DETERMINATION OF PROTEIN STRUCTURES

USING DISTANCE CONSTRAINTS

by

BAŞAK SAMUR
B.S. in Chemical Engineering, Boğaziçi University, 1996

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science
in
Chemical Engineering

Boğaziçi University
1998

ACKNOWLEDGMENTS

I would like to express my gratitude to my thesis advisor, Prof. Dr. Ivet Bahar, for her guidance, helpful criticisms and stimulating encouragement throughout my study. I am also grateful to Prof. Dr. Burak Erman for his interest, help and constructive suggestions in my work.

I am also thankful to Doç. Dr. Türkan Haliloğlu for the time she has devoted to reading and commenting on my thesis.

I wish to remember Doç. Dr. Ali Rana Atılgan for his moral support.

I am much indebted to Neşe Kurt, Melik Cumhur Demirel, Taner Zafer Şen, and Nevin Gerek who were always there to support me. They shared the good days and the bad ones. I would also like to thank Özlem Keskin and Banu Özkan for their comforting smiles that I always feel inside.

Finally, I want to dedicate this thesis to *my family* who believe in me and let me do things my way.

ABSTRACT

Constraints of different types are applied to a protein chain of 30 residues by using a low resolution model adopted from the cooperative kinematics (CK) model of polymers. The aim was to investigate the importance of the location and number of the distance restraints in reducing the conformational space accessible to the chain.

Initially, one constraint is applied to the chain ends. Then, the place of the constraint is changed to the middle of the chain. It was observed that the place of the constraint is very important in the formation of the tertiary structure. Also, the effect of applying a constraint between two residues which are close to each other along the chain is investigated. At the second stage of calculations, the number of constraints is increased to two, and the results confirm that increasing the number of constraints effectively restricts the conformational space.

A helical protein, referred to as the Rop dimer, is explored as a last case by applying three different distance restraints, which are originally calculated from the real distances of α-carbons in the Protein Databank structure. The resulting lowest energy conformation showed that a real protein structure can also be predicted with a reasonable accuracy level by using a limited number of distance restraints.

ÖZET

30 residüden oluşan bir protein zincirine değişik kısıtlamalar koyarak, orijinali polimerler için geliştirilmiş olan kooperatif kinematik (CK) modeli uygulanmıştır. Amaç proteine konulan uzaklık kısıtlamalarının yerinin ve sayısının zincirin konformasyon uzayını azaltmakta ne kadar önemli olduğunu araştırmaktı.

Önce, zincirin uçlarına tek bir tane kısıtlama getirilmiştir, daha sonra kısıtlamanın yeri zincirin ortası olarak değiştirilmiştir. Bunun sonucunda konulan kısıtın yerinin üç boyutlu yapının oluşumunda çok önemli olduğu görülmüştür. Son olarak zincirde birbirine yakın olan iki residüye konulan kısıtlama incelenmiştir. İkinci aşamada, kısıt sayısı ikiye çıkarılmıştır ve kısıtlama sayısının konformasyon uzayını azaltmakta önemli olduğu gözlenmiştir.

Son çalışma olarak Rop adlı dimerik proteinin yapısını geliştirilen analitik yöntemle elde etmeye çalışılmıştır. Bu işlem için, 1rop'un α-karbonları arasındaki (protein yapısı Bilgi Bankasından elde edilen) gerçek mesafelerden yararlanılarak üç kısıtlama kullanılmıştır. Sonuç olarak elde edilen en düşük enerjili konformasyon gerçek bir proteinin üç boyutlu yapısı kabul edilebilir bir doğrulukla, sınırlı mesafe kısıtlamaları kullanılarak tahmin edilebileceğini göstermiştir.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

E	Energy
E_{LJ}	Lennard-Jones potential
l_i	Bond vector of magnitude l_i pointing from the $(i-1)^{th}$ α-carbon
l_i	Length of i^{th} backbone bond
n	Number of residues
q	Set of generalized coordinates
R	Gas constant
r_{ij}	Distance between the i^{th} and j^{th} C^{α} atoms
r_{ij}^2	Squared distance between atoms i and j
t	Time
T_i	Transformation matrix for the i^{th} backbone bonds to the i^{th} α-carbon
W	Distribution function
W_ξ	Work done by chain against friction
Å	Angstrom
ΔE	Change in total energy
Δr_{1n}^{ext}	External constraint applied to end-to-end vector
ΔR_{ij}	Change in the distance between the i^{th} and j^{th} C^{α} atoms
$\mathcal{F}$	Rayleigh's dissipation function
Φ, Ψ, χ	Euler angles
ϕ_i	Dihedral angle of the i^{th} bond
L	Lagrangian
λ	Vector of the three Lagrange multipliers, λ_x, λ_y, λ_z
θ_i	Bond angle between l_i and l_{i+1}

ABBREVIATIONS

Ala	Alanine
Arg	Arganine
Asn	Asparagine
Asp	Aspartic acid
Cys	Cysteine
Gln	Glutamine
Glu	Glutamatic acid
Gly	Glycine
His	Histidine
Ile	Isoleucine
Lys	Lysine
Met	Methionine
NMR	Nuclear Magnetic Resonance
PDB	Protein Data Bank
Phe	Phenylalanine
Pro	Proline
RMS	Root-mean-square
Ser	Serine
SVD	Singular Value Decomposition
Thr	Threonine
Trp	Tryptophan
Tyr	Tyrosine
Val	Valine
1aan	Amicyanin
1aoz	Ascorbate oxidase
1azb	Azurin (Cu removed)
1cob	Superoxide dismutase (Co substituted)
1lfi	Lactoferrin
1mda	Methylamine dehydrogenase
1pcy	Plastocyanin

1rop	ROP dimer
1tho	*Escherichia coli* thioredoxin

1. INTRODUCTION

The estimation of the compact three-dimensional structure of proteins from knowledge of their amino acid sequences, shortly referred to as the "protein folding problem," is the major unsolved problem in structural molecular biology. Understanding the principles of protein folding is of great importance for engineering new proteins, decoding genetic information, designing new drugs and predicting the functions of thousands of proteins. The difficulty of the protein folding problem lies in the fact that an enormous number of conformations are possible for a given amino acid sequence. As first pointed out by Levinthal, despite the astronomically large number of accessible wrong conformations [1], proteins achieve the task of finding the correct native structure in milliseconds to seconds. Obviously, they must employ one or more strategies for consistently beating what appear to us to be enormous odds stacked against them [2].

From the large number of research papers published over the past ten years, one can infer that there is no shortage of experiments that can be done to study protein folding. The mainstays of structural analysis of folded proteins are X-ray crystallography and high resolution NMR spectroscopy. They often provide insight into the function of a protein. However, detailed questions on many dynamic aspects of enzymatic mechanisms such as regulation or substrate entry, remain unanswered when only static structures are available. Dynamic processes are crucial steps in the functioning of enzymes. Therefore, detailed information on the dynamics of a protein is necessary for a complete understanding of its function [3].

Simulations can help to obtain information on dynamics which cannot be provided by experimental techniques in a straightforward manner. By providing an initial guess of a structure for subsequent refinement, a successful method appears to be the NMR-based model building [4, 5]. This method could be used in homology modeling procedures, as well as in the construction of models based on a set of experimentally provided or theoretically predicted distance restraints. This class of models is complementary to more standard NMR structure determination algorithms [6,

7, 8, 9, 10]. There, one starts from a very large number of restraints and, while explicitly neglecting intraprotein interactions, the structure is derived from purely geometrical considerations. The most prevalent variants of this approach are distance geometry [10], and distance geometry supplemented by molecular dynamics refinement [9]. When one has a large number of distance restraints, there are clearly techniques of choice. On the other hand, especially on the early stages of NMR refinement when a relatively small number of restraints per residue are known, or when significant line broadening at the NMR spectrum refinement is present, one requires a protein model that already has many built-in protein-like regularities and interactions [11], [12].

This study began with the idea of elucidating the mechanism of folding of proteins to their native state by identifying the structural regions that may play a critical role in initiating the folding process. These structural regions would then be constrained by using different number of constraints so as to assess their importance in controlling the overall folding process.

One of the structural regions known to constitute such folding nuclei is metal ion binding sites in proteins [13]. These possess distinctive coordination geometries which are useful for specific purposes. So, an extensive initial study is presently done on copper-binding proteins. Eight copper-binding proteins from Protein Data Bank [14] are chosen. High resolution proteins were selected as a test group. The proteins investigated were amicyanin, ascorbate oxidase, azurin, superoxide dismutase, lactoferrin, methylamine.*herichia coli* thioredoxin. Their respective PDB codes are 1aan, 1aoz, 1azb, 1cob, 1lfi, 1mda, 1pcy and 1tho.

For each of the selected cu-binding protein, ligands were searched near central Cu atoms, in a 7 Å radius sphere. The regularity of the ligand positions around copper was remarkable. One of the examples of the regularity can be observed in the protein superoxide dismutase, 1cob. As it is seen from Figure 1.1, copper atom stays in the plane of three histidines and a forth histidine is positioned slightly on an upper plane. This coordination geometry was common to second, third, fourth, sixth and seventh coppers of the ascorbate oxidase and the second copper binding site of superoxide dismutase presently examined. Thus, copper ligands were grouped according to their geometrical shapes. By using distance restraints to obtain these distinctive coordination geometries, a low-resolution model may be constructed. Such a reduced

protein representation consists of a backbone of α-carbons. Local bond length, bond angles and Lennard-Jones interactions between non-bonded residue pairs can be readily assigned.

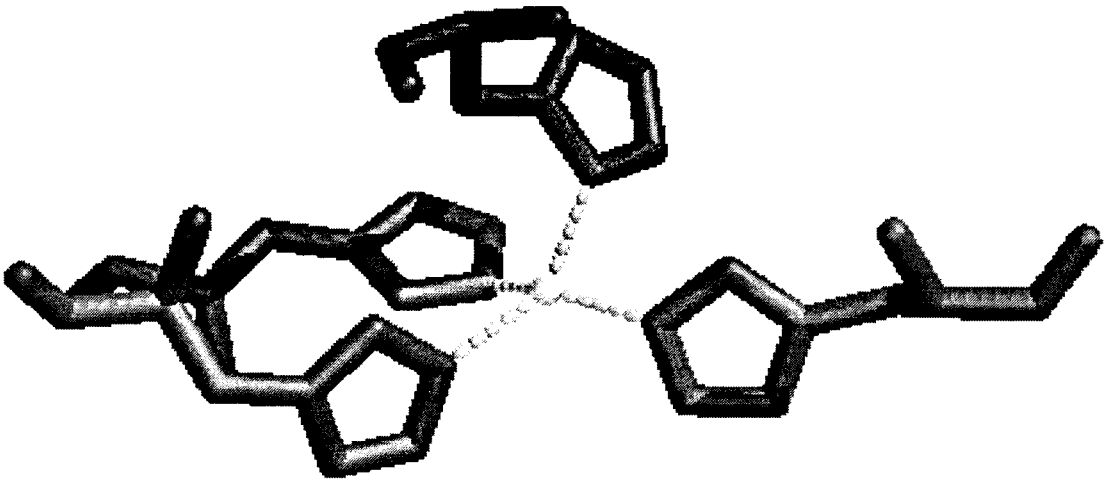

FIGURE 1.1. Ligands of a copper in superoxide dismutase (1cob)

The model is based on the original theory of Bahar and Erman developed for polymers [15]. A mathematical formulation was proposed for the study of the kinematics or geometry of motion of polymer chains with freely rotating bonds. The basic postulate of the proposed model is that, following any perturbation of an equilibrium configuration, the atoms rearrange cooperatively in space so as to minimize their square displacements. In this study, the same postulate is adopted and applied to protein chains, thus energy change is minimized during any structural transition of the protein. The configurational change is applied successively, in the form of linearized steps, subject to the condition that the energy change during each step is a minimum.

The plan of the present thesis is as follows: In the following section, an outline of structural characteristic of proteins will be presented. Simulation methods for investigation of polymer dynamics are discussed in the third chapter. The model and theory will be presented in the fourth and the fifth chapters, respectively. The simulations for predicting the low-resolution structure of proteins by using distance restraints will be presented in Chapter 6, and the results in Chapter 7. Conclusion and recommendations will be presented in the eighth chapter.

2. PROTEIN STRUCTURE : GENERAL INFORMATION

2.1. Definition of Proteins

Proteins are linear unbranched polymer chains made of monomer units. The monomers are the naturally occurring amino acids, which are successively linked to each other by peptide bonds. A chain composed of amino acids is called a polypeptide or simply a peptide. If a polypeptide consists of more than 30 amino acids, it is generally placed in the protein category [16].

All of the 20 amino acids have in common a central carbon atom (C^{α}) to which are attached a hydrogen atom, an amino group (NH_2), and a carboxyl group (COOH). The sidechain attached to the C^{α} through its fourth valency distinguishes one amino acid from another. (Figure 2.1)[17].

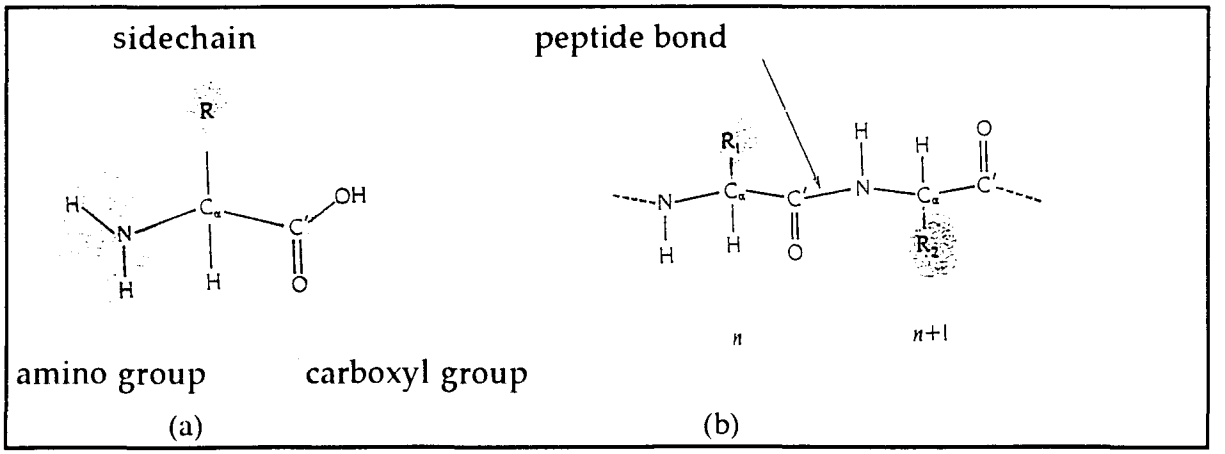

FIGURE 2.1. Schematic diagrams of an amino acid and a polypeptide chain. (a) Representation of an amino acid which consists of a central atom C^{α} bonded to an amino group NH_2, a carboxy group COOH, a hydrogen atom H, and a side chain, R. (b) Representation of polypeptide chain in which the carboxy group of amino acid n has formed a peptide bond, C-N, to the amino group of amino acid n+1, with the elimination of one water molecule.

The names of the twenty different types of amino acids are abbreviated with both a three letter and one letter code as shown in Table 2.1.

TABLE 2.1. The names and abbreviations of 20 amino acids.

Name	Three-letter code	One-letter code
Glycine	Gly	G
Alanine	Ala	A
Valine	Val	V
Isoleucine	Ile	I
Leucine	Leu	L
Serine	Ser	S
Threonine	Thr	T
Aspartatic acid	Asp	D
Asparagine	Asn	N
Glutamatic acid	Glu	E
Glutamine	Gln	Q
Lysine	Lys	K
Arginine	Arg	R
Cysteine	Cys	C
Methionine	Met	M
Phenylalanine	Phe	F
Tyrosine	Tyr	Y
Tryptophan	Trp	W
Histidine	His	H
Proline	Pro	P

It is customary to discuss proteins in terms of four levels of structure. The primary structure is the amino acid sequence. The secondary structure refers to regular local structures of linear segments of polypeptide chains, such as a helix or an extended strand. Tertiary structure is the overall topology of the folded polypeptide chain, and quaternary structure is the aggregation of the separate polypeptide chains of a protein. The crystal structure of a protein gives information about all four levels of structure, although independent knowledge of the primary structure is necessary for interpreting the electron density map.

2.2. Secondary Structures

2.2.1. Helical Structures

Helices are the most striking elements of protein secondary structure. If a polypeptide is twisted by the same amount about each of its C^{α} atoms, it assumes a helical conformation. As an alternative to specifying its ϕ and ψ angles, a helix may be characterized by the number, n, of peptide units per helical turn, and its pitch, p, the distance the helix rises along its axis per turn.

A helix has chirality; that is it may be either right-handed or left-handed (a right-handed helix turns in the direction of the fingers of a right hand curl when its thumb points along the helix axis in the direction that the helix rises).

The α-helix is a common secondary structure of both fibrous and globular proteins. For a polypeptide made from L-α-amino acid residues, the α–helix is right-handed with torsion angles of $\phi = +57^{\circ}$ and $\psi = +47^{\circ}$, n=3.6 residues per turn, and a pitch of 5.4 Å. Here the *trans* state is defined as $\phi = 180^{\circ}$.

Figure 2.2 indicates that the hydrogen bonds of an α-helix are arranged such that the peptide C=O bond of the n^{th} residue points along the helix towards the peptide N-H group of the $(n+4)^{th}$ residue. This results in a strong hydrogen bond that has the nearly optimum N··· O distance of 2.8 Å. In addition, the core of the α–helix is tightly packed such that its atoms are in van der Waals contact across the helix, thereby maximizing their association energies [18].

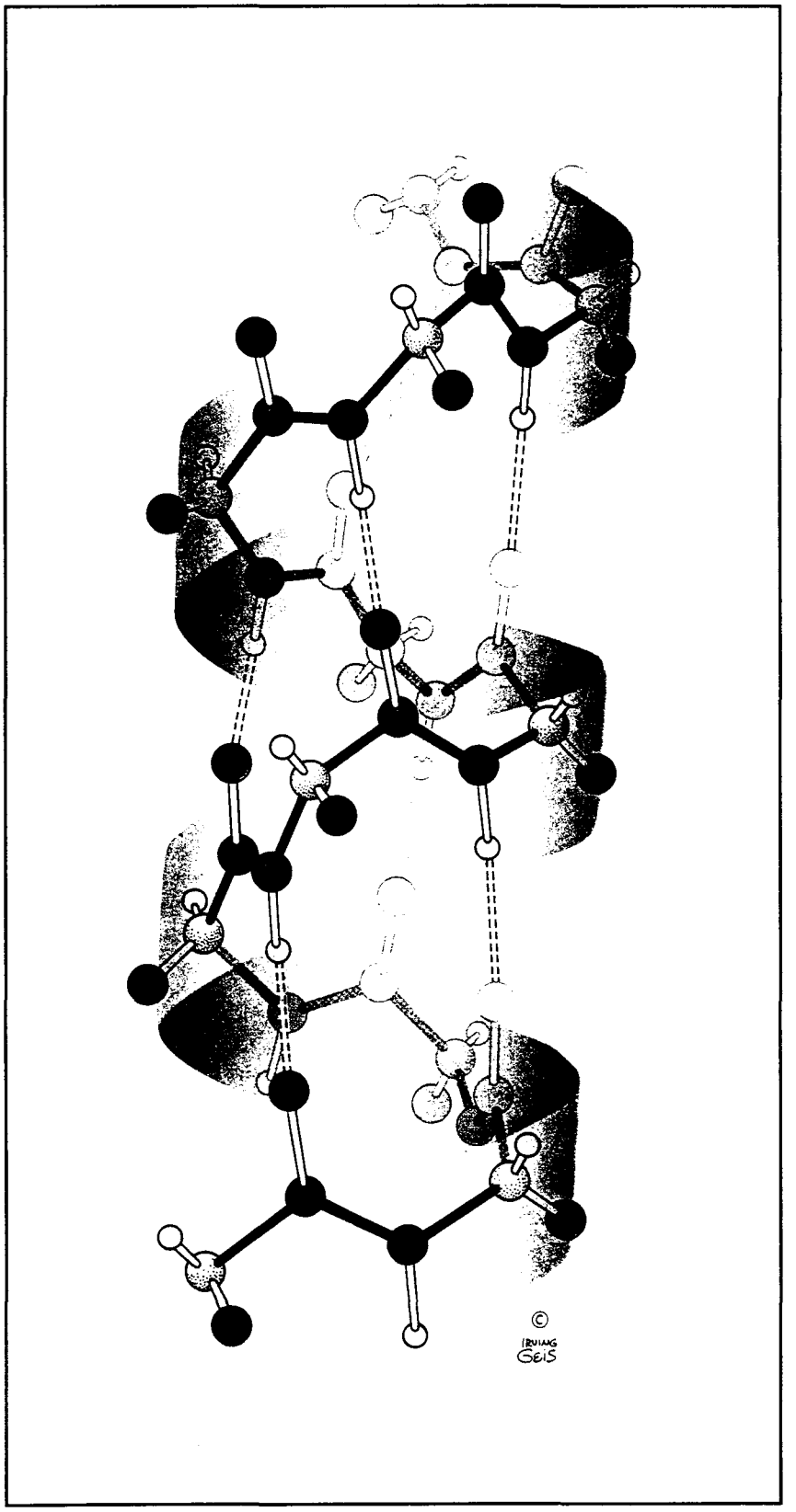

FIGURE 2.2. Right-handed α helix. Hydrogen bonds between the N-H groups and the C = O groups that are four residues apart along the polypeptide chain are indicated by dashed lines.

2.2.2. Beta Structures

As with the α–helix, the β-sheet conformation has repeating ϕ and ψ angles that fall in well-defined allowed regions of the Ramachandran diagram and utilizes the full hydrogen bonding capacity of the polypeptide backbone. In β-sheets, however, hydrogen bonding occurs between neighboring strands rather that within one strand as in α-helices.

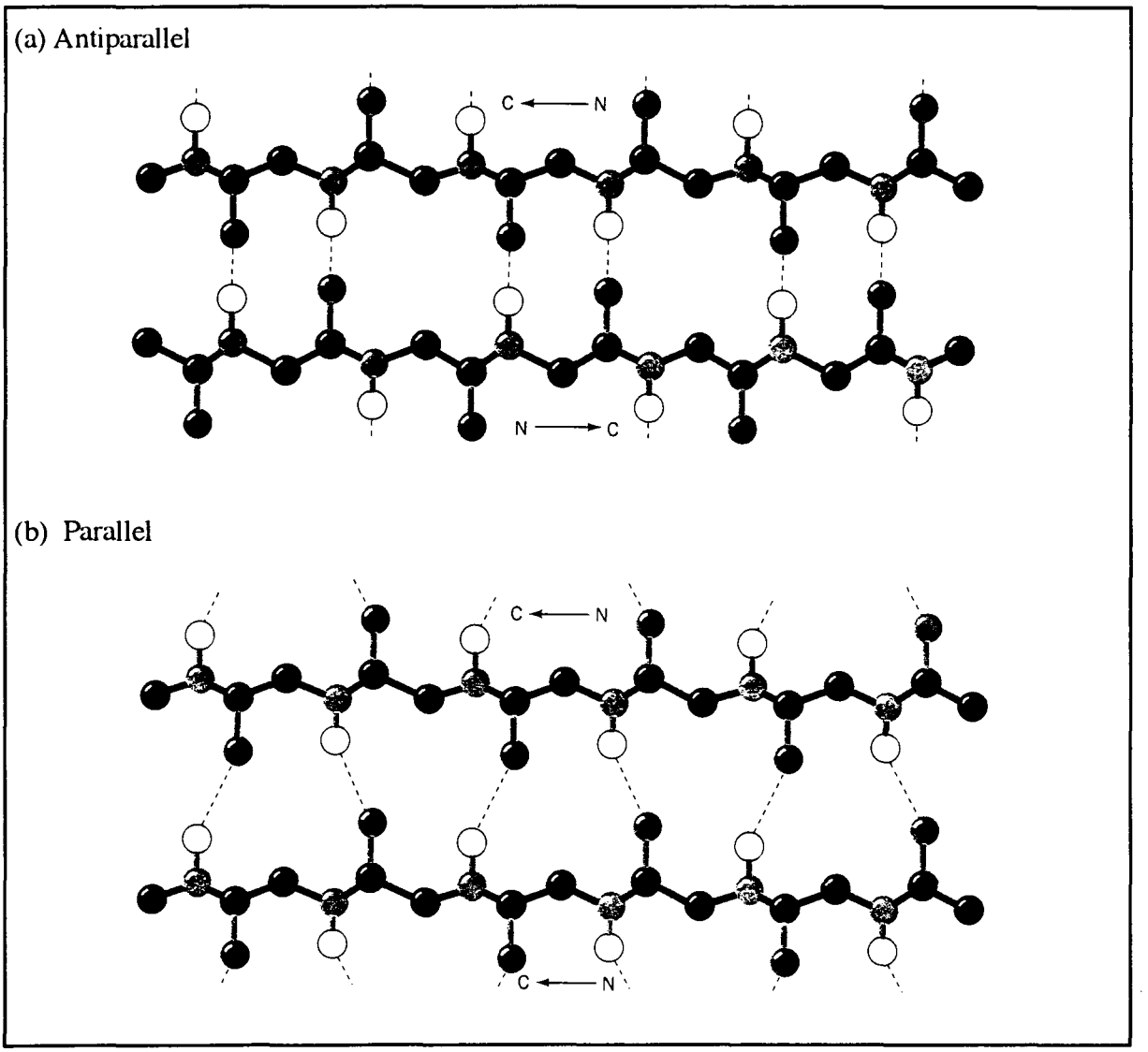

FIGURE 2.3. Hydrogen bond associations in β -sheets. Sidechains are omitted for clarity. (a) Antiparallel β-sheet. (b) Parallel β–sheet.

β-sheets exhibit in two varieties:

1. The antiparallel β–sheet, in which the neighboring hydrogen bonded strands run in opposite directions (Figure 2.3a).

2. The parallel β–sheet, in which the hydrogen bonded strands extend in the same direction (Figure 2.3b).

The conformations in which these β structures are optimally hydrogen bonded vary by about 30° from that of a fully extended polypeptide ($\phi = \psi = \pm$ 180°). Therefore, they have a rippled or pleated edge-on appearance, which accounts for the term "pleated sheet." Successive side chains of a strand extend to opposite sides of the pleated sheet with a two-residue repeat distance of 7.0 Å.

2.2.3. Nonrepetitive Structures

Regular secondary structures comprise around half of the average globular protein. The remaining polypeptide segments are said to have a coil or loop conformation. That is not to say, however, that these nonrepetitive secondary structures are any less ordered than are helices or β-sheets; they are simply irregular and hence more difficult to describe. Therefore, the term coil conformation should not be confused with the random coil, which refers to the totally disordered and rapidly fluctuating set of conformations assumed by denatured proteins and other polymers in solution.

Almost all proteins of >60 residues contain one or more loops of 6 to 16 residues that are not components of helices or β-sheets and whose end-to-end distances are < 10 Å. Such Ω loops (so named because they have the necked-in shape of the Greek upper case letter omega), which may contain reverse turns, are compact globular entities because their side chains tend to fill in the internal cavities. Since Ω loops are almost invariably located on the protein surface, they may have an important role in biological recognition processes.

2.3. Noncovalent Forces Determining Protein Structure

Noncovalent forces, discovered by van der Waals (1873) in an attempt to explain the deviation of a real gas from the ideal gas law, are of utmost importance for biological organisms. In particular, they drive the spontaneous folding of polypeptide and nucleic acid chains and the spontaneous formation of membranes. They mediate the mutual recognition of complementary molecular surfaces.

Noncovalent energies are one to three orders of magnitude smaller than covalent binding energies. They are difficult to measure and even more difficult to calculate. Moreover, a major contribution to protein stability originates from the surrounding liquid medium and is of an entropic nature. Therefore, the whole system, protein and solvent, has to be taken into account.

2.3.1. Dispersion Forces and Electron Shell Repulsion

Dispersion forces between any pair of atoms are attractive. Such forces occur between any pair of atoms, even between totally apolar ones. Each atom behaves like an oscillating dipole generated by electrons moving in relation to the nucleus. In a pair of atoms each dipole polarizes the opposing atom. As a consequence, the oscillators are coupled, giving rise to an attractive force between the atoms.

When bound in a molecule the polarizability of an atom is anisotropic. Therefore, the dispersion forces vary with the relative orientation of molecular groups. However, since the effects of orientation are small and difficult to measure, they are usually neglected.

Nonbonded electron shells repel each other. The attractive dispersion between a pair of nonbonded atoms are counterbalanced by the repulsion of the electronic shells. The repulsion is commonly approximated by a Lennard-

Jones potential using a term proportional to the m^{th} power of the reciprocal distance. For computations with proteins, the Lennard-Jones approximation with m = 12 is common approach although the use of m = 9 has the advantage of being more suitable for simulations because of avoiding excessively strong interactions at short separations. Together with London's term for dispersion forces, this results in the computationally simple 6-9 or 6-12 potentials [19].

2.3.2. Electrostatic Interactions

In proteins, the most commonly used method for calculating the electrostatic interactions is a straightforward application of Coulomb's law to all partial charges. This procedure is, however, very time consuming because of the large number of atoms in a protein and the long range of electrostatic interactions. But the calculation can be simplified. Since no monopoles, that is, free charges, are observed inside proteins (salt bridges are ion pairs and therefore dipoles), all partial charges may be expressed as dipoles. It can be seen from Equation 2.1 that the interaction between these decreases with at least the third power of the inverse distance [19].

$$E_{dip\text{-}dip} = \frac{332}{\varepsilon}\left[\frac{(\mu_1.\mu_2)}{R_{12}^3} - \frac{3(\mu_1.\mathbf{R}_{12}) \cdot (\mu_2.\mathbf{R}_{12})}{R_{12}^5}\right] \qquad (2.1)$$

Here μ_1 and μ_2 are dipole moment vectors expressed in Debye units, ε is the dielectric constant taken as 3 for nonpolar media, and $\mathbf{R}_{12}$ is the distance vector between partial charges.

Therefore, the actual range of electrostatic interactions is rather short, and the energy calculation can be restricted to the interaction between close neighbors. It could be argued that the weak long distance interactions can add up if dipoles are aligned as found with hydrogen bond dipoles in α-helices and

β-sheets. However, in β-sheets adjacent dipoles are antiparallel so that their fields cancel each other at long distances. In α-helices the dipoles form lines, canceling each other except for charges at both ends. For this reason antiparallel helices are more favorable than parallel ones. But this electrostatic contribution is small when compared with the total binding energy between α-helices.

2.3.3. Van der Waals Potentials

Van der Waals potentials comprise electron shell repulsion, dispersion forces, and electrostatic interactions. It is computationally convenient to combine all three nonbonded forces into a single, simple potential function, which is then called the "van der Waals potential" for historical reasons. To accomplish this, considerations of the electrostatic interaction have to be further simplified. First of all, only atoms sufficiently close to each other are taken as the pairs that must be considered. Moreover, electrostatic interactions are averaged over all relative orientations that are sterically allowed for two contacting groups, C = O and H - C, etc. Thus, the resulting contribution depends only on the distance between the contacting atoms. With this simplification, the van der Waals potential is isotropic and may be simply written as

$$E_{VW} = \frac{a_{ij}}{R_{ij}^{12}} - \frac{b_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{R_{ij}} \qquad (2.2)$$

Equation 2.2 contains three parameters: a_{ij} and b_{ij} are the Lennard-Jones parameters, and q_i, q_j are the effective charges of the contact partners.

The van der Waals potential allow for the definition of "van der Waals contact distances" between given atoms. Values of van der Waals radii are

usually measured from the smallest distances that can exist between neighboring, but not covalently bonded, atoms in the crystalline state; this distance is the sum of their respective van der Waals radii. Some typical values of van der Waals radii are given in Table 2.2 [20].

TABLE 2.2. Van der Waals radii of atoms found in proteins

Atom	Observed Range (Å)	Radius when singly bonded (Å)
Hydrogen	1.00 - 1.54	1.17
Oxygen	1.40 - 1.70	1.40
Nitrogen	1.55 - 1.60	1.55
Carbon	1.70 - 1.78	1.75
Sulfur	1.75 - 1.80	1.80

A range of values is given in Table 2.2 because the observed radius depends on the way in which the atoms are covalently bonded. For example, the van der Waals radius of a hydrogen atom varies from 1.0 Å when bonded to an aromatic carbon atom to 1.54 Å when bonded to a negative ion. Fortunately, these are extreme variations, and two atoms are generally in close van der Waals contact when the distance between them is approximately 0.8 Å greater than their separation when covalently bonded. The van der Waals radius is a minimal estimate of the size of an atom or molecule. Optimal van der Waals interactions generally occur at a distance that is about 1.2 Å greater than the covalent bond length [20].

2.4. Copper-Binding Proteins

2.4.1. Introduction

Recent investigations show that among the classes of interactions to be considered for designing proteins, hydrophobic interactions are not specific enough, hydrophilic interactions are typically too weak, and water interactions are always on the exterior. In terms of overall protein stability, there is a substantial advantage to concentrate on a nucleus with strong, directional interactions.

Metal ion sites in proteins exhibit strong directional preferences for their coordinate ligands. Metal binding sites can therefore constitute important structural building blocks for protein design [13]. Accordingly, metal ion sites in protein crystals, together with their coordination types, should be thoroughly investigated in order to assess their geometric characteristics. An extensive initial study is presently done on copper-binding proteins.

2.4.2. Investigation of Ligands near Copper Binding Sites of Protein

Eight high resolution copper-binding proteins from Protein Data Bank [14] were selected as a test group. The proteins investigated were amicyanin, ascorbate oxidase, azurin, superoxide dismutase, lactoferrin, methylamine dehydrogease, plastocyanin, *esclericlua coli* thioredoxin. Their respective PDB codes are 1aan, 1aoz, 1azb, 1cob, 1lfi, 1mda, 1pcy and 1tho. More information about the selected proteins is given in Table 2.3.

TABLE 2.3. Copper-binding proteins

PDB code	Name of the protein	Resolution (Å)	Residue number	Number of Coppers	Reference no
1aan	Amicyanin	2.0	105	1	21
1aoz	Ascorbate Oxidase	1.9	552	9	22
1azb	Azurin	2.2	129	2	23
1cob	Superoxide Dismutase	2.0	151	2	24
1lfi	Lactoferrin	2.1	691	2	25-26
1mda	Methylamine Dehydrogenase	2.5	268	2	27
1pcy	Plastocyanin	1.6	99	1	28-29
1tho	*Escherichia coli* Thioredoxin	2.3	109	1	30

The next step was the search for ligands near central Cu atoms, within a spherical volume of 7 Å radius. This search was done in two ways. First, only α-carbons were considered in finding the ligands of copper. As a second method, side groups of each residue were considered and new coordinates were assigned using the residue-specific R-group coordinates. It was found that the number of ligands near Cu atoms within this radius is quite large; then another search is done using 4 Å radius. There were 20 copper atoms or ions analyzed in these proteins as described in Table 2.3. The corresponding ligands are reported in Table 2.4.

Copper ligands were examined by using the package program UCSF MIDASPLUS [31], and it was seen that they could be grouped according to their geometrical shapes. They were divided into 2 groups. In the first group, the nearest ligands of copper were two histidines, a cysteine and a methionine. These four ligands were arranged approximately as a distorted tetrahedron around the copper atom. The copper atom resides in, or almost in, the plane of the two histidines and the cysteine and one of the axial positions is occupied by the methionine.

TABLE 2.4. Ligands of copper in copper-binding proteins within 4 Å radius sphere

PDB Code	Number of Cu	Ligands of Cu
1aan	1	His53, Cys92, His95, Met98
1aoz	1	His445A, Cys507A, His512A, Met517A
1aoz	2	His106A, His448A, His450A, His506A
1aoz	3	His62A, His104A, His448A, His508 A
1aoz	4	His60A, His62A, His448A, His450A
1aoz	5	His445B, Cys507B, His512B, Met517B
1aoz	6	Pro106B, His448B, His450B, His506B
1aoz	7	His62B, His104B, His448B, His508B
1aoz	8	His60B, His62B, His448B, His450B
1aoz	9	His286A, Pro287A, His286B, Pro287B
1azb	1	His46A, Cys12 A, His117A, Met121A
1azb	2	His46B, Cys112B, His117B, Met121B
1cob	1	His44A, His46A, His61A, His118A
1cob	2	His44B, His46B, His61B, His118B
1lfi	1	Asp60, Arg121, Thr122, His253.
1lfi	2	Asp395, Arg465, Thr466, His597
1mda	1	His53E, Cys92E, His95E, Met98E
1mda	2	His53F, Cys92F, His95F, Met98F
1pcy	1	His37, Cys84, His87, Met92
1tho	1	Ser1, Asp2, Lys3, Asp43

Copper of 1aan, both of the coppers of 1mda, second copper of 1azb, second and fifth coppers of 1aoz and first copper of 1pcy participate in this group (group 1). The coordination geometry of copper site for these proteins was also analyzed experimentally by different NMR techniques [32, 33]. The results are quite similar: the Cu ion or metal is immobilized within the protein framework by strong bonds with a cysteine and two histidines, being also axially coordinated by the methionine. An example of Group 1 coppers is shown in Figure 2.4.

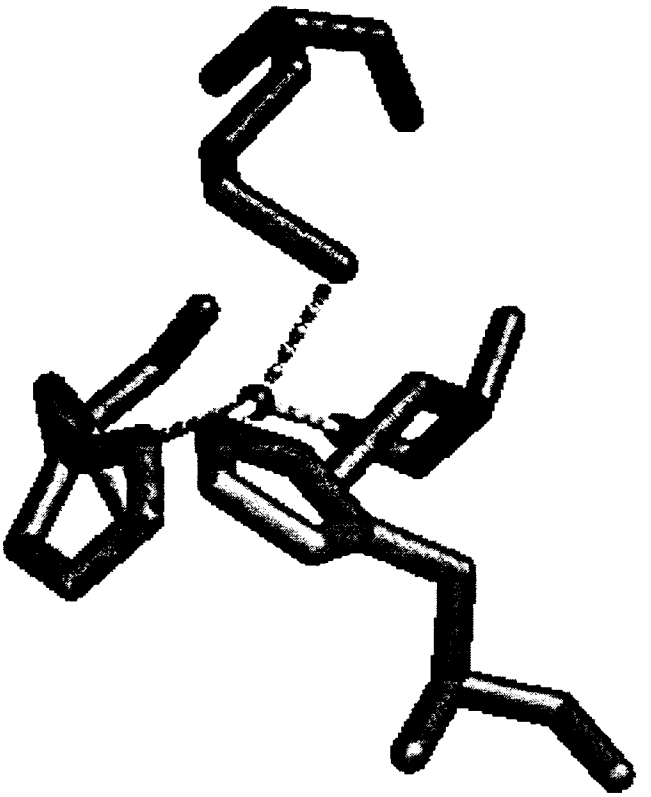

FIGURE 2.4. An example of group 1 copper ligands. Ligands of the first copper of methylamine dehydrogenase, 1mda

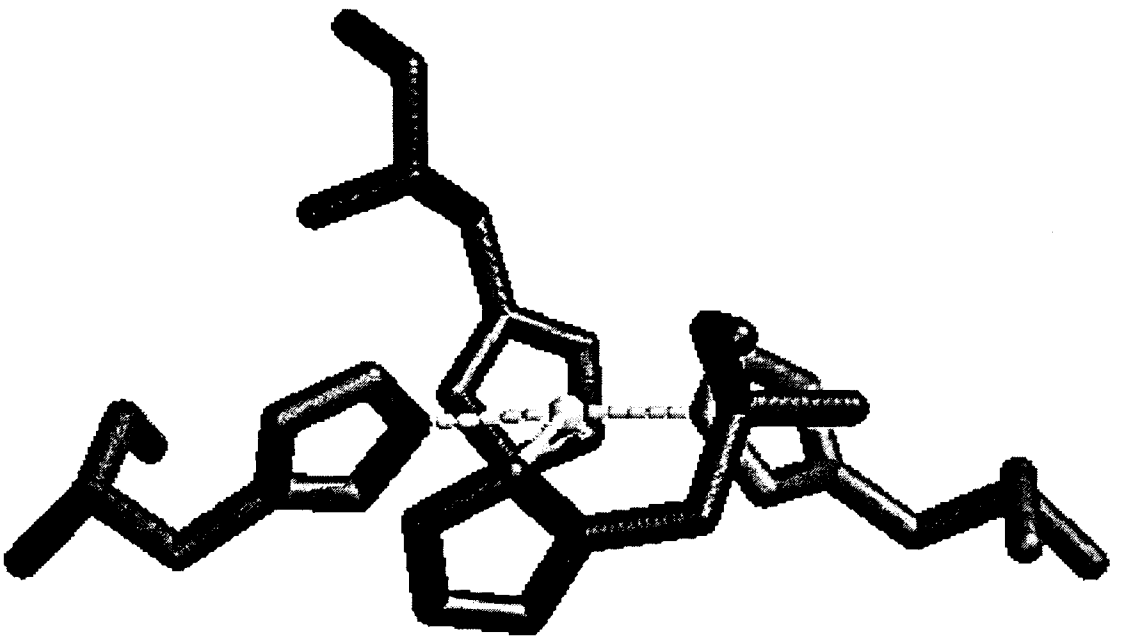

FIGURE 2.5. An example of group 2 copper ligands. Ligands of the fourth copper of ascorbate oxidase, 1aoz.

In the second group, second, third, fourth, sixth and seventh coppers of 1aoz, and both coppers of 1cob are listed. The copper atom is surrounded by four histidines in this group. For most of the coppers in this group, four histidines lie in the same plane with copper, but there are exceptions in which in one of the histidines lies slightly above the plane of the three other histidines and copper. An example of this group is shown in Figure 2.5.

Coppers of 1pcy and 1tho could not be ascribed to one of these groups. The geometrical coordination of the ligands were quite unique in these cases.

3. DYNAMICS OF POLYMERS IN BULK STATE

3.1. Computational Methods Used in Polymer Dynamics

The prerequisite for effective simulation is the utilization of an appropriate model which is complete enough to represent the true dynamics of the polymer but simple enough to extract information in reasonable computational time. A compromise between realistic models and efficient computation algorithms is necessary in order to search the configurational space of macromolecular systems. The simulation techniques used in dynamics of polymers are broadly classified as (i) those solving the real-time motion of a system of particles in space, molecular dynamics and Brownian dynamics methods. (ii) those employing analytical models to generate the most probable motion without intensive simulations. Examples of these models are the dynamic rotational isomeric states model [34, 35] and cooperative kinematics model [15, 36, 37, 38, 39].

3.2. Cooperative Kinematics

Cooperative kinematics (CK) is a computationally inexpensive method for observing the dynamic behavior in the bulk state. In a series of papers [15, 36, 38], the correlated motions of chain atoms in dense media were investigated by using cooperative kinematics. In these studies, the term dense media is used in reference to polymers in the bulk state, above the glass transition temperature. The femtosecond vibrations and fluctuations in bond lengths and bond angles are neglected. The most probable trajectory of chain atoms accompanying the rotameric transitions of a given bond from one rotational isomeric minimum to another is obtained by considering the

the frictional resistance exerted by the environment [38]. The basics of the CK are outlined below.

The Lagrange equations of motion for a chain in the presence of frictional forces is expressed as

$$\frac{d}{dt}\left[\frac{\partial L}{\partial \dot{q}_j}\right] - \frac{\partial L}{\partial q_j} + \frac{\partial \mathcal{F}}{\partial \dot{q}_j} = 0 \qquad (3.1)$$

where q represents the set of generalized coordinates, the dot denotes the time derivative, L is the Lagrangian and $\mathcal{F}$ is the Rayleigh's dissipation function whose gradient with respect to the velocity of any atom gives the frictional force experienced by that atom. Backbone atoms are numbered form 0 to n. The set q includes the following variables: (i) the position vector R_0 = col [X_0 Y_0 Z_0] of the zeroth atom relative to a laboratory-fixed coordinate system, (ii) the Euler angles, Φ, Ψ, χ, defining the absolute orientation of the chain in space, (iii) the internal torsional angles, φ_i. In a highly viscous medium, such as that experienced by a polymer chain in the bulk state, the kinetic energy term in L and the acceleration contribution may be neglected, and within the limits of this approximation, the Lagrange equation reduces to

$$\frac{\partial V}{\partial q_j} + \frac{\partial \mathcal{F}}{\partial \dot{q}_j} = 0 \qquad (3.2)$$

where V is the potential. Following this approximation, the problem of finding the optimal changes in the generalized coordinates in response to a torsional rotation leads to the solution of

$$\frac{\partial V}{\partial \varphi_i} + \frac{1}{2}\frac{\partial \delta W_\xi}{\partial \delta \varphi_i} = 0 \;\; ; \;\; 2 \leq i \leq (n-1)$$

$$\frac{\partial \delta W_\xi}{\partial \delta \Phi} = \frac{\partial \delta W_\xi}{\partial \delta \Psi} = \frac{\partial \delta W_\xi}{\partial \delta \chi} = \frac{\partial \delta W_\xi}{\partial \delta \mathbf{R}_0} = 0 \qquad (3.3)$$

Here δW_ξ denotes the differential work done by the chain against friction during small incremental displacements δR_i in chain units occurring within short time intervals δt. The potential V, being a function of internal coordinates only, does not appear in the second line of Equation 3.3. Detailed mathematical formalism of the CK approach can be found from the work of Bahar et al [38].

According to the CK approach, a bond in the middle of the chain is rotated by small increments, and the new values of the remaining degrees of freedom are determined at each step from the simultaneous solution of n+4 equations implicit in Equation 3.3. The incremental variation procedure is repeated until the rotating bond completes a full isomeric transition. This operation leads to the cooperative response of all the chain elements in the particular original configuration to the rotational isomerization of the given bond.

4. MODEL

4.1. Geometry of the Chain

We consider a protein chain which is represented by only its α-carbons. Therefore, a protein of n residues consists of n-1 virtual bonds, connecting the successive α-carbons. l_i designates the bond vector of magnitude l_i pointing from the $(i-1)^{th}$ α-carbon, C^{α}_{i-1}, to the i^{th}, C^{α}_{i}; θ_i is the bond angle between l_i and l_{i+1}, and ϕ_i is the torsional angle defining the rotation about bond l_i as shown in Figure 5.1. Bond lengths and bond angles are fixed. Bond lengths are taken to be 3.8 Å and a perfect tetrahedral structure is assumed for the bond angles. The dihedral or torsional angle ϕ_i of bond i may assume any value in the range $-180° \leq \phi_i \leq 180°$. The chain is allowed to translate and rotate in space, and the coordinates of chain atoms relative to a laboratory-fixed reference frame OXYZ are observed.

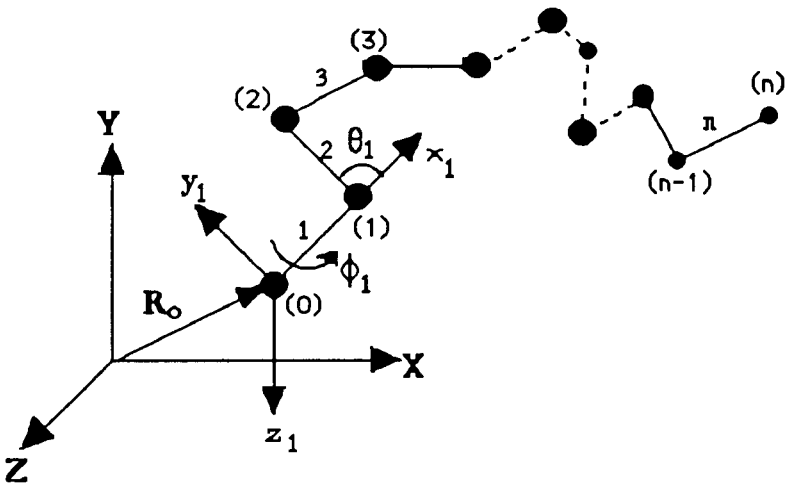

FIGURE 4.1. A random configuration of a protein chain. The C^{α} atoms are numbered from 0 to n in the laboratory-fixed frame, OXYZ, as shown in parentheses. The first bond-based coordinate system $x_1y_1z_1$ is shown.

The internal motions of the chain may be mathematically described in terms of bond-based coordinate systems. The system $Ox_1y_1z_1$ affixed to the first bond, for example, is shown in Figure 5.1. Following the notation introduced by Flory [40], the x-axis of the i^{th} bond-based coordinate system is chosen in the direction of the i^{th} bond vector and the y-axis lies in the plane of the i^{th} and the $(i-1)^{st}$ bonds, and makes an acute angle with the extension of bond (i-1). The z-axis completes a right-handed frame.

Chain atoms are indexed from 0 to n. $\mathbf{r}_i$ denotes the position vector of the i^{th} atom along the chain relative to the local frame $Ox_1y_1z_1$. The absolute position of atom i in space is indicated by the vector $\mathbf{R}_i$ connecting the origin of the frame OXYZ to the i^{th} atom. The absolute location of the chain in space is specified by the position vector $\mathbf{R}_o = (X_0\ Y_0\ Z_0)^T$ of the zeroth atom in the frame OXYZ. Here the superscript T denotes the transpose [15].

4.2. Energetics

The total energy of the protein chain is expressed by a summation over all pairwise non-bonded interactions,

$$E = \sum_{i=1}^{n-3} \sum_{j=i+3}^{n} E_{LJ}(r_{ij}) \tag{4.1}$$

where E_{LJ} is the Lennard-Jones potential between units i and j, and r_{ij} is the distance between these units and the Lennard-Jones potential is defined as,

$$E_{LJ}(r_{ij}) = -\frac{a_{ij}}{r_{ij}^{6}} + \frac{b_{ij}}{r_{ij}^{12}} \tag{4.2}$$

or

$$E_{LJ}(r_{ij}) = 4\varepsilon\left[-\left(\frac{\sigma}{r_{ij}}\right)^6 + \left(\frac{\sigma}{r_{ij}}\right)^{12}\right] \tag{4.3}$$

where the parameters a_{ij}, b_{ij}, ε and σ are related as,

$$a_{ij} = 4\varepsilon\sigma^6 \tag{4.4}$$

$$b_{ij} = 4\varepsilon\sigma^{12} \tag{4.5}$$

In the following, the subscript LJ of $E_{LJ}(r_{ij})$ will be omitted for simplicity.

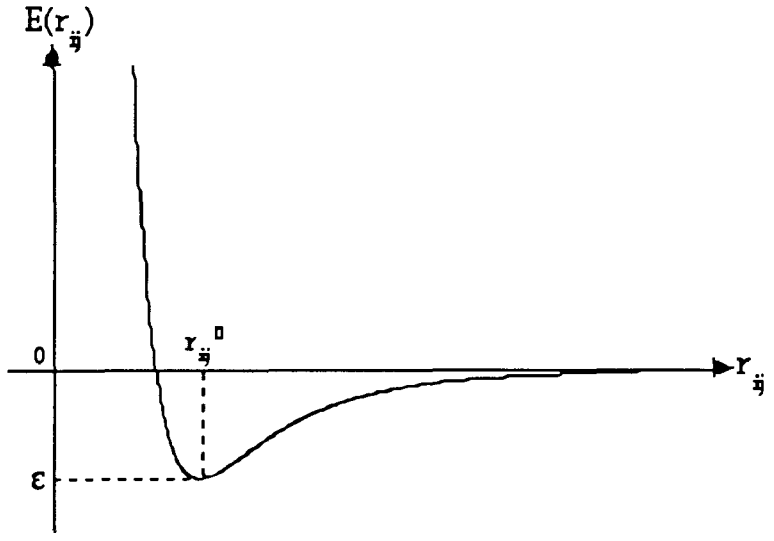

FIGURE 4.2. Lennard-Jones potential vs. r_{ij}

The physical meaning of ε and σ is better understood from Figure 4.2. r_{ij}^0 is the special r_{ij} where the energy reaches a global minimum value. σ is

related to r_{ij}^0 as $r_{ij}^0 = 2^{1/6}\sigma$. ε is the absolute value of energy at the minimum. Thus, ε defines the depth of the potential well.

In order to derive the relationship between r_{ij}^0 and σ, the derivative of the total energy with respect to distance r_{ij} has to be taken. This derivative is zero when r_{ij} is equal to r_{ij}^0 since the Lennard-Jones potential reaches a global minimum number at that point (Figure 4.2), i.e.

$$\frac{dE}{dr_{ij}} = 0 \text{ at } r_{ij} = r_{ij}^0 \tag{4.6}$$

Using Equation 4.2, the derivative of energy with respect to distance is replaced by

$$\frac{dE}{dr_{ij}} = \frac{6a_{ij}}{(r_{ij}^0)^7} - \frac{12b_{ij}}{(r_{ij}^0)^{13}} = 0 \tag{4.7}$$

Equation 4.7 is solved for the unknown r_{ij}^0,

$$r_{ij}^0 = \left(\frac{2b_{ij}}{a_{ij}}\right)^{\frac{1}{6}} \tag{4.8}$$

with further arrangements, r_{ij}^0 takes the form,

$$r_{ij}^0 = 2^{\frac{1}{6}}\sigma \tag{4.9}$$

5. THEORY

5.1. Basic Postulate

According to the basic postulate adopted in the present study, the energy change should be minimized during any structural transition of the protein chain. The configurational change is applied successively, in the form of linearized steps, subject to the condition that the energy change during each step is minimum. Thus, at any step the energy either decreases with the maximum possible amount, or increases with the minimum possible amount.

It should be noted that the dissipation term of Equation 3.2, Section 3, is not used in the application of theory to the proteins. The main problem to be solved is to find how the other residues rearrange when we apply a constraint on two or more residues, respecting the condition of $\frac{\partial \Delta E}{\partial \Delta \phi_k} = 0$. If there was no such condition, then there will be infinite number of solutions. But, here we arrange the other residues in such a way that at each step the decrease in energy is maximized. Therefore, we both fulfill the constraint exerted on specific residues, and at the same time efficiently decrease the overall non-bonded energy of the protein.

The change in the total energy when the positions of atoms change may be written as

$$\Delta E = \sum_i \sum_j \frac{\delta E}{\delta(r_{ij}^2)} \Delta(r_{ij}^2) \qquad (5.1)$$

where r_{ij}^2 is the squared distance between atoms i and j. In the following notation, $\Delta(r_{ij}^2)$ will be adopted for small incremental changes in squared distance, $\delta(r_{ij}^2)$.

The derivative of energy with respect to squared distance r_{ij}^2 will be taken in order to implement it back into the Equation 5.1. This derivative may be written as

$$\frac{dE}{d(r_{ij}^2)} = \frac{d}{d(r_{ij}^2)}\left[4\varepsilon\left(-\frac{\sigma^6}{(r_{ij}^2)^3} + \frac{\sigma^{12}}{(r_{ij}^2)^6}\right)\right] \tag{5.2}$$

which leads to

$$\frac{dE}{d(r_{ij}^2)} = 4\varepsilon\left[\left(\frac{3\sigma^6}{(r_{ij}^2)^4} - \frac{6\sigma^{12}}{(r_{ij}^2)^7}\right)\right] \tag{5.3}$$

Equation 4.9 is inserted into Equation 5.3 to obtain the equality

$$\frac{dE}{d(r_{ij}^2)} = \frac{6\varepsilon}{(r_{ij}^0)^2}\left[\left(\frac{r_{ij}^0}{r_{ij}}\right)^8 - \left(\frac{r_{ij}^0}{r_{ij}}\right)^{14}\right] \tag{5.4}$$

5.1.1 Change in Squared Distances Expressed in Terms of Generalized Coordinates

The change in squared distance is defined as

$$\Delta(r_{ij}^2) = r_{ij}^2 - (r_{ij}^0)^2 \tag{5.5}$$

where r_{ij}^0 and r_{ij} are the original and final vectors, joining the atoms i and j.

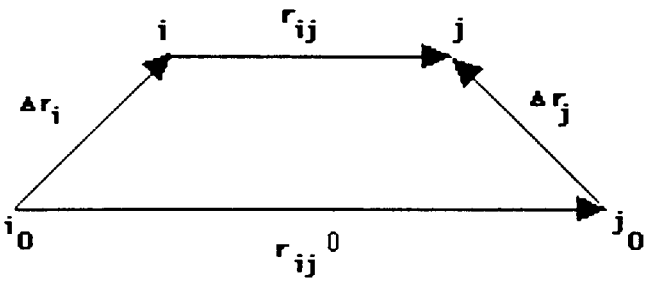

FIGURE 5.1. Representation of r_{ij}

The positions of atoms i and j after an incremental step are shown in Figure 5.1. Their positions before this step were i_0 and j_0 respectively. Δr_i and Δr_j are the displacement vectors of i and j during the incremental change. Thus, r_{ij} can be expressed as,

$$r_{ij} = r_{ij}^0 + \Delta r_j - \Delta r_i \tag{5.6}$$

where the accompanying change in distance between atoms i and j will be indicated as,

$$\Delta \mathbf{r}_{ij} = \Delta \mathbf{r}_j - \Delta \mathbf{r}_i \tag{5.7}$$

The square of $\mathbf{r}_{ij}$ may be expressed as the dot product.

$$r_{ij}^2 = (\mathbf{r}_{ij}^0 + \Delta \mathbf{r}_j - \Delta \mathbf{r}_i).(\mathbf{r}_{ij}^0 + \Delta \mathbf{r}_j - \Delta \mathbf{r}_i) \tag{5.8}$$

With some algebraic manipulations, we obtain

$$r_{ij}^2 = (\mathbf{r}_{ij}^0)^2 + (\Delta \mathbf{r}_j - \Delta \mathbf{r}_i).(\Delta \mathbf{r}_j - \Delta \mathbf{r}_i) + 2\mathbf{r}_{ij}^0.(\Delta \mathbf{r}_j - \Delta \mathbf{r}_i) \tag{5.9}$$

and finally using Equation 5.5, $\Delta(r_{ij}^2)$ is expressed as,

$$\Delta(r_{ij}^2) = (\Delta \mathbf{r}_j.\Delta \mathbf{r}_j) + (\Delta \mathbf{r}_i.\Delta \mathbf{r}_i) - 2(\Delta \mathbf{r}_j .\Delta \mathbf{r}_i) + 2\mathbf{r}_{ij}^0.\Delta \mathbf{r}_j - 2\mathbf{r}_{ij}^0.\Delta \mathbf{r}_i \tag{5.10}$$

$\Delta(r_{ij}^2)$ may be related to the changes in dihedral angles by using the linear transformation between position and dihedral angle changes. To achieve this, $\Delta \mathbf{r}_i$ ($2 \leq i \leq n$) is written in compact notation as a linear combination of $\Delta\phi_k$ for $1 \leq k \leq i-1$ [40], i.e.

$$\Delta \mathbf{r}_i = \sum_{k=1}^{i-1} \mathbf{a}_{ik} \Delta\phi_k = \sum_{k=1}^{n-1} H(k,i)\, \mathbf{a}_{ik} \Delta\phi_k \tag{5.11}$$

where H(k,i) is defined as

$$H(k,i) = \begin{cases} 1 & k<i \\ 0 & k \geq i \end{cases} \tag{5.12}$$

For each i in the range $2 \leq i \leq n$, $\mathbf{a}_{ik}$ is a 3-dimensional vector given by [41]

$$\mathbf{a}_{ik} = \left[\prod_{m=1}^{k-1} \mathbf{T}_m\right] \mathbf{A}\, \mathbf{T}_k [\mathbf{E}\ 0] \left[\prod_{m=k+1}^{i-1} \mathbf{G}_m\right] \begin{bmatrix} \mathbf{l} \\ 1 \end{bmatrix} \tag{5.13}$$

where l is the local bond vector *col*(l 0 0) for bonds of length l. Here *col* denotes the column. E is the identity matrix of order 3. The transformation matrix $\mathbf{T}_m$ for expressing vectorial or tensorial quantities of the bond-based frame $(0xyz)_{m+1}$ in frame $(0xyz)_m$ reads

$$\mathbf{T}_m = \begin{bmatrix} \cos\theta_m & \sin\theta_m & 0 \\ \sin\theta_m\cos\phi_m & -\cos\theta_m\cos\phi_m & \sin\phi_m \\ \sin\theta_m\sin\phi_m & -\cos\theta_m\sin\phi_m & -\cos\phi_m \end{bmatrix} \tag{5.14}$$

where θ_m is the supplement of the tetrahedral bond angle at the m^{th} backbone atom. A is a matrix of order 3, defined as

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \tag{5.15}$$

and $\mathbf{G_m}$ is the conventional generator matrix for calculating $\mathbf{r_i}$ [41]

$$\mathbf{G_m} = \begin{bmatrix} \mathbf{T_m} & \mathbf{I} \\ 0 & 1 \end{bmatrix} \tag{5.16}$$

By changing the indices of Equation 5.11 we may obtain $\Delta\mathbf{r_j}$, as

$$\Delta\mathbf{r_j} = \sum_{l=1}^{j-1} \mathbf{a_{jl}}\, \Delta\phi_l = H(l,j)\, \mathbf{a_{jl}}\, \Delta\phi_l \tag{5.17}$$

where the last equality is written using Einstein's summation convention over repeated indices (here l).

Upon multiplication by $\Delta\mathbf{r_i}$, Equation 5.17 yields

$$\Delta\mathbf{r_i}.\Delta\mathbf{r_j} = \sum_{k=1}^{n-1} \sum_{l=1}^{n-1} H(k,i)\, H(l,j)\, \Delta\phi_k\, \Delta\phi_l\, \mathbf{a_{ik}}.\mathbf{a_{jl}} \tag{5.18}$$

Equation 5.18 may be rewritten again using Einstein's convention as

$$\Delta \mathbf{r}_i . \Delta \mathbf{r}_j = \mathbf{B}_{ijkl} \, \Delta\phi_k \, \Delta\phi_l \tag{5.19}$$

where $\mathbf{B}_{ijkl}$ is defined as

$$\mathbf{B}_{ijkl} = H(k,i) \, H(l,j) \, \mathbf{a}_{ik} . \mathbf{a}_{jl} \tag{5.20}$$

The term $\mathbf{r}_{ij}^0 . \Delta \mathbf{r}_i$ in Equation 5.10 may be written in the explicit form

$$\mathbf{r}_{ij}^0 . \Delta \mathbf{r}_i = \mathbf{r}_{ij}^0 . \sum_{k=1}^{i-1} \mathbf{a}_{ik} \, \Delta\phi_k \tag{5.21}$$

Equation 5.21 may be rewritten as

$$\mathbf{r}_{ij}^0 . \Delta \mathbf{r}_i = \mathbf{C}_{ijk} \, \Delta\phi_k \tag{5.22}$$

where $\mathbf{C}_{ijk}$ is defined as

$$\mathbf{C}_{ijk} = H(k,i) \, r_{ij}^0 \, . \, \mathbf{a}_{ik} \tag{5.23}$$

Inserting Equations 5.18 and 5.21 into the Equation 5.10 will result in

$$\Delta\left(r_{ij}^2\right) = \mathbf{B}_{jjkl}+\mathbf{B}_{iikl}-2\mathbf{B}_{ijkl}\,\Delta\phi_k\,\Delta\phi_l - 2\,\mathbf{C}_{ijk}\,\Delta\phi_k + 2\,\mathbf{C}_{ijl}\,\Delta\phi_l \tag{5.24}$$

5.1.2 Final Expression for the Differential Change in Energy

The change in energy is obtained with the implementation of Equations 5.4 and 5.24 in Equation 5.1, which yields in explicit notation,

$$\Delta E = \sum_{i=1}^{n}\sum_{j=1}^{n}\left[\sum_{k=1}^{n-1}\sum_{l=1}^{n-1}\left(\mathbf{B}_{jjkl}+\mathbf{B}_{iikl}-2\mathbf{B}_{ijkl}\right)\Delta\phi_k\,\Delta\phi_l\,\frac{\delta E}{\delta\left(r_{ij}^2\right)}\right]+\kappa \tag{5.25}$$

where κ is

$$\kappa = \sum_{i=1}^{n}\sum_{j=1}^{n}\left[\sum_{k=1}^{n-1}\mathbf{r}_{ij}^{0}.\mathrm{H}(k,i)\,\mathbf{a}_{ik}\,\Delta\phi_k - \sum_{k=1}^{n-1}\mathbf{r}_{ij}^{0}.\mathrm{H}(l,j)\,\mathbf{a}_{jl}\,\Delta\phi_l\right]\frac{\delta E}{\delta\left(r_{ij}^2\right)} \tag{5.26}$$

We notice that the term κ will vanish since indices i and j are interchangeable. Thus ΔE reduces to

$$\Delta E = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n-1}\sum_{l=1}^{n-1}\left(\mathbf{B}_{jjkl}+\mathbf{B}_{iikl}-2\mathbf{B}_{ijkl}\right)\Delta\phi_k\,\Delta\phi_l\,\frac{\delta E}{\delta\left(r_{ij}^2\right)} \tag{5.27}$$

or

$$\Delta E = \sum_{k=1}^{n-1} \sum_{l=1}^{n-1} \left[\sum_{i=1}^{n} \sum_{j=1}^{n} (B_{jjkl} + B_{iikl} - 2B_{ijkl}) \frac{\delta E}{\delta(r_{ij}^2)} \right] \Delta\phi_k \, \Delta\phi_l \tag{5.28}$$

Using Equation 5.20,

$$B^*_{kl} = \sum_{i=1}^{n} \sum_{j=1}^{n} (B_{jjkl} + B_{iikl} - 2B_{ijkl}) \frac{\delta E}{\delta(r_{ij}^2)} \tag{5.29}$$

ΔE takes the form of

$$\Delta E = \sum_{k} \sum_{l} B^*_{kl} \, \Delta\phi_k \, \Delta\phi_l \tag{5.30}$$

5.2. Introduction of an External Constraint

Initially, let us introduce only one external constraint: the displacement of the end-to-end vector is constrained at each step to be Δr_{1n}^{ext}. Then the new condition introduced is defined as $\Delta r_{1n} = \Delta r_{1n}^{ext}$. The constraint of the fixed displacement of chain ends is accounted for the introduction of three Lagrange multipliers λ_x ,λ_y and λ_z. Equation 5.30 will then read as

$$\Delta E = \sum_{k} \sum_{l} \mathbf{B}^{*}_{kl} \Delta\phi_k \Delta\phi_l + \lambda.[\Delta \mathbf{r}_{1n} - \Delta \mathbf{r}_{1n}^{ext}] \qquad (5.31)$$

Here, λ is the vector of the three Lagrange multipliers which are conveniently written as $\lambda = [\lambda_x \lambda_y \lambda_z]^T$. We recall that Δr_{1n} may be written as

$$\Delta r_{1n} = \Delta r_n = \sum_{k=1}^{n-1} \mathbf{a}_{nk} \Delta\phi_k \qquad (5.32)$$

5.3. Minimization of Energy Change in the Presence of External Constraints

The minimization of the energy change, ΔE, with respect to differential changes in dihedral angles and with respect to λ, is performed by taking the partial derivatives of ΔE with respect to (i) $\Delta\phi_m$, where $2 \leq m \leq n-1$, and (ii) λ.

We first consider the derivatives with respect to changes in dihedral angles, which using Equation 5.31 and 5.32 may be written as

$$\frac{\delta \Delta E}{\delta \Delta \phi_m} = \sum_{l=1}^{n-1} \mathbf{B}^{*}_{ml} \Delta\phi_l + \sum_{k=1}^{n-1} \mathbf{B}^{*}_{km} \Delta\phi_k + \lambda \,.\, \mathbf{F}_m(n) = 0 \qquad (5.33)$$

where $\mathbf{F}_m(n)$ is defined as

$$F_m(n) = H(m,n)\ a_{nm} \tag{5.34}$$

The derivative with respect to λ reads

$$\frac{\delta \Delta E}{\delta \lambda} = \sum_{k=1}^{n-1} F_k(n)\, \Delta\phi_k - \Delta r_{1n}^{ext} = 0 \tag{5.35}$$

The simultaneous solution of the above set of (n-2) equations of Equation 5.33, for m in the range $2 \le m \le n-1$, together with the three identities of Equation 5.35 yields the (n+1) unknowns, $\Delta\phi_m$ with $2 \le m \le n-1$, λ_x, λ_y, λ_z. The set of (n+1) equations are conveniently organized in matrix notation as

$$\begin{bmatrix} \mathbf{D}_{(n-1)x(n-1)} & \mathbf{F}(1,n)_{(n-1)x3} \\ \mathbf{F}^T(1,n)_{3x(n-1)} & 0_{3x3} \end{bmatrix} \begin{bmatrix} \Delta\phi \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ \Delta r_{1n}^{ext} \end{bmatrix} \tag{5.36}$$

where the subscripts indicate the sizes of the matrices. Here $\Delta\phi$ represents the array of incremental changes in dihedral angles

$$\Delta\phi = \mathrm{col}\left(\Delta\phi_1\ \Delta\phi_2 \ldots\ldots \Delta\phi_{n-1}\right) \tag{5.37}$$

and the elements D_{ij} of D are given by

$$\mathbf{D}_{ij} = \mathbf{B}^{*}_{ij} + \mathbf{B}^{*}_{ji} \tag{5.38}$$

and $\mathbf{F}^{T}(1,n)$ is the 3 x (n-1) matrix defined as

$$\mathbf{F}^{T}(1,n) = [\ \mathbf{F}_1(1,n)\ \ \mathbf{F}_2(1,n)\ \ \mathbf{F}_3(1,n)\ \ \ldots\ldots\ \ \mathbf{F}_{n-1}(1,n)\ ]_{3x(n-1)} \tag{5.39}$$

Alternatively, we may write its transpose $\mathbf{F}(1,n)$ as

$$\mathbf{F}(1,n) = \begin{bmatrix} F_{1x}(1,n) & F_{1y}(1,n) & F_{1z}(1,n) \\ F_{2x}(1,n) & F_{2y}(1,n) & F_{2z}(1,n) \\ \vdots & \vdots & \vdots \\ F_{n-1x}(1,n) & F_{n-1y}(1,n) & F_{n-1z}(1,n) \end{bmatrix}_{(n-1)x(3)} \tag{5.40}$$

5.4. Implementation of Multiple Constraints

Additional constraints, other than the restriction of the end-to-end distance, may also be imposed in Equation 5.36. As an example of a second constraint, we can give,

$$\Delta \mathbf{r}_{ps} = \Delta \mathbf{r}_{ps}^{ext} \tag{5.41}$$

Then, we define F(p,s) as

$$\mathbf{F}(p,s) = H_p(k,s)\, \mathbf{a}_{sk} \tag{5.42}$$

where the subscript p is added in order to keep track of the starting point of the serial multiplication. Then, Δr_{ps} may be written as

$$\Delta \mathbf{r}_{ps} = \sum_{k=1}^{s-1} H_p(k,s)\, \mathbf{a}_{sk}\, \Delta\phi_k \tag{5.43}$$

where $H_p(k,s)$ is given by

$$H_p(k,s) = \left\{ \begin{array}{c} 0 \text{ if } k<p \text{ or } k\geq s \\ 1 \quad \text{otherwise} \end{array} \right\} \tag{5.44}$$

The counterpart of Equation 5.36 in the presence of two constraints becomes

$$\begin{bmatrix} \mathbf{D}_{(n-1)x(n-1)} & \mathbf{F}(1,n) & \mathbf{F}(p,s) \\ \mathbf{F}^T(1,n) & 0_{3x3} & 0_{3x3} \\ \mathbf{F}^T(p,s) & 0_{3x3} & 0_{3x3} \end{bmatrix} \begin{bmatrix} \Delta\phi_1 \\ \Delta\phi_2 \\ \Delta\phi_{n-1} \\ \lambda_1(1,n) \\ \lambda_2(1,n) \\ \lambda_3(1,n) \\ \lambda_1(p,s) \\ \lambda_2(p,s) \\ \lambda_3(p,s) \end{bmatrix} = \begin{bmatrix} 0 \\ \Delta\mathbf{r}_{1n}^{ext} \\ \Delta\mathbf{r}_{ps}^{ext} \end{bmatrix} \tag{5.45}$$

The n+4 unknowns, $\Delta\phi_m$ with $2 \le m \le n-1$, λ_1, λ_2, λ_3 for the first constraint, Δr_{1n}^{ext}, and λ_1, λ_2, λ_3 for the second constraint, Δr_{ps}^{ext}, are obtained by inverting the square matrix on the left-hand side and inserting the result in

$$\begin{bmatrix} \Delta\phi_1 \\ \Delta\phi_2 \\ \Delta\phi_{n-1} \\ \lambda_1(1,n) \\ \lambda_2(1,n) \\ \lambda_3(1,n) \\ \lambda_1(p,s) \\ \lambda_2(p,s) \\ \lambda_3(p,s) \end{bmatrix} = \begin{bmatrix} \mathbf{D}_{(n-1)x(n-1)} & \mathbf{F}(1,n) & \mathbf{F}(p,s) \\ \mathbf{F}^T(1,n) & 0_{3x3} & 0_{3x3} \\ \mathbf{F}^T(p,s) & 0_{3x3} & 0_{3x3} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \Delta\mathbf{r}_n^{ext} \\ \Delta\mathbf{r}_{ps}^{ext} \end{bmatrix} \tag{5.46}$$

It should be noted that Equation 5.46 presents the unique solution for a particular internal configuration subject to two constraints. If more constraints are imposed, the number of unknowns increases, along with the size of the matrix to be inverted.

6. SIMULATION METHOD

6.1. General Simulation Procedure

Tetrahedrally bonded protein chains of 30 bonds, each of length 3.8 Å, are taken into consideration in these simulations. The set of distance constraints that are to be fulfilled by the generated structures is initially defined.

The original conformations of the chains are generated by assigning all torsional angles as *trans* state. Literally, being in all *trans* state means that all the torsional angles of the chain are equal to zero. But, in order to give a certain amount of flexibility, these angles are taken arbitrarily in the range of [-20°, 20°].

Then, the distances between all atoms are calculated and the distances between the constrained atoms are checked. If all the constraints are satisfied initially, then the generated conformation is the one that is searched for. If not, the next step is to approach the constrained atoms with a constant step size of 0.04 Å. It should be recalled that this displacement should be small enough for representing the change in conformation by discrete steps.

The energy of the instantaneous conformation of the system is calculated by using the definition of energy in Equation 4.3 which is described in Section 4, ε and σ are taken as 6.0 J and 1.8 Å. Pairwise and tertiary interactions are neglected.

The next step is the minimization of the energy of the instantaneous conformation with respect to torsional angles and Lagrange multipliers. If there is only one constraint to be satisfied, Equation 5.36 is solved for the unknowns, $\Delta\phi_m$ with $2 \le m \le n-1$, λ_1, λ_2, and λ_3. If there are two constraints, n+4 unknowns, $\Delta\phi_m$ with $2 \le m \le n-1$, λ_1, λ_2, λ_3 for the first constraint, Δr_{1n}^{ext}, and λ_1, λ_2, λ_3 for the second constraint, Δr_{ps}^{ext}, are obtained uniquely from the solution of Equation 5.45. By this approach, new torsional angles and

Lagrange multipliers are calculated at each step. The energy of the system at each step with the new torsional angles is also calculated. If the energy of the system is greater than that in the preceding step, then the step size is reduced. If the energy is 0.5 J greater than that of the previous, then the step size is divided by 2. If ΔE is between 0.5 J and 1.0 J , then the step size is divided by 4, between 1.0 J and 1.5 J by 8.

If the energy is increasing by more than 1.5 J, the step size is reduced by a factor of 10^{-6}. In this case, the effect of this move on torsional angles is almost set to zero.

If there is more than one constraint and the energy is increasing more than 1.5 J, then this movement is canceled and the conformation of the system is relaxed by imposing a series of successive small increments on the torsional angles. Approximately, 30 moves are made at a time and at each move the torsional angles are changed by 10°. The acceptance of a move is based on the Metropolis criterion [42]. Accordingly, a conformation is accepted if it is of lower energy than the current one; otherwise the new conformation is accepted conditionally, using $\exp[\, a\, (E\{\phi\} - E\{\phi\}^0) / RT] \leq p$, with the acceptance probability, number p. Here, a/RT is taken as 1.

This procedure is continued until all the distance criteria are fulfilled.

For the purpose of statistical analysis, 500 independent chains are systematically generated. Their motions in the presence of constraints are examined. The runs which fulfill all the distance constraints and also have a negative (attractive) energy are accepted as successful results.

Further information on the computer program is presented in Appendix A. Also, the program code and the input files are given in Appendix B.

6 .2. Systems Studied

The following cases with different constraints were examined:

Case 1. Application of a single constraint to chain ends. Fully extended (all *trans*) protein chains of 30 residues with a kink in the middle are taken into consideration. A kink in the middle is ensured by a *gauche* bond at the 15^{th} residue and the chain ends are constrained to coincide with each other.

Case 2. Study of the effect of a single constraint to the chain ends. All *trans* chain with an applied distance constraint of 5.5 Å between chain ends is considered.

Case 3. Study of a chain with a constraint in the middle of the chain. A constraint of 5.5 Å is applied between the 8^{th} and 22^{nd} residues. The chains are initially in all *trans* conformation.

Case 4-. Study of an all *trans* protein structure which is subjected to a constraint of 5.5 Å between the 8^{th} and 15^{th} residues.

Case 5. Study of a chain with 2 constraints. The generated conformations are expected to fulfill two constraints. The first is applied to chain ends, and the second to residues 8 and 15.

Case 6. Study with a 30 residue protein chain with 3 constraints. In this case, one constraint of 6.5 Å is applied to the ends of the chain, a second one of 7.5 Å to the 8^{th} and 22^{nd} residues, and a third one of 11.5 Å between the 8^{th} and 15^{th} residues.

Case 7. A real helical protein *eschericlua coli* repressor of primer, with PDB code of 1rop. The protein has 56 residues. Results of Case 6 are used as input for the central segment of 30 residues and one more constraint is applied to the ends of the chain with a distance restraint of 5.5 Å.

7. RESULTS and DISCUSSIONS

7.1. General Evaluation of Generated Conformations

In the present study, the aim was to generate a set of protein structures using different number of distance restraints, and to have a better understanding of the effect of distance restraints on protein folding with the extensive analysis of the obtained conformations. As a result more than 500 independent runs were performed from each case. The number of conformations that fulfill the distance restraints and have an attractive overall energy are listed below in Table 7.1.

TABLE 7.1. Number of accepted conformations for cases in Section 6.2

Case number	Number of accepted conformations
1	100
2	34
3	35
4	31
5	21
6	8
7	11

In order to quantify the similarity between the structures obtained with the same type of distance restraints, the following root-mean-square deviation (RMS) value was used:

$$RMS = \left[\frac{\sum_{i=2}^{N} (\mathbf{r}_i - \mathbf{r}_i^0)^2}{N-1} \right]^{\frac{1}{2}} \qquad (7.1)$$

where r_i is the position of the atom i in a given accepted structure, and r_i^0 is its counterpart in another accepted structure with the same constraint, the two structures being optimally superposed on each other. These RMS deviations are found by using the package program UCSF MIDASPLUS [31]. The mean RMS deviations attained for each case are presented in Table 7.2.

TABLE 7.2. Mean RMS deviation of each case studied

Case Number	Mean RMS deviation (Å)
2	6.03
3	6.18
4	6.25
5	5.77
6	5.69 (5.86)
7	9.37 (8.12)

The RMS deviation for Case 1 is not reported in the Table 7.2, because Case 1 is an extreme case of bringing the end of chain very close to each other (0 Å). The terms in parentheses indicate the RMS deviation of successful conformations from the real protein.

7.2. Simulation Results and Discussion

7.2.1. Examination of Local Structure Formation

In Case 1, the time evolution of the accepted conformations were examined. The behavior of a sample conformation can be clearly seen from Figure 7.1. The chain starts to fold at the time step of 450 with the help of the end-to-end constraint. At 900 steps, it takes a helical shape on a local scale, and at 1500 steps, the kinked protein chain has almost completed its local

structures, which are helices in this case. A helical hairpin, which is a common structural motif in folded proteins, is reached.

Figure 7.1 clarifies that the long-range constraint fulfillment is accompanied by a regular structure formation such as helices on a short-range (local) scale.

This observation is in agreement with the framework model of the protein folding [43]. This view of protein folding assumes that the "local" interactions among the near neighbors along the sequence, i.e. the interactions that form helices and turns, are the main determinants of protein structure. The model has predicted three stages of protein folding: the formation of secondary structures around their native positions, the merging of these regions into a compact state with the native-like folding pattern, and the transition of this intermediate state into the rigid native state. The important aspect of this model is that the secondary structures of the intermediate fluctuates around the native positions of α-helices and β-strands. Such intermediates have been found in kinetic and equilibrium experiments [44].

In our study, by giving an initial constraint to the system, in a way we force it to fold, and to assume structural organizations on a local scale, similarly to intermediates in the folding mechanism.

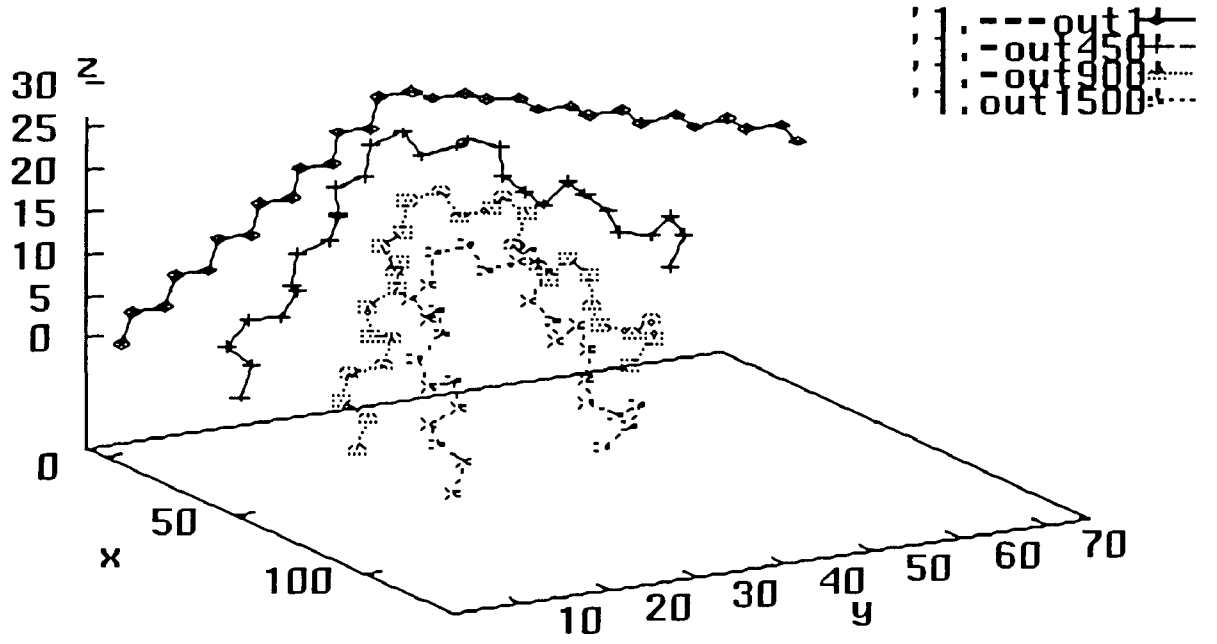

FIGURE 7.1. Time behavior of initially all *trans* protein chain with a kink in the middle

7.2.2. Investigation of Energy and Local Structure Formation Relationship

In order to examine the role that energy plays in local structure formation, a more realistic distance restraint is considered in Case 2. The distance restraint is again applied to the ends of the chain but this time it is taken as 5.5 Å. The resulting conformations are analyzed in order to determine the relationship between energy and local structure formation. Conformations with low energies are observed to have local structure formations such as helices, whereas there is no local structure formation in the conformations that have higher energies. This feature is illustrated in Figure 7.2. Therefore, a systematic study shows that those conformations exhibiting regularities (helices) on a local scale are energetically more favorable.

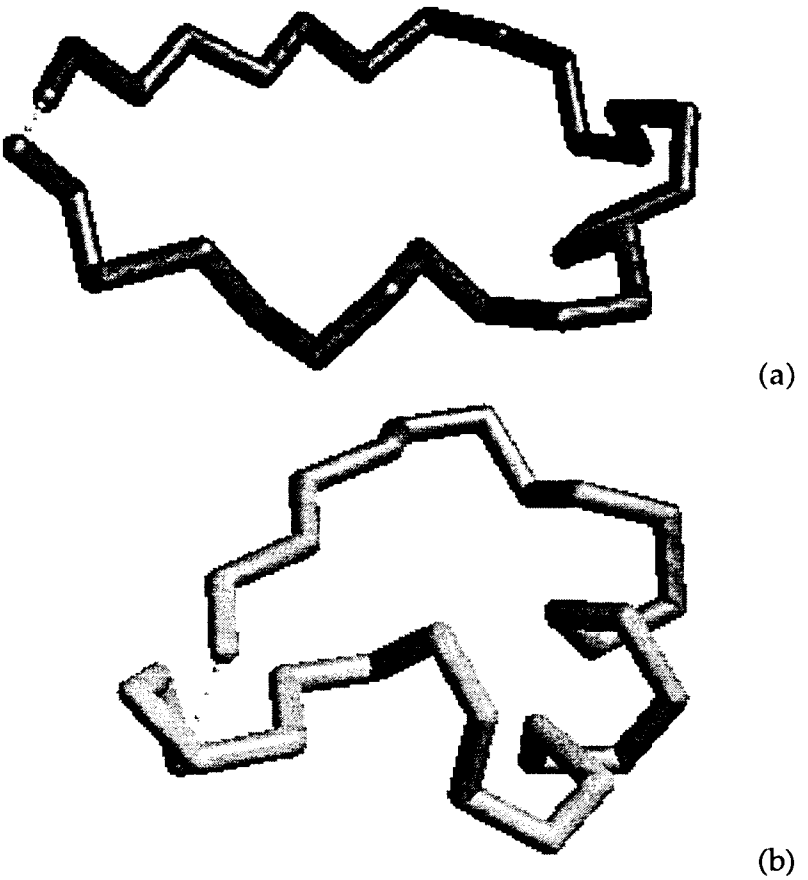

FIGURE 7.2. Two example conformations found from the solution of CK theory with single constraint. The conformation in part (a) has relatively high energy (-0.02 kJ/residue), and that in part (b) has a low energy (-1.58 kJ/residue) and more structure on a local scale.

The time evolution of conformations with lowest and highest energy during the time steps is also investigated. Although both of the conformations satisfy the end-to-end constraint, one ends up with a high energy while the other reaches a low energy configuration. The energy changes of both conformations are presented in Figure 7.3.

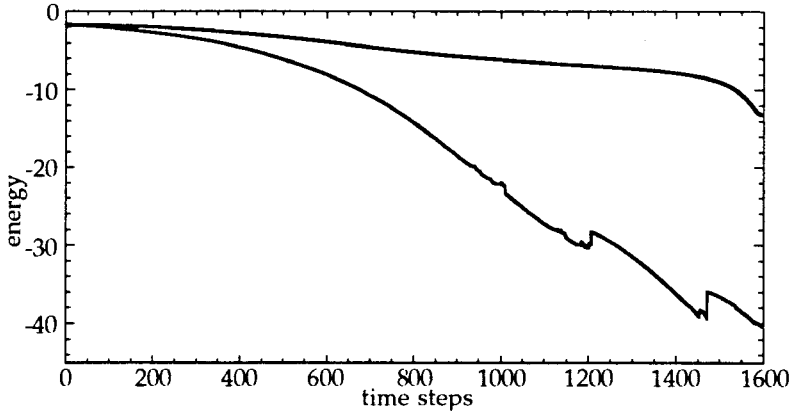

FIGURE 7.3. Time evolution of energy in the lowest and highest energy conformations obtained with $\Delta r_{1,n}^{ext} = 5.5$ Å.

7.2.3. The Importance of Constraint Location

After investigating the effect of a constraint applied on the chain ends, one might examine the sensitivity of the results on the focus of imposition of the constraint. To this aim a 5.5 Å constraint is now applied to the middle of the chain, between the 8^{th} and 22^{nd} residues, and the resulting conformations are compared with those obtained by applying the end-to-end constraint in order to find an answer to the question of what kind of an effect does the place of constraint has.

The average RMS deviation found when a constraint is applied to the chain ends was 6.03 Å, but an increases to 6.18 Å is observed when the constraint is exerted on the middle of the chain. The probability distribution curves for RMS deviations between pairs of accepted results were investigated,

and it was seen that the probability curve shifts to the right when the constraint is applied to the middle of the chain as opposed to chain ends (Figure 7.4). Also, RMS deviation is increased apparently when the constraint is applied to the middle of the chain.

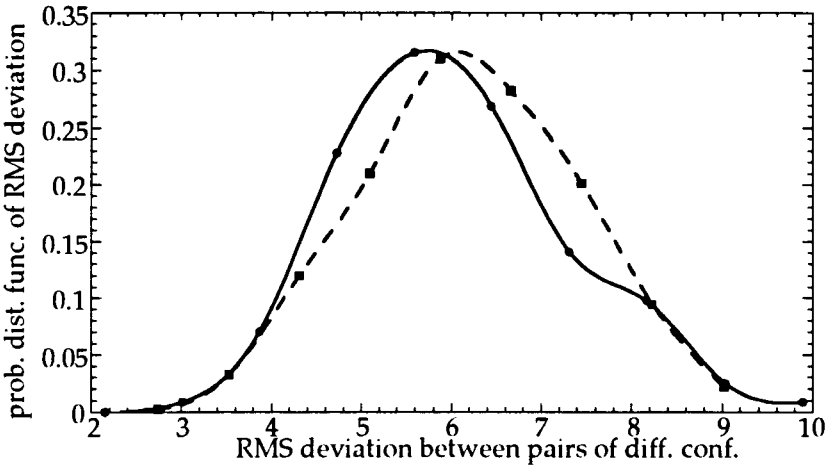

FIGURE 7.4. Probability distribution function of RMS deviations between pairs of conformations satisfying the end-to-end constraint (solid curve) and the constraint applied to the middle of the chain (dashed curve).

As a result, we can say that the application of a single constraint to chain ends is more effective in restricting the conformational space of the chain of 30 residues than its application to the two internal points.

Also, the RMS deviations were evaluated for the constrained parts of the chain only. When the 15 residues that lie between the 8^{th} and 22^{nd} residues are considered, the average RMS deviation for the constrained region is substantially lowered that the mean RMS for the constrained region is only 3.6 Å while it is 6.18 Å for 30 the complete chain of 30 residues.

The probability distributions for the respective RMS deviations are shown in Figure 7.5.

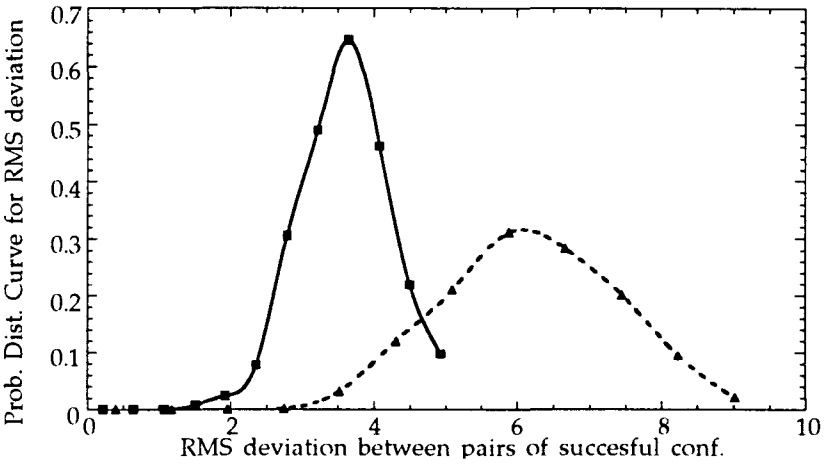

FIGURE 7.5. Probability distributions for RMS deviations between pairs of conformations for the portion of the chain between constrained residues 8 and 22 (solid curve), and for the whole chain (dashed curve).

7.2.4 The Effect of Application of a Constraint to Close Residues Along the Chain Connectivity

In order to further investigate the effects of constraint position on the accuracy and reproducibility of final configuration, a constraint is applied between close residues along the chain connectivity, between the 8^{th} and 15^{th} residues.

The final configurations that successfully satisfy the constraint are superimposed and the same structural pattern is observed in all of them. In all chains, the shortest tail is being readjusted while the longer tail remains almost unaffected. The final superimposed views of three resulting conformations can be seen from Figure 7.6.

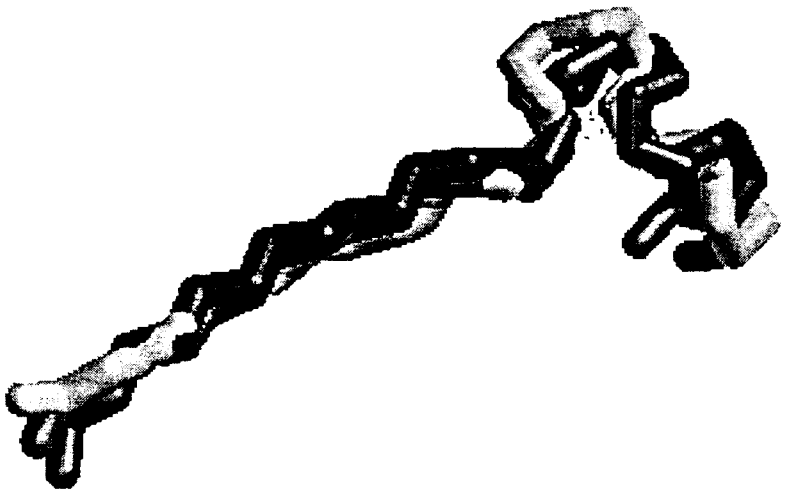

FIGURE 7.6. Superposition of three conformations satisfying the constraint applied between the 8th and 15th residues.

7.2.5. The Effect of Increasing the Number of the Constraints- Investigation of the Case with Two Constraints.

After examination of the effect of one constraint, we investigated the result of increasing the number of constraints from one to two. The two constraints to be satisfied were applied on the chain ends (5.5 Å) and on the 8th and 22nd residues (5.5 Å).

When accepted conformations were superimposed, the structures looked quite alike, in accordance with the fact that using two constraints restricts the conformational space more efficiently than using one constraint (Figure 7.7).

The mean RMS deviation attained in this case is lower than as expected. The conformations which are subject to two constraints have a mean RMS value of 5.77 Å while those with one constraint have mean RMS value over 6 Å (Table 7.2). Also, the probability distribution curve of RMS deviations is shifted to the left with respect to the curve found with one constraint since

using two constraints restricts the conformational space more than using one constraint. (Figure 7.8)

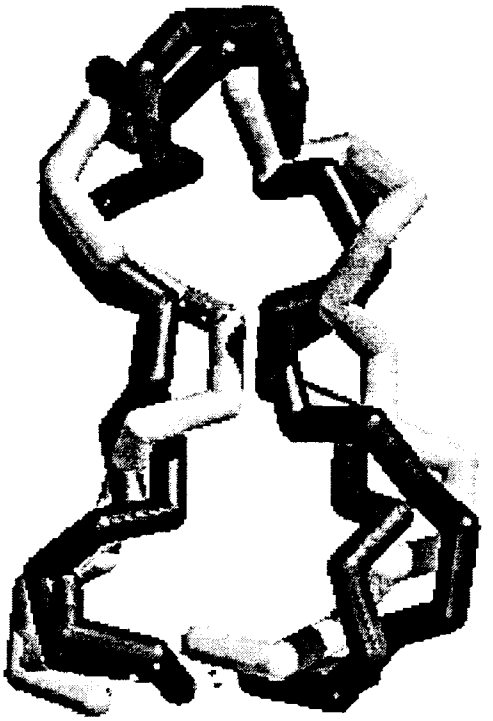

FIGURE 7.7. Superposition of three conformations satisfying two constraints, one at chain ends and the other in the central portion of the structure.

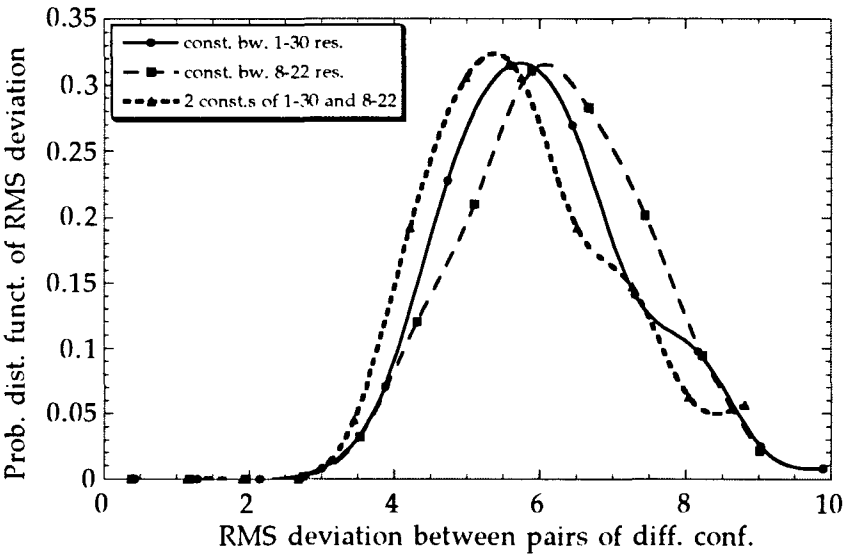

FIGURE 7.8. Probability distribution curve of RMS deviations for conformations subject to one constraint and two constraints.

Also, the consequences of having two constraints are energetically investigated. Having two constraints undoubtedly decreases the overall energy of the conformations. The probability distribution curve for the energy of conformations with two constraints is sharper and narrower than the curves for conformations with one constraint.

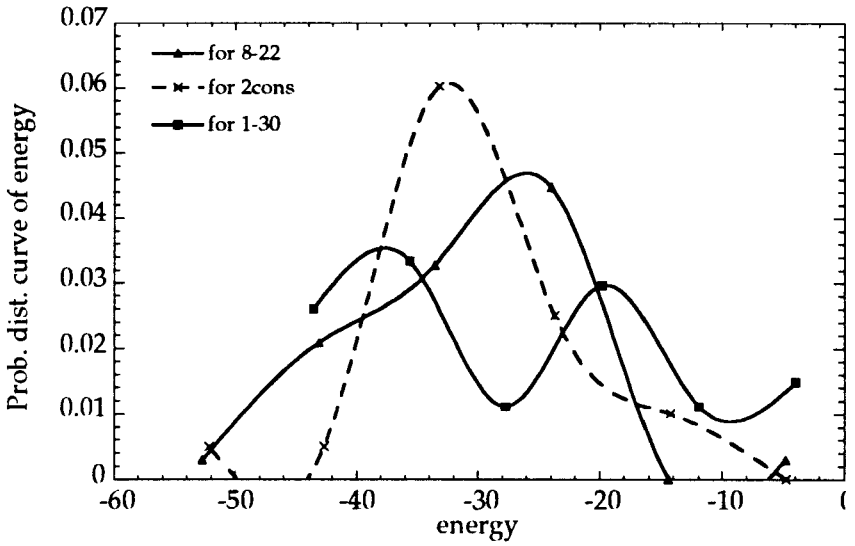

FIGURE 7.9. Distribution of energies of final conformations found with one constraint and two constraints.

7.2.6. Application of Three Distance Restraints

In the present section, the number of distance restraints is increased to three and the distances are adopted from a real protein, an *esclierichia coli* repressor of primer with a PDB code of 1rop. There are 56 α-carbons in 1rop. First, the 30 carbons between the 16^{th} to 45^{th} are considered. The distance between the 23^{rd} and 30^{th} α-carbons is 10.49 Å and the distance restraint taken between the 8^{th} and 15^{th} α-carbons is 11.50 Å. The distance restraint for the middle of the chain is taken as 7.5 Å where the distance for the real protein is 8.92 Å. As a last restraint, the end-to-end distance is taken as 6.5 Å,

approximating the distance between the α– carbons 16 and 44, which is actually 5.0 Å.

When the predicted conformations are compared with mentioned section of the native structure, the results are quite satisfactory. The mean RMS deviation from the native structure is only 5.86 Å which is an impressive result for such a simple and coarse-grained model. Also, the restraint imposed on the middle of the chain was different from the corresponding distance in the actual structure.

The probability distribution of RMS deviations obtained in this case is displayed in Figure 7.10. Since there is a sharp, narrow peak in the probability distribution curve, this implies that the predicted conformations are quite alike and in good agreement with the structure. Also, Figure 7.11 illustrates how well the imposition of three constraints can lead to conformation matching the native structure.

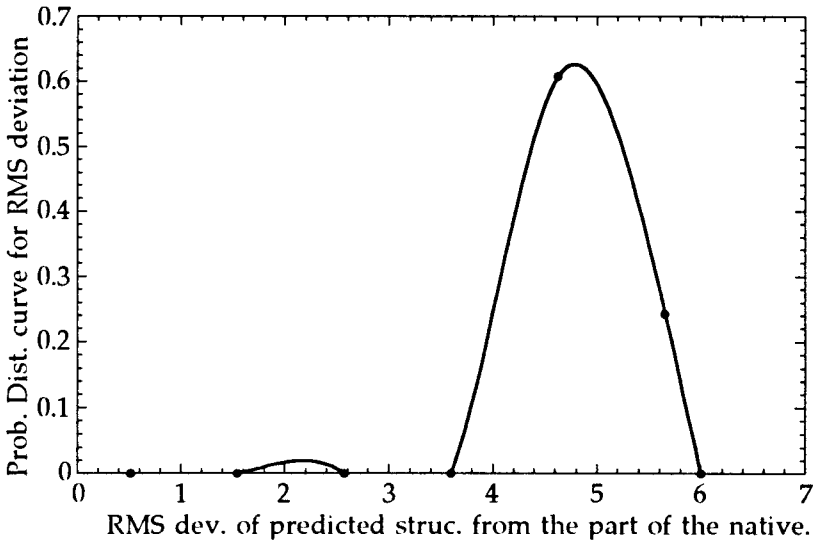

FIGURE 7.10. Probability distribution of RMS deviations between predicted structure and the 30-residue segment of the native structure.

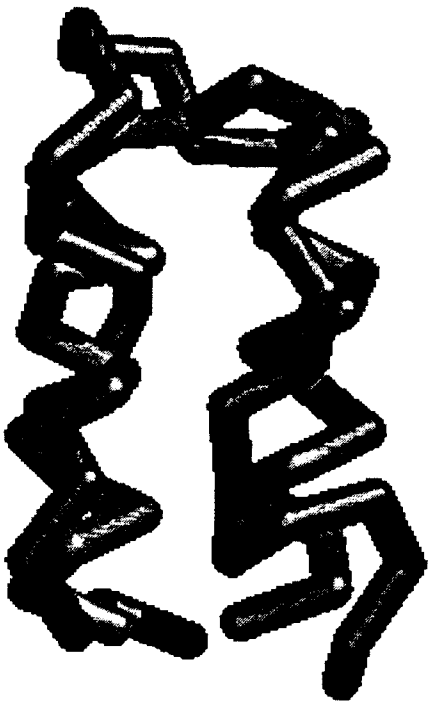

FIGURE 7.11. The superimposed structure of predicted and 30-residue segment of 1rop. The light-grey structure is the native structure and the darker one is the predicted tertiary structure. The 3-D structure of the backbone is shown.

7.2.7. Folding of a Helical Protein

In this final part of the calculation, the dihedral angles of the conformations obtained in the above case with the three constrains case is used as input in mimicking the folding of a helical protein, 1rop. Since there are 56 residues in the original protein, the remaining dihedral angles of the starting configuration are taken arbitrarily. This procedure does not exactly correspond to applying four distance restraints, since a hierarchical approach is adopted as adding one more constraint on the originally determined solutions. The resulting conformations tend to be slightly more expanded relative to the X-ray structure, Figure 7.12 displays one of the resulting conformations, superimposed on the native structure. This figure illustrates how well the native structure is predicted with such a simple model.

When the conformations found in the present case were investigated energetically, it was seen that the mean energy of the conformations with three distance restraints is -52.2 kJ while that with four constraints decreases to -56.1 kJ. When the probability distribution of energies is taken into consideration, the difference is more striking. The probability distribution curve for the energies of conformations mimicking the real protein is much narrower and shifted to the left relative to the curve for conformations with three distance restraints. See Figure 7.13.

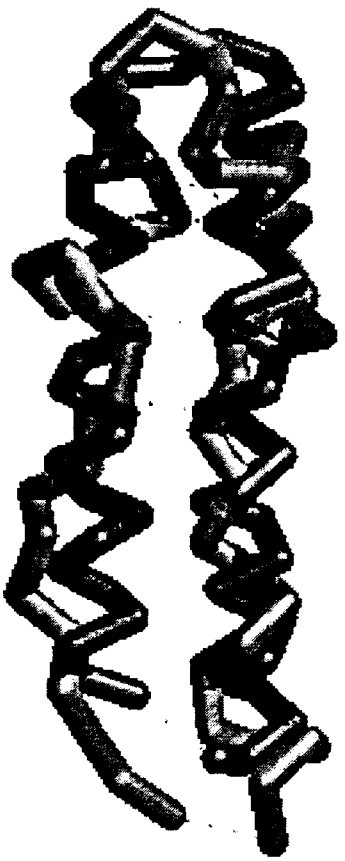

FIGURE 7.12. The superimposed structure of predicted and native configuration of 1rop. The darker structure is the native structure and the lighter one is the predicted tertiary structure.

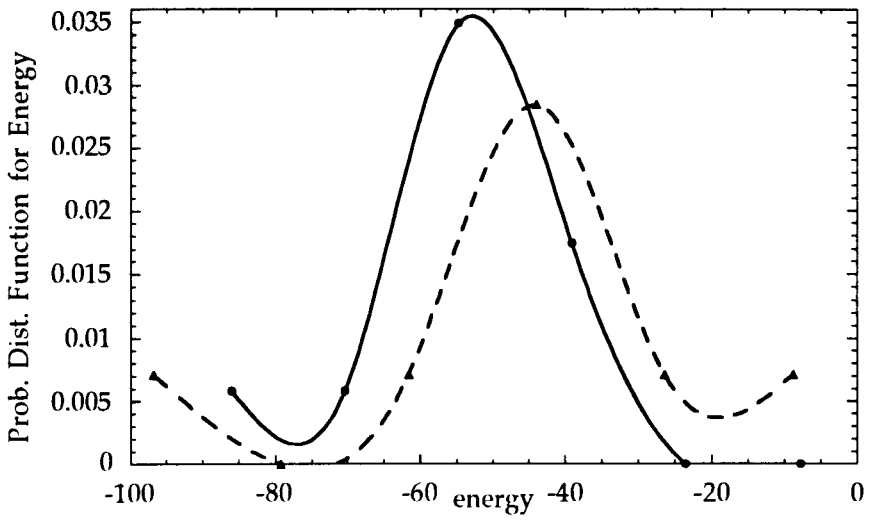

FIGURE 7.13. Probability distribution of the energies of ROP conformations predicted using three constraints (dashed) and four constraints (solid)

8. CONCLUSIONS AND RECOMMENDATIONS

8.1 Conclusions

In this study, different numbers of distance restraints were applied to variable parts of model proteins of 30 residues. Application of different types and numbers of restraints give us information about the importance of such constraints in locating native-like conformations. Present computations for predicting the native three-dimensional structure of proteins using distance restraints lead to the following conclusions:

(a) After generating the fully extended (all *trans*) chains with a kink (*gauche* bond) in the middle and constraining the chain to have end-to-end distance equal to zero, it is concluded that the long-range constraint fulfillment is accompanied by a regular structure formation on a short range scale.

(b) Those conformations forming regularities (in our case, helices) on a local scale are systematically found to be energetically more favorable than others.

(c) Application of a single constraint to chain ends is more effective in restricting the conformational space of the chain of 30 residues than application at two internal points. This shows the importance of the place of the constraint that is applied.

(d) An increase in the number of constraints results in a decrease in the number of the conformations. Also, as the number of constraints increases, the overall energy of the conformations decreases, indicating that the stability of the predicted structures is also enhanced.

(e) When the distance restraints are directly taken from structures available in the Protein Databank, RMS deviation of a tertiary structure predicted for a helical protein, 1rop, is reduced down to 5.95 Å. Therefore, it

can be concluded that a real protein structure can also be predicted with a reasonable accuracy level by using a limited number of distance restraints.

7.2. Recommendations

The results show that the number of constraints and their positions of application are very important in the prediction of protein tertiary structure. It was also shown that it is possible to closely reproduce the structure of a helical protein, 1rop, by using a limited number of distance restraints. Since these calculations were performed for only one protein, it would be desirable to increase the number of proteins used in the calculations and check the validity of the model and the method.

The starting aim of this study was to investigate the possibility of predicting the tertiary structure of small globular proteins, and in particular the metal-binding ones, given that the exact coordination geometry of the ligands about the metal are used as distance restraints. To this aim we considered Cu-binding proteins in Protein Data Bank and found that these obey well-defined structural rules near the metal-binding sites. The use of this information in the present computation algorithm could not be reached during this study. Since distinct coordination geometries of copper ligands are identified and the cooperative kinematics approach is presently shown to satisfactorily produce conformations compatible with a given number of distance restraints, the results can be used to investigate metal-binding proteins as the next step. Although copper-binding proteins were explored as a first approach, this study can also be extended to other metal-binding proteins, such as zinc-binding proteins or calcium-binding proteins.

APPENDIX A

The program developed to predict the native structure of proteins using distance restraints is named as constraint.f. It is written by using FORTRAN 77 programming language. It consists of a main part and a subroutine. The code of the program is given in Appendix B.

Description of variables used in the program is given below before going on with the explanation of the program.

e:	3x3 unit matrix
el:	components of the bond vector
t:	transformation matrix, separately identified for each bond
h:	matrix of zeros and ones
fi(i):	torsional bond angle for the i'th bond
a0:	matrix appearing in differentiation of t(i,j)
g(j,k,l):	4x4 super matrix
rj:	row matrix pre multiplying g,
cj:	column matrix post multiplying g
tt(i,j,k):	product of transformation matrices, ending at the i'th bond
gg(j,k,l,m):	4x4 super matrix starting at j'th bond, ending at k'th bond
a(i,j,k):	the aij vector appearing in the delta R, k=1,3
v(i,j):	vector appearing in the r matrix
u(m,j):	scalar appearing in the r matrix
r(i,j):	the r matrix, R1(i,j): the r matrix to be inverted returns from the subroutine in inverted form. it is equated to cc(i) before entering the subroutine.

In the main part of the program, first filenames and constants that are to be needed in the program are read from the input file. These constants are ntot, which is the number of residues found in the chain, theta, which is the constant bond angle, eq1n, which is the equilibrium distance between the constrained atoms and req, epsilon which are Lennard-Jones parameters for all pairs of atoms.

Then identity matrix, e and h matrices which is found in in Equation 5.12 of Section 5 are defined in the program. Generation of the original conformation in terms of dihedrals are made next. Then program enters to the main loop which generates the conformation satisfying the distance constrain by using known dihedrals. For this, t matrix of Equation 5.14, g matrix of Equation 5.16 and a matrix of Equation 5.15 of Section 5 are outlined. Distance vectors between each residue is also calculated. Next, calculation of a(i,j,1-3) of Equation 5.13 in Section 5 is done. Then, overall energy of the instantaneous conformation is calculated. Then, B matrix of Equation 5.29 is formed. Determination of the correction step size is made according to the energy instantaneous conformation. As a last step, the d matrix found in Equation 5.36 is formed and it is inverted by using the subroutine inverse. From this inversion, new dihedral angles are formed and distance vector between the residues to which a constraint is applied is calculated. Finally, if the aimed distance restraint is satisfied, the program terminates by writing the resulting dihedral angles and Lagrange multipliers to the output file.

Two input files are used to run the program. First one is filename.dat in which the filenames are listed. Second input file is fluctloop.dat in which constants are given.

FLUCTLOOP.DAT:

In the first line of fluctloop.dat , a dummy variable, initial number, and total number of steps are present. In the second line, number of residues, fixed bond angle, and bond lengths in x, y, z directions are written respectively. In the third line, the distance that is to be achieved at the end of the run is written. Fourth line is the Lennard-Jones parameters, epsilon and sigma.

It should be noted that initial number can be only 1 or 2. If it is 1, then initial dihedral angles are calculated in the program, but if initial is given as 2, then the dihedral angles should be given at the end of the fluctloop.dat.

FILENAME.DAT:

This is a input file that the filenames of the output files are listed. For the sample program in Appendix B, there are 14 files, and each filename should be 14 characters.

It should be remembered that output files can be made for different snapshots thorough out the simulation. For practical purposes, number of output files can be changed, but then filename.dat should be changed, too.

APPENDIX B: PROGRAM CODES

FLUCTLOOP.DAT

```
22148        1      8001
30     70.5  3.80   0.0    0.0
5.5
6.0    1.8
```

PROGRAM CONSTRAINT

```
c     This program finds the dihedral angles and Lagrange multipliers for
c     chain of 30 residues which is constrained from chain ends.

      real e(3,3),h(50,50),el(3),sfun(7000),ener(50,50)
      real a0(3,3),g(50,4,4),rj(3,4),cj(4)
      real gg(50,50,4,4),a(50,50,3),s33(3,3),s34(3,4),s3(3)
      real sum(3),v(50,3),u(50,50),r(50,50),rk(30,30)
      real bb(50,50),cc(50),q1(50,50),q2(50,50),p(3)
      real fi(50),t(50,3,3),vav(3,3),erk(7005)
      real tt(50,3,3),ri(50,3),b(50,50,50,50)
      real rij(50,50,3),rij0(50,50)
      real f(50,3),bp(50,50),sij(50,50),fi0(50)
      real ep(50,50),d(50,50),final(50)
      character*14 filename(14)

c*************************************************************************
c  description of variables
c  e: 3x3 unit matrix,
c  el: components of the bond vector
c  t: transformation matrix,separately identified for each bond,
c  h:  matrix of zeros and ones,
c  fi(i): bond angle for the i'th bond obtained by monte carlo,
c  a0: matrix appearing in differentiation of t(i,j)
c  g(j,k,l): 4x4 supermatrix
```

```
c   rj: row matrix premultiplying g,
c   cj: column matrix postmultiplying g
c   tt(i,j,k): product of transformation matrices, ending at the i'th bond
c   gg(j,k,l,m): 4x4 supermatrix starting at j'th bond,ending at k'th bond
c   a(i,j,k): the aij vector appearing in the deltaR, k=1,3
c   v(i,j): vector appearing in the r matrix
c   u(m,j): scalar appearing in the r matrix
c   r(i,j): the r matrix, R1(i,j): the r matrix to be inverted returns from the
c   subroutine in inverted form. it is equated to cc(i) before entering
c   the subroutine.
c*************************************************************************
      open(9,file='idum.dat')
      read(9,*) idum
      open(11,file='filenam.dat')
1     format(A14)
      do j=1,14
      read(11,1)filename(j)
      enddo

      open (1,file = 'fluctloop.dat')
      open (2,file = filename(1))
      open (110,file  = filename(2))
      open (111,file  = filename(3))
      open (112,file  = filename(4))
      open (113,file  = filename(5))
      open (114,file  = filename(6))
      open (115,file  = filename(7))
      open (116,file  = filename(8))
      open (117,file  = filename(9))
      open (118,file  = filename(10))
      open (119,file  = filename(11))
      open (120,file  =  filename(12))
      open (127,file  = filename(13))
      open (128,file  = filename(14))

      read (1,*) initial,ntot
      read (1,*) n,theta,el(1),el(2),el(3)
```

```
c       reading the final end-to-end vector

        read(1,*) eq1n
c       this is the equilibrium distance between the two terminal atoms
c       which is to be reached at the end of the run

        write(6,*)'equilibrium separation between chain ends=',eq1n
        write(2,*)'equilibrium separation between chain ends=',eq1n

        read (1,*) req,epsilon
c       these are the Lennard-Jones parameters for all pairs of atoms
        write(2,*)' LJ parameters=',req,epsilon

        i111=0
        rprev=1000.
        if(initial.ne.1) then
        do 90001 i=2,n-1
90001   read (1,*)kk,fi0(i)
        endif

5921    close (1)
        do 6654 i=1,n
        do 6654 j=1,n
        sij(i,j)=epsilon
        rij0(i,j)=req
6654    continue

        write (2,*) 'n= ',n,' theta= ',theta
        pi=3.141592654
        theta=theta*pi/180.

        m1=initial

* e is the identity matrix

        do 10 i=1,3
        do 1156 j=1,3
1156    e(i,j)=0.
```

```
10      e(i,i)=1.

* h is the matrix defined in Equation 5.12 of Section 5

        do 20 m=1,n
        do 21 i=1,n
        h(m,i)=0.
21      if(m.lt.i) h(m,i)=1.
20      continue

        write(2,*)'original  bond angles'
        write(2,*)' '

c       GENERATION OF THE ORIGINAL CONFORMATION IN TERMS OF
c       DIHEDRALS

        do 40 i=2,n-1
        random=ran(idum)
        fi(i)=random*0.34906585*2-(0.1744533925*2)
        write(*,*) fi(i)*180/pi
        write(2,*) fi(i)*180/pi
40      continue

        if(initial.ne.1) then
        do 90002 i=2,n-1
90002   fi(i)=fi0(i)*pi/180
        endif

        alfa=cos(theta)
        beta=sin(theta)

c*************************************************************************
c  THIS IS THE MAIN LOOP

c       GENERATION OF A CONFORMATION USING KNOWN DIHEDRALS

c       the t matrix in Equation 5.14 in Section 5
5555    fi(1)=0.
```

```
      do 750 i=1,n-1
      fii=fi(i)
      t(i,1,1)=alfa
      t(i,1,2)=beta
      t(i,1,3)=0.
      t(i,2,1)=beta*cos(fii)
      t(i,2,2)=-alfa*cos(fii)
      t(i,2,3)=sin(fii)
      t(i,3,1)=beta*sin(fii)
      t(i,3,2)=-alfa*sin(fii)
      t(i,3,3)=-cos(fii)
750   continue

c  the a matrix of Equation 5.15 in Section 5
      do 70 i=1,3
      do 70 j=1,3
70    a0(i,j)=0.
      a0(2,3)=-1.
      a0(3,2)=1.

c  the g matrix of Equation 5.16 in Section 5
      do 90 i=1,n-1
      do 80 j=1,3
      do 80 k=1,3
      g(i,4,k)=0.
      g(i,k,4)=el(k)
80    g(i,j,k)=t(i,j,k)
90    g(i,4,4)=1.

c  forming the row and column matrices for calculation of a(i,j,k)
c  of Equation 5.13 of Section 5.

      do 100 i=1,3
      do 102 j=1,4
102   rj(i,j)=0.
100   continue
      rj(1,1)=1.
      rj(2,2)=1.
```

```
      rj(3,3)=1.
      cj(1)=el(1)
      cj(2)=el(2)
      cj(3)=el(3)
      cj(4)=1.

C     determining the position vectors ri(k,1-3) of all k atoms
      do 5213 i=1,3
5213  ri(1,i)=el(i)

      do 110 k=1,3
      do 110 l=1,3
      vav(k,l)=e(k,l)+t(1,k,l)
110   tt(1,k,l)=t(1,k,l)

      do 7713 i=1,3
7713  ri(2,i)=vav(i,1)*el(1)+vav(i,2)*el(2)+vav(i,3)*el(3)

      do 120 i=2,n-1
      do 122 k=1,3
      do 122 l=1,3
      xxx=tt(i-1,k,1)*t(i,1,l)+tt(i-1,k,2)*t(i,2,l)
      tt(i,k,l)=xxx+tt(i-1,k,3)*t(i,3,l)
122   vav(k,l)=vav(k,l)+tt(i,k,l)
      do 7813 i1=1,3
7813  ri(i+1,i1)=vav(i1,1)*el(1)+vav(i1,2)*el(2)+
     : vav(i1,3)*el(3)
120   continue

      if (m1.eq.2) then
c     this is taken as the original conformation and written to fluctloop.out0
      do 41583 i=1,n
c     write(*,*) fi(i),i
41583 write(110,1055) (ri(i,j),j=1,3)
      endif

      if (m1.eq.(initial+500)) then
      do 41783 i=1,n
```

```
c       write(*,*)fi(i)
41783   write(111,1055) (ri(i,j),j=1,3)
        close (111)
        endif

        if (m1.eq.(initial+1000)) then
        do 42783 i=1,n
c       write(*,*) fi(i),i
42783   write(112,1055) (ri(i,j),j=1,3)
        close (112)
        endif


        if (m1.eq.(initial+1500)) then
        do 43783 i=1,n
c       write(*,*) fi(i),i
43783   write(113,1055) (ri(i,j),j=1,3)
        close (113)
        endif

        if (m1.eq.(initial+2000)) then
        do 44783 i=1,n
c       write(*,*) fi(i),i
44783   write(114,1055) (ri(i,j),j=1,3)
        close (114)
        endif

        if (m1.eq.(initial+2500)) then
        do 22783 i=1,n
c       write(*,*) fi(i),i
22783   write(115,1055) (ri(i,j),j=1,3)
        close (115)
        endif

        if (m1.eq.(initial+3000)) then
        do 24783 i=1,n
c       write(*,*) fi(i),i
24783   write(116,1055) (ri(i,j),j=1,3)
```

```
      close (116)
      endif

      if (m1.eq.(initial+3500)) then
      do 25783 i=1,n
c     write(*,*) fi(i),i
25783 write(117,1055) (ri(i,j),j=1,3)
      close (117)
      endif

      if (m1.eq.(initial+4000)) then
      do 25775 i=1,n
c     write(*,*)fi(i),i
25775 write(118,1055) (ri(i,j),j=1,3)
      close (118)
      endif
      if (m1.eq.(initial+4500)) then
      do 35783 i=1,n
c     write(*,*) fi(i),i
35783 write(119,1055) (ri(i,j),j=1,3)
      close (119)
      endif

      if (m1.eq.ntot) then
      write(2,*)m1,'ntot'
      do 49783 i=1,n
c     write(*,*) fi(i),i
49783 write(120,1055) (ri(i,j),j=1,3)
      close (120)
      endif
1055  format (3f7.2)

c  calculating the vector from the ith to the jth atom for all pairs
      do 6911 i=1,n-1
      do 6911 j=i+1,n
      rij(i,j,1)=ri(j,1)-ri(i,1)
      rij(i,j,2)=ri(j,2)-ri(i,2)
6911  rij(i,j,3)=ri(j,3)-ri(i,3)
```

```
      do 6912 i=1,n-1
      do 6912 j=i+1,n
      do 6912 k=1,3
6912  rij(j,i,k)=rij(i,j,k)

c  NOTE: throughout, the zeroth and first bond is excluded from calculations

317   continue

c-----------------------------------------------------------------------

c     CALCULATION OF A(I,J,1-3) of EQUATION 5.13

C     CALCULATION OF THE 4X4 MATRIX G FOR EACH BOND
      do 140 i=2,n-1
      do 140 k=1,4
      do 140 l=1,4
140   gg(i,i,k,l)=g(i,k,l)

C     CALCULATION OF SERIAL MULTIPLICATION OF G MATRICES
      do 150 i=2,n-2
      do 150 j=i+1,n-1
      do 152 k=1,4
      do 152 l=1,4
      xxx=gg(i,j-1,k,1)*g(j,1,l)+gg(i,j-1,k,2)*g(j,2,l)
152   gg(i,j,k,l)=xxx+gg(i,j-1,k,3)*g(j,3,l)+gg(i,j-1,k,4)*g(j,4,l)
150   continue

      do 200 i=3,n
      do 200 j=2,i-1
      do 210 k=1,3
      do 210 l=1,3
      zzz=tt(j-1,k,1)*a0(1,l)+tt(j-1,k,2)*a0(2,l)
210   s33(k,l)=zzz+tt(j-1,k,3)*a0(3,l)
      do 220 k=1,3
      do 220 l=1,4
      zzz=rj(k,1)*gg(j,i-1,1,l)+rj(k,2)*gg(j,i-1,2,l)
220   s34(k,l)=zzz+rj(k,3)*gg(j,i-1,3,l)+rj(k,4)*gg(j,i-1,4,l)
```

```
      do 230 k=1,3
      zzz=s34(k,1)*cj(1)+s34(k,2)*cj(2)+s34(k,3)*cj(3)
230   s3(k)=zzz+s34(k,4)*cj(4)
      do 240 k=1,3
240   s3(k)=s3(k)-el(k)
      do 250 k=1,3
250   a(i,j,k)=s33(k,1)*s3(1)+s33(k,2)*s3(2)+s33(k,3)*s3(3)
200   continue

      do 4443 i=1,2
      j=1
c     do 4443 j=2,n-1
      do 4443 k=1,3
4443  a(i,j,k)=0.

c***********end of calculation of a(ijk)**************************

c  CALCULATION OF E(I,J) AND ITS FIRST DERIVATIVE WRT RIJ2
c  (Equation 5.4 of Section 5)

      energy=0.

      ep(n,n)=0.
      do 8222 i=1,n-1
      ep(i,i)=0.
      do 8222      j=i+1,n
      rho=rij0(i,j)/(rij(i,j,1)**2+rij(i,j,2)**2+rij(i,j,3)**2)**.5
c     this rho is the inverse of the one defined on page 5
      ep(i,j)=6.*sij(i,j)/rij0(i,j)**2*(-rho**14+rho**8)
      ener(i,j)=sij(i,j)*(rho**12-2.*rho**6)
8222  continue
      do 57220 i=1,n-1
      do 57220 j=i+1,n
57220 ep(j,i)=ep(i,j)

      do 48102 i=1,n-1
      ener(i+1,i)=0.
      ener(i,i+1)=0.
```

```
      ep(i+1,i)=0.
48102 ep(i,i+1)=0.
      do 48103 i=1,n-2
      ener(i+2,i)=0.
      ener(i,i+2)=0.
      ep(i+2,i)=0.
48103 ep(i,i+2)=0.
      do 49103 i=1,n-3
      ener(i+3,i)=0.
      ener(i,i+3)=0.
      ep(i+3,i)=0.
49103 ep(i,i+3)=0.

c     OVERALL ENERGY OF THE INSTANTANEOUS CONFORMATION
      do 82202 i=1,n-3
      do 82202 j=i+3,n
82202 energy=energy+ener(i,j)


c     CALCULATION OF B(K,L)
      topii=0.
      topjj=0.
      topij=0.
      do 3257 k=2,n-1
      do 3257 l=2,n-1
      sum7=0.
      do 3256 i=1,n
      do 3256 j=1,n
      topii=a(i,k,1)*a(i,l,1)+a(i,k,2)*a(i,l,2)+a(i,k,3)*a(i,l,3)
      topii=topii*h(k,i)*h(l,i)
      topjj=a(j,k,1)*a(j,l,1)+a(j,k,2)*a(j,l,2)+a(j,k,3)*a(j,l,3)
      topjj=topjj*h(k,j)*h(l,j)
      if(i.le.k.or.j.le.l) then
      topij=0.
      goto 7721
      endif
      topij=a(i,k,1)*a(j,l,1)+a(i,k,2)*a(j,l,2)+a(i,k,3)*a(j,l,3)
7721  sum7=sum7+ep(i,j)*(topii+topjj-2.*topij)
```

```
3256  continue
      bp(k,l)=sum7
3257  continue

C-------------------------------------------------------------


      write(6,*)'end-to-end vector   ',rij(1,n,1),rij(1,n,2),
     :  rij(1,n,3),'***********energy**',energy, 'at step=',m1
      erk(m1)=energy
      write(127,*) m1,erk(m1)

      if (energy.gt.(olden))then
      jbol=100000
      if (energy.lt.(olden+0.5)) then
      jbol=2
      endif
      if ((energy.lt.(olden+1.0)).and.(energy.gt.(olden+0.5))) then
      jbol=4
      endif
      if ((energy.lt.(olden+1.5)).and.(energy.gt.(olden+1.0))) then
      jbol=8
      endif
      else
      jbol=1
      endif

      olden=energy

      do 4011 k=2,n-1
      f(k,1)=h(k,n)*a(n,k,1)
      f(k,2)=h(k,n)*a(n,k,2)
      f(k,3)=h(k,n)*a(n,k,3)
4011  continue

c     END-TO-END VECTOR MAGNITUDE
      r2nmag=rij(1,n,1)*rij(1,n,1)+rij(1,n,2)*rij(1,n,2)+
    :rij(1,n,3)*rij(1,n,3)
```

```
      r2nmag=r2nmag**0.5
      write(128,*)m1,r2nmag

      if(i111.eq.1) go to 2009
      if (r2nmag.le.5.6) then
      i111=1
      ix=m1
      ntot=ix+1
      endif

c     DETERMINATION OF THE CORRECTION STEP SIZE
c     U1-2-3 REPRESENTS THE END-TO-END VECTOR AIMED
c     AT THIS STEP

2009  u1=rij(1,n,1)/r2nmag
      u2=rij(1,n,2)/r2nmag
      u3=rij(1,n,3)/r2nmag

      delx=(-u1/jbol)*0.04
      dely=(-u2/jbol)*0.04
      delz=(-u3/jbol)*0.04

c     FORMING THE LEFT HAND SIDE OF EQUATION 5.36
c     with dimensions (N+1)x(N+1)
c     (N-2 dihedrals and 3 lambdas)

c     the last three columns
      do 8112 k=1,n-2
      d(k,n-1)=f(k+1,1)
      d(k,n)=f(k+1,2)
8112  d(k,n+1)=f(k+1,3)

c     the upper left (n-1)x(n-1) block
      do 8912 k=1,n-2
      do 8912 l=1,n-2
c     write(9,*)bp(k+1,l+1)
      d(k,l)=bp(k+1,l+1)+bp(l+1,k+1)
8912  continue
```

```
c       lower left and right portions
        do 8113 k=1,n-2
        d(n-1,k)=d(k,n-1)
        d(n,k)=d(k,n)
8113    d(n+1,k)=d(k,n+1)
        do 8114 i=n-1,n+1
        do 8114 j=n-1,n+1
8114    d(i,j)=0.

        kkk=n+1
        call inv(d,kkk)
c       the inverse comes back with the same name (d)

c   creating the constant column matrix of Equation 5.36
        do 6612 i=1,n-2
6612    cc(i)=0.
        cc(n-1)=delx
        cc(n)= dely
        cc(n+1)=delz

c       write(6,*)' '
c       write(6,*)'the components of the constant column matrix '
c       write(6,*)' '
c       write(6,*)(cc(i),i=1,n+1)

c-----------------------------------------------------------------------
c       multiplying the column matrix with the d^-1 matrix. The first n-1
c       elements of the resulting column matrix are the deltafi's, and the last
c       three are the three Lagrange multipliers (force required to move
c       satisfy the constraints, here approaching terminal atoms)

        do 5411 i=1,n+1
        sum9=0.
        do 5412 j=1,n+1
5412    sum9=sum9+d(i,j)*cc(j)
5411    final(i)=sum9
c       write(6,*)'Lagrange ',final(n-1),final(n),final(n+1)
```

```
      m1=m1+1
      do 7333 i=2,n-1
7333  fi(i)=fi(i)+final(i-1)
      if(m1.le.ntot) goto 5555

      write (*,*) ntot,'ntot'
      write(2,*)'final dihedrals '
      write(6,*)'final dihedrals '
      do 72292 i=2,n-1
      write(6,4993)i,fi(i)*180./pi
72292 write(2,4993)i,fi(i)*180./pi
      write(2,*)'final Lagrange multipliers'
      write(2,*)final(n-1),final(n),final(n+1)
      write(6,*)'final Lagrange multipliers'
      write(6,*)final(n-1),final(n),final(n+1)
4922  format(3f15.10)
4993  format(i4,f8.1)
      close (2)

      stop
      end


      subroutine inv(W,N)
      real A(50,100),W(50,50),BL(50,50)

      DO 31 I=1,N
      DO 31 J=1,N
31    A(I,J)=W(I,J)
      EPSIL=1.0E-20
      N1=2*N
      DO 3999 I=1,N
      DO 3999 J=1,N
      IF (I-J) 1999,2999,1999
1999  A(I,J+N)=0.0
      GO TO 3999
2999        A(I,J+N)=1.0
3999  CONTINUE
```

```
      DO 5099 IP=1,N
      IM=IP
      IST=IP+1
      IF(IP-N) 4999,1009,4999
4999  DO 1099 I=IST,N
      IF (ABS(A(IM,IP))-ABS(A(I,IP))) 9999,1099,1099
9999  IM=I
1099  CONTINUE
1009  IF (ABS (A(IM,IP))-EPSIL) 1199,1399,1399
1199  continue
1299  FORMAT ('    PIVOT ELEMENT=',E10.3)
      IF(A(IM,IP)) 1399,7099,1399
1399          IF (IM-IP) 1499,2099,1499
1499   DO 1599 J=IP,N1
     AL=A(IP,J)
     A(IP,J)=A(IM,J)
1599   A(IM,J)=AL
2099   AL=A(IP,IP)
      A(IP,IP)=1.0
      DO 2599 J=IST,N1
2599   A(IP,J)=A(IP,J)/AL
     DO 4099 I=1,N
     IF (I-IP) 3099,4099,3099
3099   AL=A(I,IP)
     DO 3599 J=IP,N1
3599   A(I,J)=A(I,J)-AL*A(IP,J)
4099   CONTINUE
5099   CONTINUE
      NP=N+1
7099  CONTINUE
102   FORMAT(12F9.3)
      DO 1072 I=1,N
      DO 1072 J=1,N
      SUM=0.
      DO 1074 K=1,N
      SUM=SUM+W(I,K)*A(K,J+N)
1074  BL(I,J)=SUM
1072  CONTINUE
```

```
      DO 6677 I=1,N
      DO 6677 J=1,N
6677  W(I,J)=A(I,N+J)

      RETURN
      END
```

REFERENCES

1. Levinthal, C., "Are There Pathways For Protein Folding?," *Journal of Chimie Physique*, Vol. 65, pp. 44-45, 1968.

2. Shortle, D., Y. Wang, J. R. Gillespie, and J. O. Wrabl, "Protein Folding for Realists: A Timeless Phenomenon,"*Protein Science*, Vol. 5, pp. 991-1000, 1996.

3. Groot, B. L., D. M. F. van Aalten, R. M. Scheek, A. Amadei, G. Vriend, and H. J.C. Berendsen, "Prediction of Protein Conformational Freedom From Distance Constraints," *PROTEINS: Structure, Function and Genetics*, Vol. 29, pp. 240-251, 1997.

4. Aszodi, A., M. J. Gradwell, and W. R. Taylor, "Global Fold Determination From a Small Number of Distance Restraints," *Journal of Molecular Biology*, " Vol. 251, pp. 308-326, 1995.

5. Smith-Brown, M. J., D. Komino, and R. M. Levy, "Global Folding of Proteins Using a Limited Number of Distance Restraints," *Protein Engineering*, Vol. 6, pp. 605-614, 1993.

6. Braun, W., and N. Go, "Calculation of Protein Conformations by Proton-Proton Distance Constraints. A New Efficient Algorithm," *Journal of Molecular Biology*, Vol. 186, pp. 611-626, 1985.

7. Clore, G. M., M. A., Robien, and A. M. Gronenborn, "Exploring the Limits of Precision and Accuracy of Protein Structures Determined by Nuclear Magnetic Resonance Spectroscopy," *Journal of Moleculare Biol.*ogy, Vol. 231, pp. 82-102, 1993.

8. Gronenborn, A. M., and G. M. Clore, "Where is NMR Taking Us?," *Proteins: Structure Function and Genetics*, Vol. 19, pp. 273-276, 1994.

9. Guentert, P., W. Braun, and K. Wutrich, "Efficient Computation of Three-Dimensional Protein Structures in Solution from Nuclear Magnetic Resonance Data Using the Program DIANA and the Supporting Programs CALIBA, HABAS and GLOMSA," *Journal of Molecular Biology*, Vol 217, pp. 517-530, 1991.

10. Havel, T.F., and K. Wutrich, "An Evaluation of the Combined Use of Nuclear Magnetic Resonance and Distance Geometry for the Determination of Protein onformation in Solution," *Journal of Molecular Biology*, Vol. 182, pp. 281-294, 1985.

11. Kolinski, A., and J. Skolnick, "Monte Carlo Simulations of Protein Folding. I. Lattice Model and Interaction Scheme,"*Proteins: Structure, Function and Genetics* , Vol. 18, pp. 338-352, 1994.

12. Skolnick, J., A. Kolinski, and A. R. Ortiz, "MONSSTER: A Method for Folding Globular Proteins with a Small Number of Distance Restraints,"*Journal of Molecular Biology*, Vol. 265, pp. 217-241, 1997.

13. Jernigan, R. L., G. Raghunathan, and I. Bahar, "Characterization of Interactions and Metal Ion Binding Sites in Proteins," *Current Opinion in Structural Biology*, Vol. 4, pp. 256-263, 1994.

14. Bernstein, F. C., T. F. Koetzle, G. J. B., Williams, J. E. F., Meyer, M. D. Brice, J. R. Rodgers, O., Kennard, T. Shimanuchi, and M. J. Tasumi, "The Protein Data Bank: A Computer-based Archival File for Macromolecular Structures," *Journal of Molecular Biology*, Vol. 112, pp. 535-542, 1977.

15. Bahar, I., and B. Erman, "Kinematics of Polymer Chains with Freely Rotating Bonds in Restrictive Environment. 1. Theory ," *Macromolecules*, Vol. 25, pp. 6309-6314, 1992.

16. Bohinski, R. C., *Modern Concepts in Biochemistry*, Allyn and Bacon, Inc., Boston, 1983.

17. Branden, C., and J. Tooze, *Introduction to Protein Structure*, Garland Publishing, Inc., New York, 1991.

18. Voet, D., and J. G. Voet, *Biochemistry*, John Wiley & Sons, Inc., New York, 1995.

19. Schulz, G. E., and R. H. Schirmer, *Principles of Protein Structure*, Springer -Verlag, Inc., New York, 1979.

20. Creighton, T. E., *PROTEINS Structures and Molecular Properties*, W. H. Freeman and Company, New York, 1996.

21. R. Durley, L. Chen, L. W. Lim, F. S. Mathews, and V. L. Davidson, "Crystal Structure Analysis of Amicyanin and Apoamicyanin From *Paracoccus Denitrificants* at 2.0 Angstroms and 1.8 Angstroms Resolution," *Protein Science*, Vol 2, pp. 739-744, 1993.

22. Messerschmidt, A., R. Ladenstein, R. Huber, M. Bolognesi, R. Avigliano, R. Petruzzelli, A. Rossi, and A. Finazzi, "Refined Crystal Structure of Ascorbate Oxidase at 1.9 Angstroms Resolution," *Journal of Molecular Biology*, Vol. 224, pp. 179-184, 1992.

23. Baker, E., "Structure of Azurin from *Alcaligenes Denitrificans* Refinement at 1.8 Å Resolution and Comparison of the Two Crystallographically Independent Molecules," *Journal of Molecular Biology*, Vol. 203, pp. 1071, 1988.

24. Djinovic, K., K. Gatti, A. Coda, L. Antolini, G. Pelosi, A. Desideri, M. Falconi, F. Marmocchi, G. Rotilio, and M. Bolognesi, "Structure Solution and Molecular Dynamics of Refinement 2.0 AX of the Yeast *Cu, *Zn Enzyme Superoxide Dismutase," *Acta Crystallograpy*, Section B, Vol 47, pp. 918-935, 1991.

25. Smith, C. A., H. M. Baker, and E. N. Baker, "Preliminary Crystallographic Studies of Copper (II)- and Oxalate-Substituted Human Lactoferrin," *Journal of Molecular Biology*, Vol. 219, pp. 155-167, 1991.

26. Anderson, B. F., H. M. Baker, G. E. Norris, D. W. Rice, and E. N. Baker, "Structure of Human Lactoferrin: Crystallographic Structure Analysis and Refinement at 2.8 Angstroms Resolution," *Journal of Molecular Biology*, Vol. 209, pp. 711-716, 1989.

27. Chen, L., R. Durley, and F. S. Mathews, "Crystal Structure of an Electron-Transfer Complex Between Methylamine Dehydrogenase and Amicyanin," *Biochemistry*, Vol. 31, pp. 4959-4963, 1992.

28. Chapman, G. V., P. M. Colman, H. C. Freeman, J. M. Guss, M. Murata, V. A. Norris, J. A. M. Ramshaus, and M. P. Venkatappa, "Preliminary Crystallographic Data for Copper-Containing Protein, Plastocyanin," *Journal of Molecular Biology*, Vol 110, pp. 187-194, 1977.

29. Gus, J. M., and H. C. Freeman,"Structure of Oxidized Poplar Plastocyanin at 1.6 Angstroms Resolution," *Journal of Molecular Biology*, Vol 169, pp. 521-526, 1983.

30. Katti, S. K., D. M. Lemaster, and H. Eklund, "The Refined Structure of *Escherichia coli* Thioredoxin," *Journal of Molecular Biology*, Vol. 212, pp. 167-170, 1990.

31. Ferrin, T. E., C. C. Huang, L. E. Jarvis, and R. Langridge, "The MIDAS Display System," Journal of Molecular Graphics, Vol. 6, pp. 13-27, 1988.

32. Kalverda, A. P., J. Salgado, C. Dennison, and G. W. Canters, "Analysis of the Copper (II) Paramagnetic Site of Amicyanin by H NMR Spectroscopy," Biochemistry, Vol 35, pp. 3085-3092, 1996.

33. Romero, A. H. Nar, R. Huber, A. Messerschmidt, A. Kalverda, A. Canters, R. Durley, and F. S. Mathews, "Crystal Structure Analysis and Refinement at 2.15 A Resolution of Amicyanin, a Type I Blue Copper Protein, from *Thiobacillus versutus*," *Journal of Molecular Biology*, Vol. 236, pp. 1196-1211, 1994.

34. Bahar, I. and B. Erman, "Investigation of Local Motions in Polymers by the Dynamic Rotational Isomeric State Model," *Macromolecules*, Vol. 20, pp. 1368-1376, 1987.

35. Bahar, I., B. Erman, and L. Monnerie, "Comparison of Dynamic Rotational Isomeric State Results with Previous Expressions for Local Chain Motion," *Macromolecules*, Vol. 22, pp. 431-437, 1989.

36. Bahar, I., B. Erman, and L. Monnerie, "Kinematics of Polymer Chains with Freely Rotating Bonds in a Restrictive Environment. 2. Conformational and Orientational Correlations," *Macromolecules*, Vol. 25, pp. 6314-6321, 1992.

37. Erman, B., and I. Bahar, "Local Dynamics of Freely Rotating Polymer Chains in Dense Systems," *Progress in Colloids and Polymer Science*, Vol. 91, pp. 16-19, 1993.

38. Bahar, I., N. Baysal, B. Erman, and L. Monnerie, "Kinematics of Polymer Chains in Dense Media. 3. Influence of Intramolecular Conformational Potentials," *Macromolecules*, Vol. 28, No. 4, pp. 1038-1048, 1995.

39. Baysal, C., "Efficient Computational Models and Methods for Investigating Local Polymer Dynamics," Ph.D. Dissertation, Boğaziçi University, 1996.

40. Flory, P. J., *Statistical Mechanics of Chain Molecules*, Interscience, New York, 1969.

41. Flory, P. J., "Foundations of Rotational Isomeric State Theory and General Methods for Generating Configurational Averages," *Macromolecules*, Vol. 7, pp. 381,1974.

42. Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *Journal of Chemical Physics*, Vol 51, pp. 1087-1092, 1953.

43. Kim, P. S., and R. L. Baldwin, "Specific Intermediates in the Folding Reactions of Small Proteins," *Annual Review of Biochemistry*, Vol.51, pp. 459-489, 1982.

44. Uversky, V. N. and O. B. Ptitsyn, ""Partly Folded", a New Equilibrium State of Protein Unfolding of β-Lactamase at Low Temperatures," *Biochemistry*, Vol. 33, pp. 2782-2791, 1994.