has played an important role in determining the fine details of the native structure. Oas and Kim (1988) have reported that, by introduction of the [30,51] disulfide bond, two fragments of BPTI, one corresponding to the short α-helix segment and the other corresponding to the β-sheet segment, can form a packed structure similar to the corresponding part of the native structure of BPTI, while when no such disulfide bond exists, the two fragments do not form a closely compacted compound. This experiment suggests an important role for the [30,51] disulfide bond in fixing the native fold of BPTI.

The minimization procedure was not successful in refolding two of a total of seven most expanded conformations obtained from separate unfolding simulations. One is the most expanded conformation from the trajectory of case 5 described at the beginning of this section; only a less expanded conformation on this unfolding trajectory (the one denoted as 7a in Table I, but which is still quite expanded, with its longest principal radius being twice that of the native) can be folded correctly. The other case is the expanded conformation on the unfolding path shown by curves 3 in Figure 3a,b. On that unfolding path, even less expanded conformations (r_c larger than 27 Å) cannot be refolded to the native structure by minimization with the dimensional constraints. Such an abnormality is perhaps implied by its energy trajectory, shown by curve 3 of Figure 3b, which is nearly 30 kcal/mol higher than that of the other trajectories.

The backbone torsional angles of the structure which failed to be refolded correctly are indicated by filled circles in Figure 7a,b. The distribution of the torsional angle deviations in the failed structure is somewhat different from those of the other expanded structures. However, no particularly large deviations appear in the torsional angles of this unfoldable conformation as compared to other expanded conformations. Within our sampling range, we did not observe a clear correlation between the unfolding simulation step sizes and the torsional angle deviations in the resulting expanded conformations. It appears that whether or not an expanded conformation can be refolded to the native structure is determined by some global conformational changes. The folding of the expanded conformations in the dimensional-constraint energy minimization is a very delicate process. By examining the stereo diagrams of the series of folded conformations (one of the series is shown in Figure 10), we found that the folding of the expanded conformations to the native structure involves a rotation of the loop around an axis passing approximately through residues 14 and 38 and a twist of the loop. In the successfully folded situations, the rotation takes place first and is followed by the twist. In the failed cases, on the other hand, this folding order is always violated in one way or another. Because the minimization procedure has limited control over how the conformational changes take place, once folding adopts a wrong path in the middle of the process, the final structure is always a nonnative conformation.

DISCUSSION

A. Unfolding Native Structures of Proteins by Dimensional Expansion. Protein folding usually occurs as a cooperative conformational transition. As a result, many obligatory intermediates in the folding process are difficult to isolate for study. However, because such intermediates hold the key to the kinetics of protein folding, it is of great interest to devise

novel experimental or theoretical methods for investigating the folding intermediates of proteins. Conventional simulation methods, such as molecular dynamics, would require an extremely large number of steps to simulate protein folding or unfolding. The present procedure takes advantage of the fact that protein unfolding and folding often involves a dimensional change of the structure. By using the dimensional variables of the proteins, *i.e.*, the principal radii, and by formulating protein unfolding as a uniform deformation of the structure in the directions of the three principal axes, a simple equation of motion has been obtained. Using the backward-Euler scheme to solve this dynamic equation, the unfolding process of a protein can be simulated in a small number of steps, usually several tens to a few hundreds. Thus, our procedure provides an effective tool for studying the unfolding process of proteins and for gaining information about the folding intermediates.

Compared to other studies, our simulations provide some new information about the unfolding of BPTI. For example, Daggett and Levitt (1992) have carried out an MD simulation of BPTI unfolding in water, in which the dynamics of all atoms in the system were simulated. At a relatively high temperature (500 K), the dimensions of BPTI were increased by about 25% after a few hundreds of picoseconds simulations. The major conclusion of the Daggett–Levitt MD simulations was that the (initially) unfolded BPTI is a "molten globule" state. Our results suggest that the compact molten globule proposed earlier may represent *but one* possible form encountered in BPTI unfolding. The expanded conformations revealed here could be favorable folding intermediates, as judged by the facts that they have relatively low energies and that they can be reproducibly generated and reversibly folded back to the native structure. This is similar to what is observed experimentally in other proteins (Jeng & Englander, 1991; Haynie & Freire, 1993), where the unfolded states of the protein were characterized as a molten globule under some experimental conditions and as an expanded intermediate under other conditions. The implication of the expanded intermediate in protein folding is discussed below. As a technical point, we note that a certain similarity exists between the partially unfolded BPTI reported by Daggett and Levitt and the early unfolded structures of BPTI in our simulations when all disulfide bonds are intact (Figure 2), when the [14,38] disulfide bond is broken, and when all three disulfide bonds are broken in BPTI (results not shown). In these cases, the expansion of the native structure of BPTI initially occurs in the direction of the second largest principal radius.

An oversimplification of our procedure is that many possible side-chain rotations in protein unfolding are neglected. At normal temperatures, such rotations are expected to occur. However, simplification of the side-chain conformations by fixing them in their native states has the technical advantage that overall conformational changes in protein unfolding can be more easily demonstrated. If all of the side-chain conformational changes are simulated, it is not clear whether the large conformational unfolding observed here can be obtained. Another shortcoming of our procedure is that the temperature of the dynamic system is not rigorously defined. This problem can be overcome when the frictional coefficient of the protein is determined by other methods, as discussed in the Methods section. A limitation of the present procedure is that, because the procedure takes the dimensional parameters of a protein as the primary variables, it may not be useful for treating those unfolding processes that do not involve a dimensional change. Other technical problems with this procedure arise because the assumption of uniform deformation does not hold exactly in numerical simulations; hence, errors in the dynamic trajectories can accumulate.

B. Folding of a Protein by Minimization with Dimensional Constraints. This study has shown that the dimensional-constraint function in energy minimizations of protein structures can facilitate the location of the native structure. As described in previous sections, in minimizations without volume constraints, the structures are often trapped in local minima. In many of these situations, the structures are trapped in minor and shallow minima which are even invisible when the step size of the folding algorithm is large. By decreasing the dimensional parameters of the constraint function, the structure can be driven out of local minima, and the final minimized structures can be closer to the native structure than those obtained without the dimensional constraints.

With regard to the dimensional parameters used, it is useful to point out that the three principal radii of the native structure of a protein may be calculated empirically from the number of residues in a protein, on the basis of a survey of the crystal structures of proteins (Hao *et al.*, 1992). In the present calculations, when the dimensional parameters of the constraint function were taken as the empirical values resulting from our previous procedure (Hao *et al.*, 1992), the minimized BPTI structures had folding patterns similar to that of the native structure, such as is shown in Figure 10d. Folding from the expanded conformations to the compact structure can be driven by a set of increasingly compact dimensional constraints, and the dimensional parameters derived from the average protein structure (Hao *et al.*, 1992) are sufficient for this purpose. However, the rms deviations of the minimized structures, obtained with the empirical constraint parameters, are higher than those obtained with the dimensional constraints of the crystal structure of BPTI. It appears that the empirical dimensional parameters are useful for folding expanded conformations of proteins to compact structures by dimensional-constraint minimization, but are not good enough for promoting the correct structural rearrangements in compact structures to reach the native structure.

The contributions made by the dimensional constraint are global effects in the native structures of proteins. Such effects arise from interactions between atoms or residues which are distant in sequence but proximal in space. Global effects cannot be completely accounted for by interactions between neighboring residues and even general pairwise interactions; such effects involve cooperative interactions among a large number of residues. In an earlier study, Burgess and Scheraga (1975) carried out some related calculations on BPTI. They initially assigned the local conformational *states* for all residues of BPTI to be the same as those of the native structure, with the dihedral angles ϕ and ψ in each conformational state being average values rather than the experimental values. They minimized the energy of the resulting structure. The minimized structure that they obtained, however, was very extended and the overall chain folding was not native. Clearly, even though the local chain states had been treated correctly, a major element in the native structure of a protein, *viz.*, the global effect, was neglected. Conceivably, had the global effects been considered in the earlier study, the calculated structure of BPTI might have been closer to the native structure. The volume constraint is a simplistic but effective way to account for the global features of the native structures of proteins.

C. The Native Structure of BPTI and Its Stable Intermediate without the [30,51] Disulfide Bond. Our folding study provides some insights into the nature of the native structure of BPTI. The energy valley in which the native structure lies is probably quite broad, if trivial local energy

minima that have the same qualitative conformations as the native structure are ignored. At the scale of the step size of our minimizations, the energy profiles of folding from the various expanded conformations to the native structure do not show significant deep local energy minima. All incorrectly folded conformations are found to have distinctly higher energies than does the native structure. The predicting power of local minimization techniques for protein structures is quite limited, because such procedures are very easily trapped in local minima. However, if a minimization approach does produce a large conformational change in a certain starting structure and leads to the global minimum, it usually indicates that there is a strong tendency for the starting structure to move downhill toward the energy minimum. The various open loop conformations obtained by the dynamic procedure show such a tendency toward the native structure when minimized with the volume constraint.

The results obtained here also shed some light on the question of why the intermediate of BPTI without the [30,51] disulfide bond can fold to a structure that is very similar to the native structure (States *et al.*, 1984; Kosen *et al.*, 1992). Minimization of the energies of the expanded conformations of this intermediate easily produces the native structure, without the constraint of the [30,51] disulfide bond. The interloop interactions in the expanded structure dictate the folding path that leads quite accurately to the native conformation. This result is consistent with, and may be a partial explanation of, the high yield (over 80% under appropriate conditions) of the folded disulfide intermediate of BPTI without the [30,51] disulfide bond obtained in oxidizing reduced BPTI (States *et al.*, 1984). Furthermore, the refolded structure of the BPTI intermediate lacking the [30,51] disulfide bond is very similar to the intact native structure, indicating that the lowest energy conformation of this intermediate must be similar to the native structure of BPTI. This is in agreement with the solution NMR structure of the disulfide intermediate of BPTI without the [30,51] disulfide bond (States *et al.*, 1984) and with the crystal structure of a mutant BPTI that has residues 30 and 51 replaced by alanines, which is very close to the native structure of wild-type BPTI, with the maximum atomic displacements of atoms between the two structures being less than 0.5 Å (Eigenbrot *et al.*, 1990). The minimization results explain the stability of the intermediate of BPTI without the [30,51] disulfide bond, which can be stable for weeks under appropriate solvent conditions (States *et al.*, 1984; Creighton, 1990; Weissman & Kim, 1991).

D. Folding Pathway of Proteins to the Native Structures. In contrast to earlier molecular dynamics simulations on BPTI unfolding (Daggett & Levitt, 1992), the present study reveals an extended intermediate in BPTI unfolding. It is interesting to note that both the molten globule and the extended intermediates have been proposed as possible folding intermediates from experimental observations (Jeng & Englander, 1991; Haynie & Freire, 1992). It had been suggested that the extended intermediates would occur earlier than the molten globular state in protein folding (Jeng & Englander, 1991). The main characteristics of extended intermediates are that the seconday structural elements of the native state are intact or largely intact, while the tertiary structural contacts are mostly or completely missing. This is exactly what is observed in the expanded conformations obtained in this study. The existence of the extended intermediates implies that some native structure units of the protein would be formed in the early events of protein folding. Such a result supports the hierarchic model of protein folding in which folding proceeds with progressive initiation of folding units, midsize motifs, and finally the compact native structure (Baldwin, 1986; Montelione & Scheraga, 1989).

It is of interest to discuss the folding of BPTI in more detail. Since most native secondary structures are preserved in the expanded conformations of BPTI, these structural units should be quite stable when they are formed. It may be speculated that the expanded conformations, such as the one shown in Figure 5f, can be more easily formed than tightly folded structures in the initial stages of BPTI folding. A loop is produced naturally once the [5,55] disulfide bond is formed. This loop would then be helpful in stabilizing some of the secondary structures. After the correct local structures are formed, the final folding to the native conformation might proceed automatically when hydrophobic interactions favor the compact conformation, as in the process simulated by the dimensional-constraint minimization.

The movements of the structural fragments in the expanded conformations of BPTI are not random, but predominate in certain directions. The longest dimension of the starting conformations is about 2.5 times the length of the native structure. The dimensional constraints, on the other hand, are only nonspecific compression forces; many different compact conformations can satisfy the same dimensional constraints. On the path from the expanded conformation to the native structure, there are certainly many local energy minima that can deflect the motion to a different path and lead to a wrong conformation. Yet in the majority of cases, the expanded conformations, obtained from the dynamic procedure under various simulation conditions, could be folded reasonably accurately to the native conformation. This suggests that there must be internal interactions in the expanded conformation which are strong enough to overcome local energy barriers and guide the conformation to fold in the correct direction.

A determining factor for the expanded conformations of BPTI to find their correct folding pathway is the interatomic interactions in the protein. Presumably, similar energetic factors might also exist in the folding processes of other proteins as well. Theoretically, it is of interest to distinguish the energetic contributions from entropic effects which originate from ring closure (Jacobson & Stockmayer, 1950; Schellman, 1955; Poland & Scheraga, 1965; Flory & Semlyen, 1966; Chan & Dill, 1989b). Some theories (Matheson & Scheraga, 1978; Fiebig & Dill, 1993) have emphasized the entropic effects of the ring closure factor and nonspecific interactions between long-range hydrophobic residues in determining the folding paths of a protein. But many simulations on reasonably long chains of simplified protein models have found that, in addition to those nonspecific interactions, some *energetic biases* on local chain states must be provided in order to fold a chain molecule into a unique "native" structure (Skolnick & Kolinski, 1991; Kolinski & Skolnick, 1992; Honeycutt & Thirumalai, 1992; Guo *et al.*, 1992; Siepmann & Sprik, 1992). Experimental data have also indicated that local chain interactions have important effects on the folding paths of some proteins (Milburn & Scheraga, 1988; Montelione & Scheraga, 1989). The problem of interest is to elucidate the concrete roles of the energetic driving forces in the folding processes of different proteins. Our refolding simulations on the expanded conformations of BPTI provide an example where intramolecular interactions determine the paths in the folding of a protein under nonspecific compression forces.

ACKNOWLEDGMENT

We thank Dr. D. M. Rothwarf for helpful comments on this manuscript.

APPENDIX A: THE KINETIC ENERGY OF PROTEIN UNFOLDING

By analogy with an analysis of the deformation of a continuous body, we may describe the unfolding of the native structure of a protein by the deformation rate of the structure in three orthogonal directions. For simplicity, we neglect the shearing components of the deformations in the protein structure. The deformation strains of the structure are defined as follows: if ρ is the relative displacement of two general volume elements from the undeformed state in the structure, the deformation strains, after neglecting the shearing components, can be expressed by the vector

$$\mathbf{D} \equiv \{D_x, D_y, D_z\} = \left\{\frac{\partial \rho}{\partial x}, \frac{\partial \rho}{\partial y}, \frac{\partial \rho}{\partial z}\right\} \tag{A1}$$

For simplicity, we consider here only the situation of a uniform deformation of the protein structure along each coordinate direction. The model is then a linear deformation scheme for the unfolding of protein structures, and the parameters (D_x, D_y, D_z) are similar to the expansion factors in Flory's theory (1953) of network swelling in the cases of anisotropic expansion.

The rate of deformation of the structure can be characterized by

$$\frac{d\mathbf{D}}{dt} = \{\dot{D}_x, \dot{D}_y, \dot{D}_z\} \tag{A2}$$

We fix the center of mass of the structure at the origin of the coordinate frame. Then, the apparent velocity of an atom with coordinates x_i, y_i, and z_i in the structure can be written as follows:

$$\begin{aligned} \dot{x}_i &= \dot{D}_x x_i \\ \dot{y}_i &= \dot{D}_y y_i \\ \dot{z}_i &= \dot{D}_z z_i \end{aligned} \tag{A3}$$

The kinetic energy of protein unfolding is given by

$$T = \frac{1}{2}\sum_{i=1}^{N} m_i(\dot{x}_i^2 + \dot{y}_i^2 + \dot{z}_i^2) \tag{A4}$$

where N is the total number of atoms in the structure. For simplicity, we define an average mass for the atoms in the protein by

$$\langle m \rangle = \frac{\sum_{i=1}^{N} m_i(\dot{x}_i^2 + \dot{y}_i^2 + \dot{z}_i^2)}{\sum_{i=1}^{N}(\dot{x}_i^2 + \dot{y}_i^2 + \dot{z}_i^2)} \tag{A5}$$

Then, from eqs A3 and A5, the kinetic energy of the protein can be written as

$$T = \frac{\langle m \rangle}{2}[\dot{D}_x^2\sum_{i=1}^{N} x_i^2 + \dot{D}_y^2\sum_{i=1}^{N} y_i^2 + \dot{D}_z^2\sum_{i}^{N} z_i^2] \tag{A6}$$

As it stands, the kinetic energy depends on the coordinates of the atoms in the structure.

However, if the coordinate axes of the structure are chosen to coincide with the three principal axes of the structure, then the three principal radii can be calculated as

$$\begin{aligned} r_a^2 &= \frac{1}{N}\sum_{i}^{N} x_i^2 \\ r_b^2 &= \frac{1}{N}\sum_{i}^{N} y_i^2 \\ r_c^2 &= \frac{1}{N}\sum_{i}^{N} z_i^2 \end{aligned} \tag{A7}$$

and the velocities of variation of the principal radii are given, similarly to eq A3, as

$$\begin{aligned} \dot{r}_a &= \dot{D}_x r_a \\ \dot{r}_b &= \dot{D}_y r_b \\ \dot{r}_c &= \dot{D}_z r_c \end{aligned} \tag{A8}$$

Now, applying the relationships of eqs A7 and A8 to eq A6, and after some rearranging, we obtain a simple expression for the kinetic energy of protein unfolding:

$$T = N\frac{\langle m \rangle}{2}[\dot{r}_a^2 + \dot{r}_b^2 + \dot{r}_c^2] \tag{A9}$$

The significance of this result is that the kinetic energy depends only on the velocities of the three principal radii, except for the approximation about the average mass. Then, by defining a Lagrangian as $L = T - U$, with U being the potential energy of the structure, the equation of motion for the three principal radii can be written as (Moore, 1983)

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{r}_\alpha}\right) - \frac{\partial L}{\partial r_\alpha} = 0 \qquad \alpha = a, b, c \tag{A10}$$

This leads directly to eq 3 of the main text.

The condition for eq A10 to hold is that the deformation of the protein should be uniform, as specified above. However, because the protein structure is not a perfectly continuous body and the deformations are not exactly uniform, errors in the dynamic trajectory due to deviations from the assumed conditions can accumulate during numerical simulations. To reduce the errors, the procedure should be used in simple unfolding processes, e.g., when the structure undergoes only expansion.

APPENDIX B: THE PRINCIPAL RADII AND THEIR DERIVATIVES

To speed up the convergence of minimization, the functional values as well as the first derivatives are needed. Because the function being minimized contains the dimensional-constraint function, the values of the principal radii as well as their first derivatives must be evaluated. They are calculated as follows. In a general coordinate frame, the principal radii of a structure are the square roots of the eigenvalues of the following symmetrical matrix:

$$[A] = \frac{1}{N}\begin{pmatrix} \sum_{i=1}^{N} x_i x_i & \sum_{i=1}^{N} x_i y_i & \sum_{i=1}^{N} x_i z_i \\ \sum_{i=1}^{N} y_i x_i & \sum_{i=1}^{N} y_i y_i & \sum_{i=1}^{N} y_i z_i \\ \sum_{i=1}^{N} z_i x_i & \sum_{i=1}^{N} z_i y_i & \sum_{i=1}^{N} z_i z_i \end{pmatrix} \tag{B1}$$

where (x_i, y_i, z_i) are the coordinates of the atoms with the center of mass of the structure as the origin and N is the number of atoms in the structure. The principal radii of the protein ellipsoid are equal to the square roots of the eigenvalues of the above matrix multiplied by a factor of $5^{1/2}$ (Hao *et al.*, 1992). Once the coordinates of a conformation are known, the principal radii of the structure are calculated directly according to eq B1. To calculate the derivatives of the principal radii, the eigenvalues, λ, of eq B1 are first written in the form of a cubic equation:

$$\lambda^3 + A\lambda^2 + B\lambda + C = 0 \tag{B2}$$

where A, B, and C can be expressed as explicit functions of the coordinates of the atoms. By differentiating eq B2 with

respect to the x-coordinate of the ith atom, x_i, after rearrangement the derivatives of λ with respect to x_i can be expressed as

$$\frac{\partial\lambda}{\partial x_i} = -\frac{(\partial A/\partial x_i)\lambda^2 + (\partial B/\partial x_i)\lambda + \partial C/\partial x_i}{3\lambda^2 + 2A\lambda + B} \quad \text{(B3)}$$

Similar expressions can be obtained for the y and z components of the coordinates of all atoms. Once the derivatives of λ are known, the derivatives of the principal radii as well as the functions (they are quadratic functions in the present procedure) containing these radii can easily be calculated. The derivatives with respect to Cartesian coordinates are first calculated and then transformed to the derivatives with respect to the torsional angles used in the minimization algorithm.

REFERENCES

Baldwin, R. L. (1986) *Trends Biochem. Sci. 11*, 6–9.

Burgess, A. W., & Scheraga, H. A. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 1221–1225.

Chan, H. S., & Dill, K. A. (1989a) *Macromolecules 22*, 4559–4573.

Chan, H. S., & Dill, K. A. (1989b) *J. Chem. Phys. 90*, 492–509.

Chou, K. C., Némethy, G., Pottle, M. S., & Scheraga, H. A. (1985) *Biochemistry 24*, 7948–7953.

Covell, D. G., & Jernigan, R. L. (1990) *Biochemistry 29*, 3287–3294.

Creighton, T. E. (1977) *J. Mol. Biol. 113*, 275–293.

Creighton, T. E. (1990) *Biochem. J. 270*, 1–16.

Creighton, T. E., & Goldenberg, D. P. (1984) *J. Mol. Biol. 179*, 497–526.

Daggett, V., & Levitt, M. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 5142–5146.

Dahlquist, G., & Bjorck, A. (1974) *Numerical Methods*, Prentice-Hall, Englewood Cliffs, NJ.

Eigenbrot, C., Randal, M., & Kossiakoff, A. A. (1990) *Protein Eng. 3*, 591–598.

Fiebig, K. M., & Dill, K. A. (1993) *J. Chem. Phys. 98*, 3475–3487.

Flory, P. J. (1953) *Principles of Polymer Chemistry*, pp 576–590, Cornell University Press, Ithaca, NY.

Flory, P. J. (1972) in *Polymerization in Biological Systems*, *Ciba Foundation Symposium 7*, pp 109–121, ASP Press, Amsterdam.

Flory, P. J., & Semlyen, J. A. (1966) *J. Am. Chem. Soc. 88*, 3209–3212.

Gay, D. M. (1983) *ACM Trans. Math. Software 9*, 503–524.

Gō, N. (1980) *Proc. Jpn. Acad., Ser. B 56*, 414–419.

Gregoret, L. M., & Cohen, F. E. (1991) *J. Mol. Biol. 219*, 109–122.

Guo, Z., Thirumalai, D., & Honeycutt, J. D. (1992) *J. Chem. Phys. 97*, 525–535.

Hao, M.-H., Rackovsky, S., Liwo, A., Pincus, M. R., & Scheraga, H. A. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 6614–6618.

Hawley, S. A., & Mitchell, R. M. (1975) *Biochemistry 14*, 3257–3264.

Haynie, D. T., & Freire, E. (1993) *Proteins: Struct. Funct., Genet. 16*, 115–140.

Honeycutt, J. D., & Thirumalai, D. (1992) *Biopolymers 32*, 695–709.

Jacobson, H., & Stockmayer, W. H. (1950) *J. Chem. Phys. 18*, 1600–1606.

Jeng, M.-F., & Englander, S. W. (1991) *J. Mol. Biol. 221*, 1045–1061.

Kauzmann, W. (1959) *Adv. Protein Chem. 14*, 1–63.

Kolinski, A., & Skolnick, J. (1992) *J. Chem. Phys. 97*, 9412–9426.

Kosen, P. A., Marks, C. B., Falick, A. M., Anderson, S., & Kuntz, I. D. (1992) *Biochemistry 31*, 5705–5717.

Levitt, M. (1983) *J. Mol. Biol. 170*, 723–764.

Levitt, M., & Warshel, A. (1975) *Nature 253*, 694–698.

Liwo, A., Pincus, M. R., Wawak, R., Rackovsky, S., & Scheraga, H. A. (1993) *Protein Sci.* (in press).

Matheson, R. R., Jr., & Scheraga, H. A. (1978) *Macromolecules 11*, 819–829.

McCammon, J. A., & Harvey, S. C. (1987) *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, Cambridge, UK.

Meirovitch, H., & Scheraga, H. A. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 6584–6587.

Milburn, P. J., & Scheraga, H. A. (1988) *J. Protein Chem. 7*, 377–398.

Momany, F. A., McGuire, R. F., Burgess, A. W., & Scheraga, H. A. (1975) *J. Phys. Chem. 79*, 2361–2381.

Montelione, G. T., & Scheraga, H. A. (1989) *Acc. Chem. Res. 22*, 70–76.

Moore, E. N. (1983) *Theoretical Mechanics*, pp 35–50, 156–164, John Wiley & Sons, New York.

Nakazawa, T., Kawai, H., Okamoto, Y., & Fukugita, M. (1992) *Protein Eng. 5*, 495–503.

Némethy, G., Pottle, M. S., & Scheraga, H. A. (1983) *J. Phys. Chem. 87*, 1883–1887.

Némethy, G., Gibson, K. D., Palmer, K. A., Yoon, C. N., Paterlini, G., Zagari, A., Rumsey, S., & Scheraga, H. A. (1992) *J. Phys. Chem. 96*, 6472–6485.

Oas, T. G., & Kim, P. S. (1988) *Nature 336*, 42–48.

Peskin, C. S., & Schlick, T. (1989) *Comm. Pure Appl. Math. 42*, 1001–1031.

Poland, D. C., & Scheraga, H. A. (1965) *Biopolymers 3*, 379–399.

Schellman, J. A. (1955) *C. R. Trav. Lab. Carlsberg, Ser. Chim. 29*, 230–259.

Scheraga, H. A. (1992) in *Reviews in Computational Chemistry* (Lipkowitz, K. B., & Boyd, D. B., Eds.) Vol. III, pp 73–142, VCH Publishers, Inc., New York.

Scheraga, H. A., & Mandelkern, L. (1953) *J. Am. Chem. Soc. 75*, 179–184.

Schlick, T., & Olson, W. K. (1992) *J. Mol. Biol. 223*, 1089–1119.

Siepmann, J. I., & Sprik, M. (1992) *Chem. Phys. Lett. 199*, 220–224.

Skolnick, J., & Kolinski, A. (1991) *J. Mol. Biol. 221*, 499–531.

States, D. J., Creighton, T. E., Dobson, C. M., & Karplus, M. (1984) *J. Mol. Biol. 195*, 731–739.

Tanford, C. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 4175–4176.

Vásquez, M., & Scheraga, H. A. (1988a) *J. Biomol. Struct. Dyn. 5*, 705–756.

Vásquez, M., & Scheraga, H. A. (1988b) *J. Biomol. Struct. Dyn. 5*, 757–784.

Weissman, J. S., & Kim, P. S. (1991) *Science 253*, 1386–1393.

Wlodawer, A., Walter, J., Huber, R., & Sjolin, L. (1984) *J. Mol. Biol. 180*, 301–329.