PROTEINS: Structure, Function, and Genetics 27:410–424 (1997)

A New Method for Modeling Large-Scale Rearrangements of Protein Domains

Vladimir Maiorov and Ruben Abagyan*
The Skirball Institute of Biomolecular Medicine and Biochemistry Department, New York University Medical Center, New York, New York

ABSTRACT **A method for modeling large-scale rearrangements of protein domains connected by a single- or a double-stranded linker is proposed. Multidomain proteins may undergo substantial domain displacements, while their intradomain structure remains essentially unchanged. The method allows automatic identification of an interdomain linker and builds an all-atom model of a protein structure in internal coordinates. Torsion angles belonging to the interdomain linkers and side chains potentially able to form domain interfaces are set free while all remaining torsions, bond lengths, and bond angles are fixed. Large-scale sampling of the reduced torsion conformational subspace is effected with the "biased probability Monte Carlo-minimization" method [Abagyan, R.A., Totrov, M.M. (1994): J. Mol. Biol. 235, 983–1002]. Solvation and side-chain entropic contributions are added to the energy function. A special procedure has been developed to generate concerted deformations of a double-stranded interdomain linker in such a way that the polypeptide chain continuity is preserved. The method was tested on Bence-Jones protein with a single-stranded linker and lysine/arginine/ornithine-binding (LAO) protein with a double-stranded linker. For each protein, structurally diverse low-energy conformations with ideal covalent geometry were generated, and an overlap between two sets of conformations generated starting from the crystallographically determined "closed" and "open" forms was found. One of the low-energy conformations generated in a run starting from the LAO "closed" form was only 2.2 Å away from the structure of the "open" form. The method can be useful in predicting the scope of possible domain rearrangements of a multidomain protein. Proteins 27:410-424, 1997.**

Key words: domain movements; inter-domain linkers; conformational calculations; Monte Carlo-minimization method; Bence-Jones protein; lysine/arginine/ornithine-binding protein

INTRODUCTION

Domain rearrangements are essential for functioning of large proteins. A variety of different interdomain configurations can be realized depending on interactions with other proteins or ligands.[1–5] Several proteins in the Protein Data Bank (PDB)[6,7] are known to have different domain arrangements in oligomeric complexes or in a set of homologues.[8–17] Analysis of the underlying mechanisms of protein domain flexibility[5,18,19] revealed two main types of motions: "shear" motion involving a large number of amino acid residues distributed over extended areas, and "hinge-bending" where a relatively small portion of a polypeptide chain significantly changes mutual domain position.

Modeling of complex concerted rearrangements of protein domains at the atomic level of detail is a serious theoretical and computational problem mainly due to the large time scale of these rearrangements and the size of a multidomain protein. Earlier studies were primarily concerned with analysis of the domain motions.[20–22] Later, domain motions were modeled by molecular dynamics[23–26] where small- to medium-scale domain movements were reproduced. Most often, the difficulty of escaping from the close vicinity of a starting conformation was found to be formidable. Broader sampling of the conformational space could be achieved if the covalent or intradomain structure of a protein was preserved by a series of restraints,[26–28] and/or atomic masses were artificially increased.[29] Among other methods that may be applicable to large-scale domain rearrangement modeling, are random and friction forces molecular dynamics,[30] internal coordinate molecular dynamics,[31–34] stochastic rigid body, and brownian dynamics with simplified residue representation,[35–36] Monte Carlo calculations in space of scaled collective internal variables,[37] and "essential dynamics."[38,39]

We propose a method for generating a diverse series of low-energy conformations of a multidomain protein by means of a large-scale sampling of conformational space of the torsion angles of a polypeptide

*Correspondence to: Dr. Ruben Abagyan, The Skirball Institute of Biomolecular Medicine & Biochemistry Department, NYU Medical Center, 540 1st Avenue, New York, NY 10016. E-mail: abagyan@earth.med.nyu.edu

Received 15 July 1996; accepted 5 September 1996.

chain. The observation that only a relatively small portion of the chain, namely, an interdomain linker and side chains at the domain interfaces, undergo significant conformational changes upon domain rearrangements[1,2,5] is exploited by the method. The modeling procedure relies on internal coordinate mechanics (ICM)[40,41] earlier applied to protein design and loop prediction,[42] side-chain prediction,[43] ab initio peptide structure prediction,[41] protein docking,[44] inverse protein folding,[45] and homology modeling.[46] The proposed method was tested on two PDB structures, each of which has two crystal forms manifesting remarkable differences in the relative domain positions: Bence-Jones protein (human VL λ *Loc*)[8] with a single-stranded linker and lysine/arginine/ornithine-binding protein from *Salmonella typhimurium*[47] with a double-stranded linker. We show that the proposed method can generate a structurally diverse set of the conformations with ideal covalent geometry and low energies distributed in a relatively narrow interval during a single run, starting from one known protein conformation, and that structural overlap between the sets generated from different conformations of one protein can be achieved.

METHODS

Internal Coordinate Mechanics

All calculations and analyses were performed with the ICM program, developed for molecular modeling and structure predictions by global restrained energy optimization of arbitrarily constrained molecules.[40,41] An arbitrary set of standard torsion angles of the polypeptide chain (main-chain ϕ, ψ, ω, and side-chain χ) may be set free, while the remaining internal variables (other torsion angles, and all bond lengths and bond angles) are kept unchanged resulting in an automatically constructed set of "rigid bodies." This representation of a protein structure will further be referred to as an "ICM model." Only changeable contribution to the energy and the gradient with respect to the free torsions are considered. Therefore, the computational cost of a single energy evaluation decreases as the number of constrained torsions increases. Concurrently, the conformational space drops drastically, thus making sampling a computationally tractable task even for a relatively large protein. The conformational space of the ICM model can be efficiently sampled by a combined biased probability Monte Carlo-minimization method using optimized random moves[41] (see below). The energy is calculated with ECEPP/3 force field[48–50] extended by recently developed solvation and side-chain entropic terms.[41] The following terms were included in the energy function: van der Waals and 1–4 nonbonded interactions, hydrogen bonding, torsion, electrostatic, disulfide bond constraints, solvation energy, and side-chain entropy. Cutoff distance for truncation of van der Waals and electrostatic interactions was set to 7.5 Å, and for hydrogen bonding interactions it was set to 3.0 Å. To accelerate time-consuming redefinition of the atomic interaction lists for the calculations of the van der Waals and electrostatic interactions, the atoms close in space were associated into groups, and the redundant distance checks for atom pairs were avoided if atom group centers were too far from each other.[41] The electrostatic energy was calculated with distance-dependent dielectric constant $D_{\text{diel}} = 4r$. Modified atomic solvation parameters of Wesson and Eisenberg[51] and side-chain entropic contributions were taken into account as described earlier.[41] A special "tether" term was added to the energy function for regularization and local deformation calculations (see below). Tethers are defined as a series of harmonic restraints confining ICM model atoms to the corresponding atom locations of a given three-dimensional template. The tether term weight was 1 kcal/mol/Å, except for the regularization procedure where it was recalculated at each step of the annealing part of the procedure.

Biased Probability Monte Carlo Minimization Method

Li and Scheraga[52] demonstrated that the performance of conformational sampling can be substantially improved in a modified Monte Carlo scheme where local energy minimization follows each random move before applying the Metropolis criterion.[53] Later, it was shown that the performance of this "Monte Carlo minimization"[54–57] can further be increased by designing a set of moves of torsion angle groups with biased probabilities derived from the distributions in a survey of known protein three-dimensional structures.[41] The method was implemented in the ICM program as a combined biased probability Monte Carlo (BPMC) minimization protocol.[40,41] It is important to emphasize that BPMC is different from a conventional Monte Carlo protocol. First, "random" moves are not random, but rather "probability-biased," and, second, the condition of microscopic reversibility does not hold because of the biased moves and the minimization step, so the application of the uncorrected acceptance criterion is somewhat arbitrary. As such, BPMC has certain limitations; for example, it cannot be used for evaluation of the relative probabilities of the generated conformational states in an ensemble. However, it has been proven to be an efficient method of global optimization and conformational sampling,[40–46] and our preliminary results show that it can also be adapted for the domain rearrangement calculations.

Conformational Stack

ICM allows combination of the BPMC calculations with the generation and run-time maintenance of a list of low energy conformations, the "conformational stack."[58] Each time a new conformation is accepted

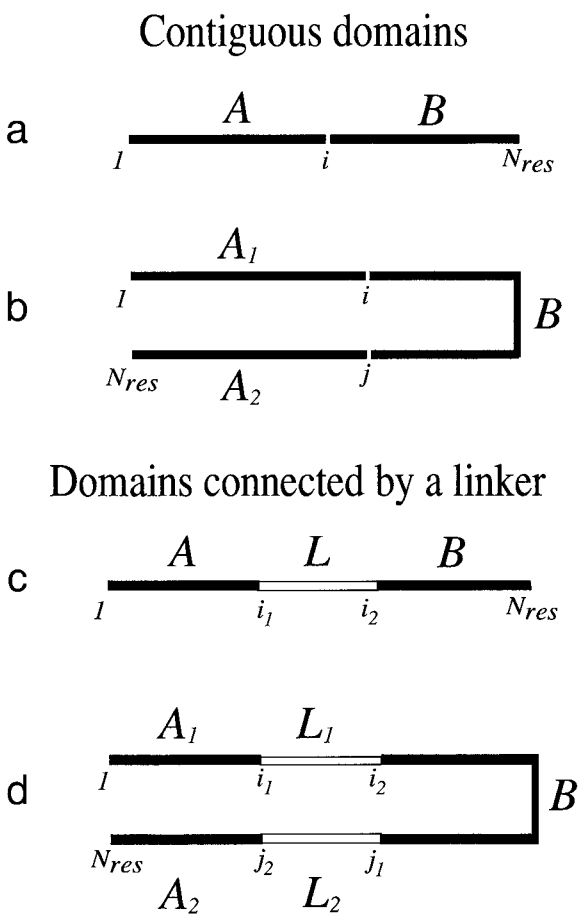

Fig. 1. Domain/linker assignments in a two-domain protein structure: (**a**) two single-segment domains *A* and *B*; (**b**) a double-segment domain *A* (A_1, A_2) and a single-segment domain *B*; (**c**) a single-stranded linker *L* connects domains *A* and *B*; (**d**) a double-stranded linker *L* (L_1, L_2) connects domain *A* (A_1, A_2) and domain *B*. The positions of the first, the last (N_{res}th) and the domain/linker boundary residues are indicated.

during a BPMC run, it is quantitatively compared to all those from the stack collected so far. Conformations are considered similar if their structural difference does not exceed a certain similarity cutoff specified by the "vicinity" parameter. It is convenient to characterize changes in the position of one domain with respect to the other in terms of a "static" root-mean-square deviation (sRMSD). For our analyses, the sRMSD was defined as a root-mean-square deviation of Cα atoms of the domain *B* provided Cα atoms of the domain *A* are optimally superimposed (Fig. 1). The parameter "vicinity" in the BPMC calculations was specified in terms of sRMSD, and its value was adjusted in a series of preliminary calculations for each considered protein structure.

If a new conformation is not similar to all those already collected, a new slot is created in the stack where the new conformation is placed. If the new conformation is within the value of the vicinity parameter of any conformation already in the stack, it substitutes for the last one if its energy is lower, and is disregarded otherwise. The size of the conformational stack defines the maximal number of conformations simultaneously residing in the stack, and should be specified before the calculations. If the stack is full, but the run is continuing and the accepted conformation differs from all those already in the stack, the new conformation replaces that with the highest energy. In summary, the conformational stack provides a flexible and convenient way to collect a structurally diverse set of low-energy conformations during a BPMC run.

Regularization of Crystal Structures

Before starting the calculations, we "regularized" a protein three-dimensional structure, that is, we built an all-atom ICM model of the structure with low-energy, idealized covalent geometry and within about 0.5 Å of all heavy atoms RMSD of the original conformation. The regularization procedure included the following steps:[40,43]

1. Generation of an extended, all-atom chain of the given protein with the standard covalent geometry
2. Imposing a set of tethers to relate the positions of the ICM model nonhydrogen atoms to the positions of corresponding crystal structure atoms
3. Building up the correct conformation starting from N-terminus of the chain by adding atoms one-by-one in a succession
4. Adjustment of the methylhydrogen positions of alanine, valine, and leucine residues by varying corresponding side-chain torsions
5. Annealing the whole structure in a multistep iterative procedure where the contribution of the tether term is gradually decreased
6. Optimization of cystein, serine, threonine, and tyrosine polar hydrogens by systematic conformational search.

The only difference from the previously described regularization[40,43] is in the definition of the tether term weight during the annealing cycle. At *k*th iteration, tether weight $W_{tz}^{(k)}$ (in kcal/mol/Å^2) was calculated as follows:

$$W_{tz}^{(k)} = W_{tz}^{(k-1)} \times 2\,E_{\text{covalent}}/E_{tz},$$

$$W_{tz}^{(k)} = 1/(2 - W_{tz}^{(k)}), \quad \text{if } W_{tz}^{(k)} < 1,$$

$$W_{tz}^{(0)} = 1, \qquad (1)$$

where E_{covalent} stands for the sum of van der Waals, 1–4, torsion and hydrogen bonding terms, and E_{tz} stands for tether term contributions. Use of Equation (1) leads to the gradual decrease of the relative significance of the tether term as van der Waals

repulsion energy decreases and allows to attain a lower final energy for the regularized structure.

Domain Rearrangement Modeling Protocol

Only two-domain protein structures were considered in this work (Fig. 1). The main assumption of the method is that a relatively small fraction of all standard torsion angles, namely, those belonging to the linker segment(s) and the side chains of the extended domain interface, is sufficient to reproduce domain rearrangements in the course of the BPMC calculations. Different low-energy conformations collected in the stack are considered as possible states of the protein structure undergoing domain rearrangements. Upon completion of the BPMC calculations, the conformations are additionally minimized ("relaxed") with respect to all standard torsion angles. The protocol can be summarized as follows.

1. Identify a single- or a double-stranded linker between the domains (see below).
2. Calculate solvent accessibilities of the isolated domains, i.e., if one domain is considered, than the other domain and the linker portion of the chain are disregarded.
3. Select torsion angles potentially essential for the domain rearrangements: (i) main- and side-chain torsion angles of the linker, and (ii) side-chain torsion angles of the previously identified solvent accessible residues (step 2), which are within a sphere of a given radius from the atoms of the other domain.
4. Set free all the torsion angles selected at step 3, and fix all remaining internal variables at their values in the crystal structure. This results in an automatic generation of a set of "rigid bodies" within which all interactions are excluded from consideration.
5. Apply the BPMC procedure to generate a series of conformations with low energy evaluated as a sum of vacuum potentials, solvation energy and side-chain entropy terms. In case of a two-stranded linker, use a special loop closure algorithm for generation of the concerted conformational changes in both strands (see below).
6. Relax the conformations collected in the conformational stack by an iterative minimization with adaptable diminishing tether restraints.

Accessible surface area[59] was calculated with water probe radius equal to 1.4 Å. For each amino acid residue X in the extended conformation of Gly-X-Gly dipeptide unit the standard exposed surface was calculated. All residues for which the ratio of their accessible surface in a given conformation to the standard one exceeded a 25% threshold were considered as accessible to solvent (see step 2). Radius of a sphere around each atom of one domain was set to 5 Å (see step 3). Atoms interacting according to this definition are assumed to constitute a potential interdomain interface, and the corresponding side-chain torsion angles were automatically identified. Relaxation of the stack conformations was performed in the same manner as the iterative annealing step in the regularization procedure, except that the starting tether weight $W_{tz}^{(0)}$ was set to 0.01 kcal/mol/Å^2 (step 6).

Interdomain Linkers

Since the results of domain rearrangement calculations may depend on a set of free torsion angles, an objective identification of the interdomain linker is required. There are several methods for protein domain identification,[60–68] which, in principle, may be adapted for this task. We used Siddiqui and Barton's method,[69] which recognizes domains on the basis of interresidue contact patterns, namely, as fragments of the polypeptide chain that manifest the maximal density of intradomain contacts and, simultaneously, the minimal density of interdomain contacts. In our version of the method, we introduced the strength of the interresidue contacts, and identified a linker portion of the polypeptide chain rather than only one separation point (Fig. 1). Correspondingly, the original form of the "split function"[69] was changed (see Appendix). The modified method was tested on a series of proteins with single- and double-stranded linkers described in the literature. For the domain rearrangement calculations and for analyses of the generated conformations, the linkers defined by this algorithm were propagated beyond their ends by one residue.

Generating Conformational Changes in a Double-Stranded Linker

ICM program builds up a polypeptide chain in succession, starting from the N-terminus. In case of a double-stranded linker, the concerted changes of main-chain torsion angles in both linker strands should be generated to prevent a relative displacement of the rest of the domain A (segment A_2, Fig. 1d) with respect to the first segment (A_1) upon conformational changes in the linker. This can be interpreted as a local deformation[70,71] of a long segment of polypeptide chain (L_1, B, L_2) containing two flexible regions (L_1, L_2) with a rigid structure (domain B) in between. A random move does not change the mutual positions of segments A_1 and A_2 if local deformation conditions are taken into account. However, local minimization may displace segment A_2 in order to resolve energetic strains resulting from a random move. Since calculations are performed at a finite temperature, the distorted conformation may still be accepted, and the distortion may build up. To prevent that, the segment A_2 has to be "tethered" to the corresponding portion of the regularized crystal structure. In our experience, the minimization following each BPMC step always guarantees

that the structure of two-segment domain *A* is preserved (with the accuracy of about 0.01 to 0.05 Å of average atom positions) if the tether term is taken into account.

Analysis of the Relative Domain Positions

In this work, we used a simplified representation of a protein structure to describe relative domain positions. Five reference points were defined through Cα-atom coordinates: the centroids of two domains ($\mathbf{P}_1$ and $\mathbf{P}_5$), the centroids of each domain's boundary residues ($\mathbf{P}_2$ and $\mathbf{P}_4$), and the centroid of all boundary residues ($\mathbf{P}_3$) (Fig. 2a). If the linker is a single-stranded one, then the points $\mathbf{P}_2$ and $\mathbf{P}_4$ simply coincide with the Cα positions of the respective domain boundary residues i_1 and i_2. In case of a double-stranded linker, the positions of $\mathbf{P}_2$ and $\mathbf{P}_4$ are defined as average positions of i_1, j_2 and i_2, j_1, respectively. In either case, the position of $\mathbf{P}_3$ is defined as an average of $\mathbf{P}_2$ and $\mathbf{P}_4$. The reference points define a plane angle $\Theta = \Theta(\mathbf{P}_1\text{–}\mathbf{P}_3\text{–}\mathbf{P}_5)$, a distance $\Delta = \Delta(\mathbf{P}_2\text{–}\mathbf{P}_4)$, and a dihedral angle $\Phi = \Phi(\mathbf{P}_1\text{–}\mathbf{P}_2\text{–}\mathbf{P}_4\text{–}\mathbf{P}_5)$ (Fig. 2b). The parameters may be interpreted as a domain bending angle (Θ), an interdomain distance (Δ), and a domain twisting angle (Θ). Positive sign of the dihedral angle Φ was chosen to correspond to the clockwise rotation as seen in the direction from $\mathbf{P}_2$ to $\mathbf{P}_4$; zero value of the angle is in the eclipsed configuration of the vectors $\mathbf{P}_2\text{–}\mathbf{P}_1$ and $\mathbf{P}_4\text{–}\mathbf{P}_5$. These parameters are obviously not sufficient to rigorously describe the positions of two rigid bodies (domains), and additional parameters might be considered (for example, plane angles $\mathbf{P}_1\text{–}\mathbf{P}_2\text{–}\mathbf{P}_4$ and $\mathbf{P}_2\text{–}\mathbf{P}_4\text{–}\mathbf{P}_5$, not shown in Fig. 2). However, for general description of the relative domain positions, the parameters Θ, Δ, and Φ seem to be characteristic and reflect essential conformational changes. Similar parameters are often used to trace global changes in a protein structure.[12,26,72,73] The values of the parameters depend on the selection of the reference points. Thus, different conformations of a protein can be correctly compared with each other only if precisely the same domain/linker assignment is used. Two-dimensional diagrams of the parameters, for example, domain twisting versus domain bending, or interdomain distance versus domain bending, may also be useful for representation of the conformational space of the generated conformations and the structural overlap between them (see below).

To test the described method, we chose two proteins known to have conformational forms with substantially different relative domain positions and with different types of interdomain connection. The first was a monomer of Bence-Jones (BJ) protein dimer crystallized from distilled water[8] (PDB code 2bjl) (Fig. 3). The crystal structure of the protein was solved to 2.8 Å resolution and refined to R factor of 0.22. Each monomer includes 217 residues and forms

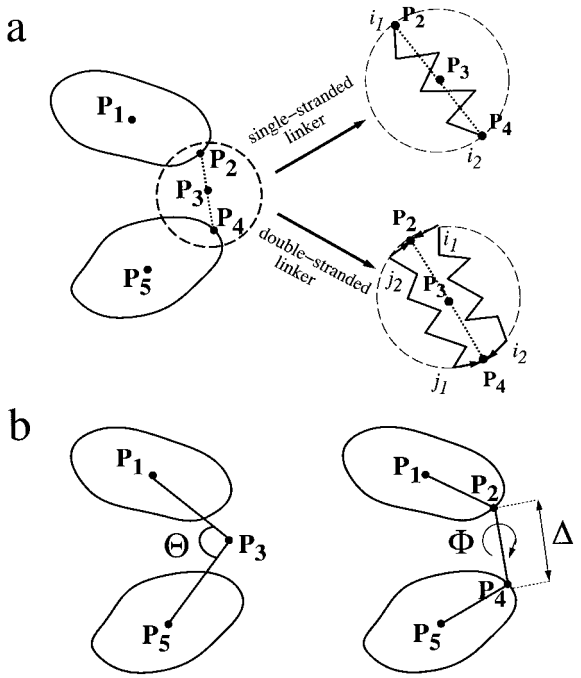

Fig. 2. Representation of the relative domain positions in a two-domain protein structure: (**a**) reference points $\mathbf{P}_1$ to $\mathbf{P}_5$; (**b**) domain bending angle Θ (*left*), domain twisting angle Φ and interdomain distance Δ (*right*).

two typical immunoglobulin β-sandwich domains connected by a single-stranded linker. The spatial structures of the individual domains in both crystal monomers are very similar: the backbone atom RMSD of the independently superimposed domains *A* and *B* equal 0.8 and 1.0 Å, respectively. The flexibility of the linker allows drastically different relative domain positions in the two crystal forms with sRMSD equal to 45.1 Å (Fig. 3). To indicate different degrees of domain bending, structures of the chains 1 and 2 of the protein will be referred to as "BJ closed" and "BJ open," respectively. The proline residue in the first position of the amino acid sequence was substituted for an N-terminal pyrrolidone carboxylic acid residue as in the crystallographic refinement,[8] and charged NH_3^+ and COO^- groups were added to the sequence at N and C termini, respectively. The structures of both crystal forms were independently regularized as described above. Nonhydrogen atom RMSD between the crystal and the corresponding regularized conformations was equal to 0.7 Å for both closed and open forms.

The second test object was lysine/arginine/ornithine-binding (LAO) protein[47] (Fig. 4). Three-dimensional structure of liganded (PDB code 1lst) and unliganded (21ao) forms of the protein were determined to 1.8 and 1.9 Å resolution and refined to R factor of 0.20 and 0.19, respectively. The backbone RMSD between individual domain structures of the two crystal forms are equal to 0.3 and 0.6 Å. The two

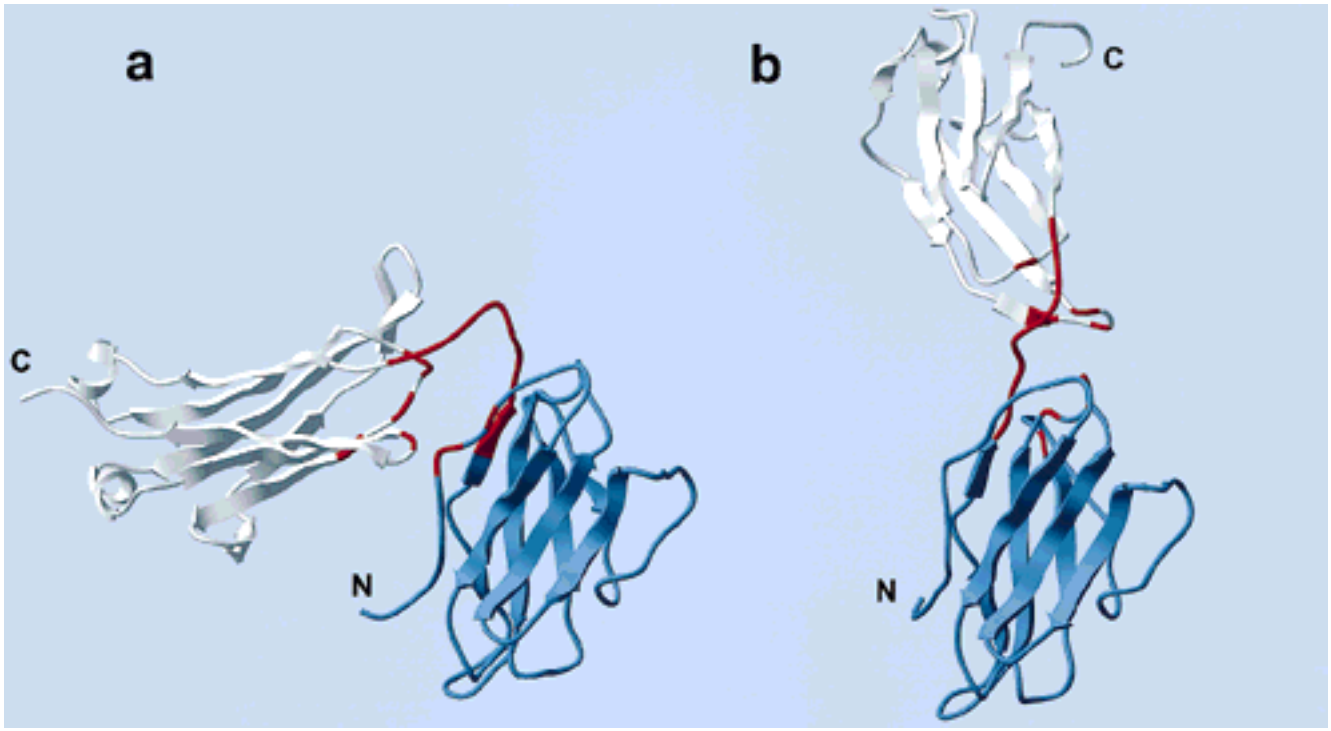

Fig. 3. Two crystal forms of Bence-Jones protein (PDB structure 2bjl) in a ribbon representation: (**a**) BJ closed form (chain 1); (**b**) BJ open form (chain 2). Two domains are colored in blue and white. Single-stranded linker and interdomain interfaces which torsion angles were set free for the calculations are shown in red. The positions of N and C termini are indicated.

domains are connected by a double-stranded linker, in contrast to other bacterial periplasmic binding proteins having triple-stranded interdomain connection.[13] Two strands of the linker form a hinge between the domains enabling the protein to accommodate "closed" (1lst) and "open" (21ao) conformations with the sRMSD between these two equal to 13.8 Å. In the following, we will refer to the 1lst and 21ao conformations as "LAO open" and "LAO closed," respectively. A dynamic exchange between these forms was suggested, and the stabilizing role of the ligand for the closed form was discussed.[47] The ligand was excluded from the consideration of the LAO closed structure, and charged NH_3^+ and COO^- groups were added to the sequence at N and C termini, respectively. After regularization, the LAO closed and the LAO open conformations were 0.5 Å RMSD from the corresponding crystal structures.

RESULTS

First, we identified the interdomain linkers in the regularized crystal conformations of the BJ and LAO proteins (Table I). Similar linker boundaries were found in both forms of each protein, in spite of the fact that the crystal conformations are quite different in their relative domain positions and, correspondingly, in the patterns of the intra- and interdomain contacts. No data on the domain/linker partition were reported for the BJ structure.[8] According to our analysis, linker boundaries derived from the closed and the open forms differed by only one residue. In the case of the LAO protein, two domain/linker assignments for the crystal conformations were consistent with each other within one residue accuracy, and the derived linkers included all but one residue reported by the authors of the crystallographic study.[47] To compare relative domain positions in the generated conformations, we used a "unified" linker/domain assignment combining the assignments independently derived from the crystal conformations.

Amino acid residues potentially involved in the interdomain interface were automatically identified as described in the Methods section, and the corresponding side-chain torsion angles as well as all those belonging to the linker residues were set free. In the case of the BJ protein, the total number of free torsion angles were 67 (BJ closed) and 66 (BJ open). Among them, 47 torsions were found to be identical in both lists, while 20 (BJ closed) and 19 (BJ open) were specific for a particular conformation (Table II). The difference between the lists were mainly in the side-chain torsions: 17 (BJ closed) and 19 (BJ open), while the main-chain ϕ, ψ, and ω torsion angles were essentially the same (Table I). In the two LAO crystal conformations, the numbers of free torsions were quite different: 151 (LAO closed) and 117 (LAO

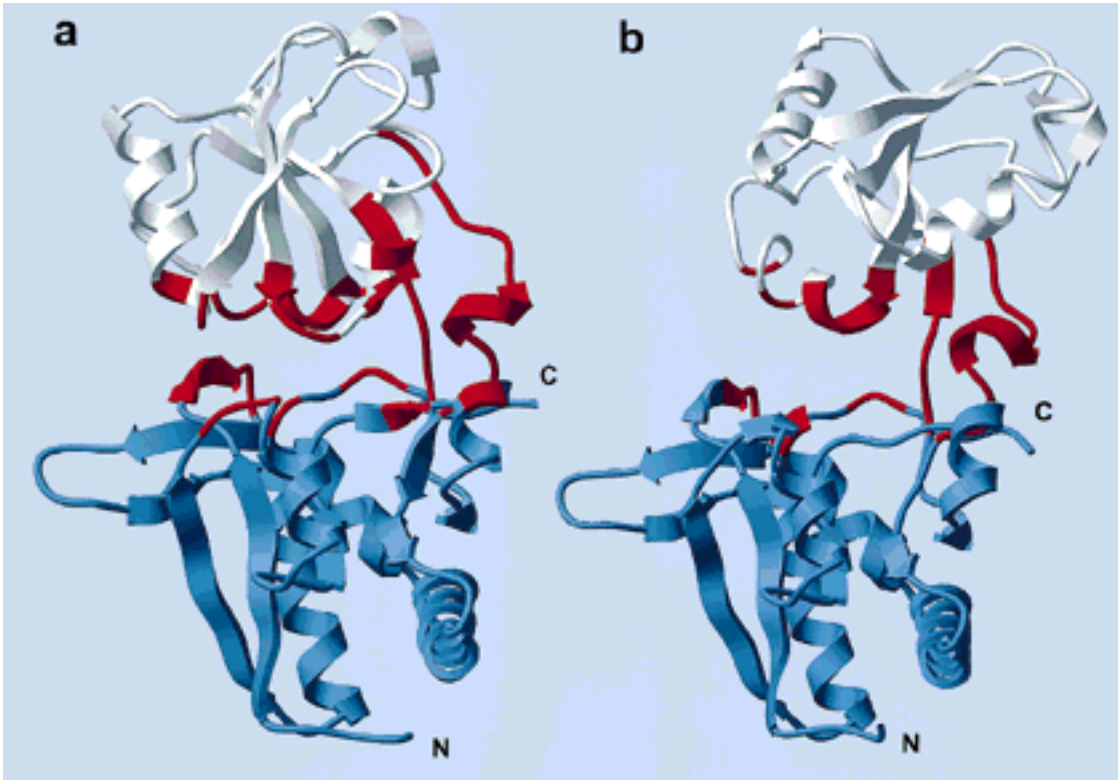

Fig. 4. Two crystal forms of lysine/arginine/ornithine-binding (LAO) protein in a ribbon representation: (**a**) LAO closed form (PDB structure 1lst); (**b**) LAO open form (2lao). Color coding and notations are the same as in Figure 3.

TABLE I. Domain/Linker Assignments in the Crystal Structures of the BJ and the LAO Proteins

Crystal form[a]	Domain *A*	Linker	Domain *B*
BJ protein			
Closed	1:107	108:116	117:217
Open	1:107	108:115	116:217
Unified[b]	1:107	108:116	117:217
LAO protein[c]			
Closed	1:88, 194:238	89:93, 182:193	94:181
Open	1:88, 195:238	89:93, 183:194	94:182
Unified[b]	1:88, 195:238	89:93, 182:194	94:181

[a]Domain/linker assignments were derived independently for each crystal form.
[b]Two assignments for two crystal forms were combined into a unified assignment for comparative analysis of the relative domain positions. One residue propagation of the linker strands was taken into account.
[c]Oh et al.[47] reported the identical linker assignments for both LAO crystal forms: 89:92, 186:194.

open). Among them, 40 torsion angles (mostly side-chain ones) were specific for the closed form as compared to only six for the open one. Comparison of the free torsion angle lists revealed that the free torsions of the LAO open structure constituted a subset of the free torsions of the LAO closed structure. This difference is due to a more extensive network of interdomain side-chain contacts in the LAO closed conformation. The residues of the BJ and the LAO proteins to which the free torsion angles belong are shown in red in Figures 3 and 4. Large-scale domain rearrangements should involve surpassing relatively high-energy barriers during a BPMC run. Therefore, the calculations were performed at a high temperature of 3000 K to enhance efficiency[75] of the sampling protocol. The size of the conformational stack was set to 50 conformations. For both proteins, two independent BPMC runs starting from the closed and the open crystal conformations were performed.

BJ Protein Calculations

The total numbers of the conformations accepted during the calculations were 2193 (BJ closed run) and 5876 (BJ open run). The maximum allowed number of conformations, 50, was collected in each stack by the end of the calculations. Stack conformations were subject to relaxation with respect to all standard torsions by 20,000 steps of local minimization (four iterations with the maximal number of function calls equal to 5000). At each iteration, the tether term contribution was weighted according to

TABLE II. Free Torsion Angles Used in the BPMC Calculations of BJ and LAO Proteins*

Protein[a]	In all[b]	In common[c]	In the closed form only[d]	In the open form only[d]
BJ	67 (26 + 41) (closed) 66 (23 + 43) (open)	47 (23 + 24)	20 (3 + 17)	19 (0 + 19)
LAO	151 (50 + 101) (closed) 117 (50 + 67) (open)	111 (47 + 64)	40 (3 + 37)	6 (3 + 3)

*Numbers in the parentheses, (m + n), indicate the number of free torsion angles: main-chain φ, ψ, and ω (m) and side-chain χ (n).
[a]Total number of standard torsion angles in the ICM models of the BJ and the LAO proteins was 1130 and 1311, respectively.
[b]Number of free torsion angles in the ICM model.
[c]Number of free torsions angles common in both closed and open forms.
[d]Number of free torsions angles specific to the given form.

Equation (1) to keep the relaxed structures in the vicinity of the corresponding nonrelaxed conformations. The values of nonhydrogen atom RMSD between the stack conformations and the corresponding relaxed ones were 0.4 to 0.6 Å.

The results of the calculations are summarized in Table III. Energies of the generated conformations were in the range 47.3 kcal/mol (BJ closed run) and 31.0 kcal/mol (BJ open run). Stack conformations manifested large structural diversity in their relative domain positions (Fig. 5): an average sRMSD with the starting crystal structure was 48.3 Å for the BJ closed run and 36.8 Å for the BJ open run. Interdomain distance Δ varied in the range of 12.7 Å (BJ closed run) and 10.6 Å (BJ open run). In both sets, a full 360° of changes in the domain twisting angle Φ were observed, and large-amplitude domain bending changes were found with the variations of Θ angle from 117° to 171° (BJ closed run) and from 131° to 178° (BJ open run). In general, the conformations tended to adopt a rather "extended" domain arrangement: average values of the domain bending angle Θ were 150° and 157° in the conformations of BJ closed and BJ open stacks, respectively. This corresponds to a more extended domain arrangement than in the BJ open crystal form (141°). The two sets of generated conformations had a substantial overlap (Fig. 5) also seen on the diagrams of interdomain distance and domain twisting angle versus domain bending angle (Fig. 6).

To learn how close the representatives of two sets may approach to each other, we calculated sRMSD for each conformation generated in the BJ closed run against all those of the BJ open run, and vice versa. Some of the conformations from two sets were found to be remarkably close to each other: in all, three pairs had sRMSD less than 3.5 Å. This seems rather close, given that the overall scale of domain rearrangement is measured in several tens of angstroms. In comparisons of one crystal structure to the stack conformations of the other form, the minimal sRMSD was 38.8 Å (BJ closed crystal structure vs the BJ open stack conformations) and 18.2 Å (BJ open crystal structure vs the BJ closed stack conformations). It is interesting to compare these results with the difference between the "closed" or "open" conformations realized in two other PDB structures of the BJ protein dimer, 3bjl and 1bjm crystallized from other solvents.[74] In comparison to the BJ closed form of 2bjl crystal structure, the closed form conformations had sRMSD 9.0 (3bjl) and 5.9 (1bjm) Å. For the BJ open form, corresponding sRMSD was equal to 20.0 (3bjl) and 14.5 (1bjm) Å. These differences may be used as a reference for the structural changes characteristic of the representatives of a particular structural form. Therefore, one may conclude that

TABLE III. Parameters of the Generated Conformations of the BJ Protein*

Parameters	Crystal	Minimal	Maximal
Start from the BJ closed crystal conformation			
Energy, kcal/mol[a]	−2002.5	−2074.2	−2026.9
Interdomain distance Δ, Å	17.4	18.0	30.6
Domain bending angle Θ, °	121	117	171
Domain twisting angle Φ, °	−99	−178	171
sRMSD, Å[b]	0.0	22.9	67.0
Start from the BJ open crystal conformation			
Energy, kcal/mol[a]	−1973.3	−2030.0	−1999.0
Interdomain distance Δ, Å	26.6	18.3	28.9
Domain bending angle Θ, °	141	131	178
Domain twisting angle Φ, °	122	−176	173
sRMSD, Å[b]	0.0	8.6	56.7

*For each parameter the value for the regularized crystal structure and the minimal and maximal values for the stack conformations are shown. Each stack contains 50 conformations.
[a]Energy after relaxation with respect to all standard torsion angles.
[b]sRMSD from the starting crystal conformation.

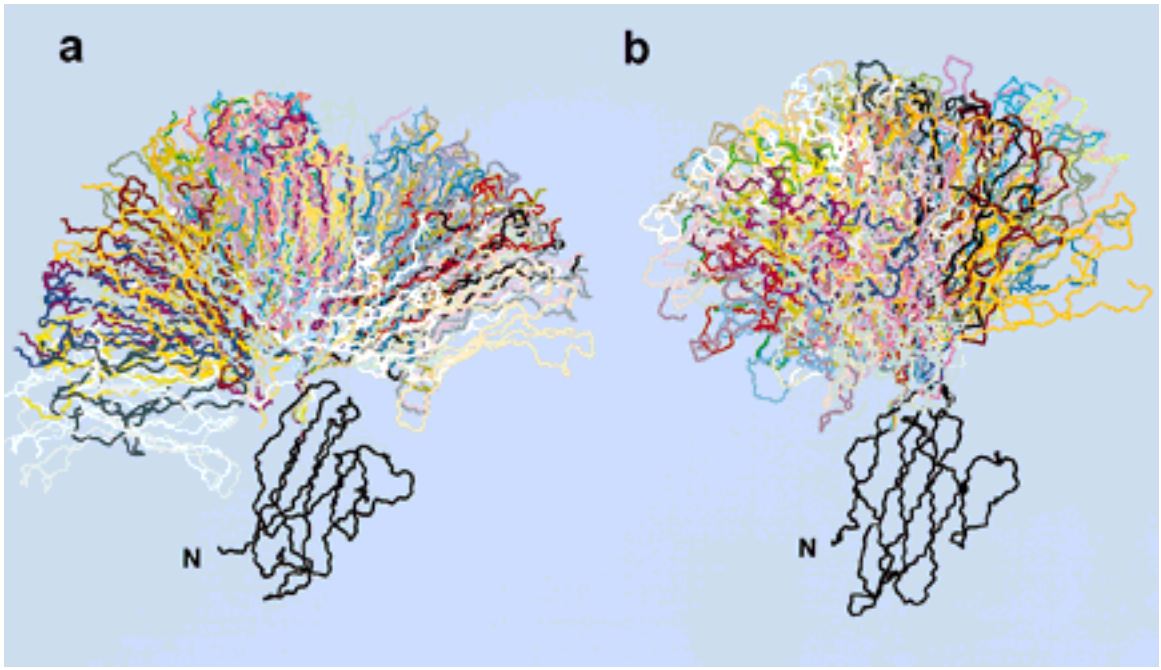

Fig. 5. Backbone display of the BJ protein conformations generated in the calculations starting from the BJ closed (**a**) and the BJ open (**b**) crystal structures. The position of N terminus is indicated.

the conformations generated from the BJ closed form could approach a region of the conformational space occupied by the open forms with an accuracy of 18.2 Å, comparable to the differences found for 3bjl and 1bjm open forms, respectively. The conformation nearest to the open crystal form was only fifth in the energy-sorted list of the stack conformations with a relatively small difference of 13.1 kcal/mol from the lowest energy stack conformation. However, in terms of the parameters of the simplified representation, the conformation was rather different from the open crystal form: 4.5 Å shorter interdomain distance, 22° greater domain bending, and 28° less domain twisting. In the open BJ stack, the conformation nearest to the BJ crystal closed form had sRMSD equal to 38.8 Å. This seems to be beyond the region of the conformational space occupied by the closed-form crystal conformations as estimated by 9.0 Å and 5.9 Å sRMSD. In terms of energy, it was 37th in the energy-sorted list of the stack conformations, with the energy 32.7 kcal/mol higher than the lowest energy conformation.

LAO Protein Calculations

The tether term was included in the energy function to prevent the double-segment domain *A* distortions during energy minimization following the loop closure procedure (see the Methods section). 963 and 1512 conformations were accepted during the LAO closed and the LAO open runs, respectively. The maximal allowed number of conformations, 50, were collected in both conformational stacks by the end of the calculations. Relaxation of the generated conformations was performed with respect to all standard torsion angles by 25,000 steps of iterative local minimization (five iterations of maximum 5000 function calls each) and led to rather small changes measured by RMSD in the range of 0.2 to 0.5 Å.

The results of the calculations are summarized in Table IV. Energies of the generated conformations were in the range of 55.9 kcal/mol (LAO closed run) and 107.7 kcal/mol (LAO open run). Structural diversity of the conformations from the LAO closed stack was quite large (Fig. 7a) with an average sRMSD with respect to the starting crystal structure equal to 14.2 Å. LAO open stack conformations manifested less diversity (Fig. 7b), with an average sRMSD equal to 6.2 Å. Figure 8 shows Θ, Δ, and Φ parameters of the relative domain positions for the conformations of both stacks. Domain bending angle Θ varied from 117° to 151° (LAO closed run) and from 131° to 156° (LAO open run). None of the conformations of both stacks was bent more than the crystal closed form with a domain bending angle of 117°: the minimal domain bending angle observed in the conformations of the LAO open stack, 131°, was 7° less than that in the crystal open form (138°). The domain twisting angle Φ varied from $-107°$ to 49° in the LAO closed stack conformations, and from 11° to 126° in the LAO open stack ones. In the conforma-

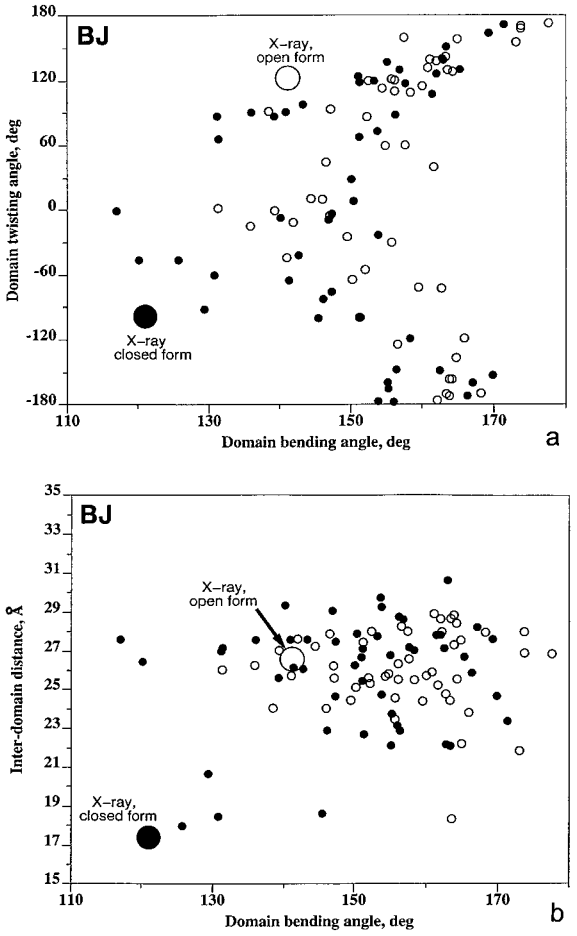

Fig. 6. Variations of the relative domain positions in the generated conformations of the BJ protein: (**a**) domain twisting angle Φ versus domain bending angle Θ; (**b**) interdomain distance Δ versus domain bending angle Θ. Filled and open circles correspond to the BJ closed and the BJ open stack conformations, respectively. Larger circles mark the positions of the corresponding starting structures.

tions generated during the LAO closed run, there was a tendency for the domain twisting angle to decrease when the domain bending angle increased (Fig. 8a). By contrast, in the conformations generated in the LAO open run, the twisting angle increased as bending angle increased. These correlations may be due to the loop closure restraints imposed on the conformations of two strands of the linker. Interdomain distance Δ varied within a range of 5.9 Å (LAO closed run) and 4.9 Å (LAO open run). In general, the distance was larger in the LAO closed stack conformations with less domain bending (Figure 8b).

To quantify the degree of structural overlap between two sets (Figs. 7 and 8), we calculated sRMSD between each conformation of the LAO closed stack and all those from the LAO open stack, and vice versa. Several pairs of the conformations were fairly close to each other: in all, eight pairs had sRMSD

TABLE IV. Parameters of the Generated Conformations of the LAO Protein*

Parameters	Crystal	Minimal	Maximal
Start from the LAO closed crystal conformation			
Energy, kcal/mol[a]	−2589.9	−2629.3	−2573.4
Interdomain distance Δ, Å	20.7	16.2	22.1
Domain bending angle Θ, °	117	117	151
Domain twisting angle Φ, °	28	−107	49
sRMSD, Å[b]	0.0	0.8	23.2
Start from the LAO open crystal conformation			
Energy, kcal/mol[a]	−2609.4	−2629.8	−2522.2
Interdomain distance Δ, Å	20.6	16.9	21.9
Domain bending angle Θ, °	137	131	156
Domain twisting angle Φ, °	57	11	126
sRMSD, Å[b]	0.0	3.1	12.6

*For each parameter the value for the regularized crystal structure and the minimal and maximal values for the stack conformations are shown. Each stack contains 50 conformations.
[a]Energy after relaxation with respect to all standard torsion angles.
[b]sRMSD from the starting crystal conformation.

less than 3.5 Å. In the comparisons between the crystal forms of the protein and the stack conformations generated from the counterpart's crystal form, even closer similarity was found: the 27th conformation from the LAO closed stack was only 2.2 Å sRMSD away from the LAO open crystal form. Such a small sRMSD corresponded to very similar relative domain positions with 0.5 Å interdomain distance difference, and 1° and 6° domain bending and twisting differences, respectively. The energy of this conformation was the highest in the stack and 55.9 kcal/mol greater than the lowest one. In the comparison of closed form crystal structure to the LAO open stack conformations, the conformation with the smallest sRMSD (9.4 Å) had energy 16.2 kcal/mol higher than the lowest energy in the stack. In this case, the interdomain distance was the same, 20.8 Å, but the generated conformation was more open and somewhat twisted compared to the crystal form with the domain bending and twisting angles larger by 13° and 6°, respectively.

DISCUSSION

The main goal of this work was to describe and to test our method for large-scale modeling of protein domain rearrangements in torsion angle space starting from a single known structure. We attempted to implement a fully automated procedure that identifies an interdomain linker consisting of one or two

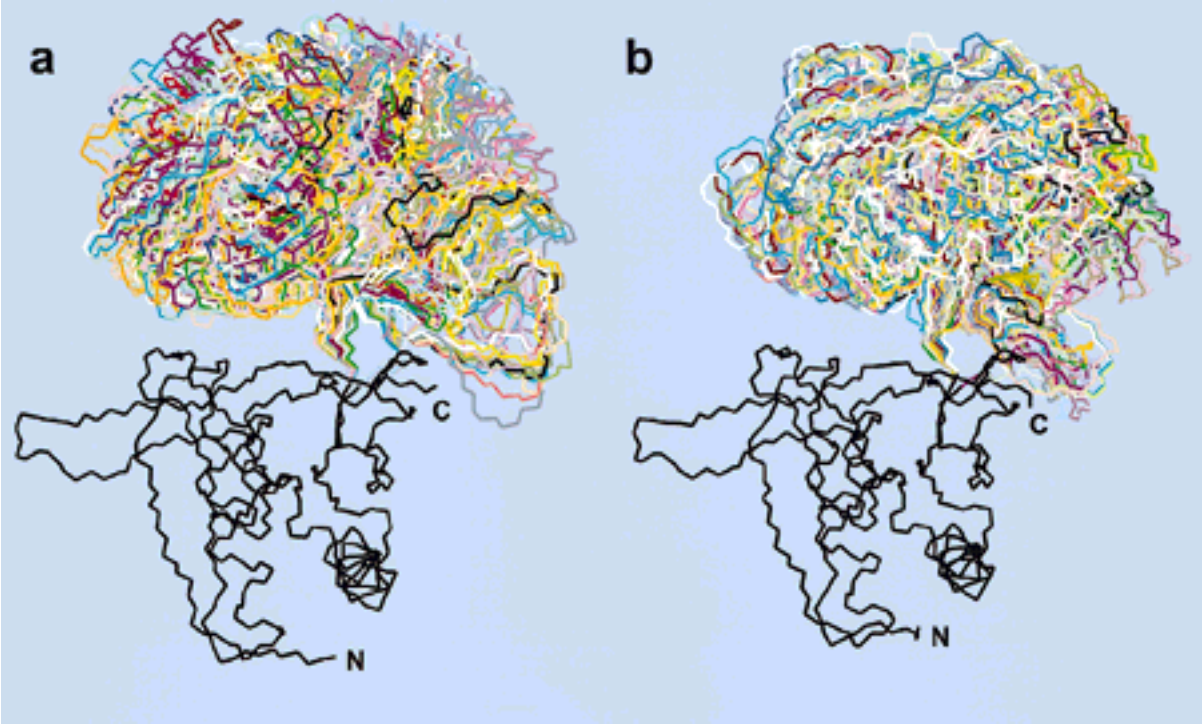

Fig. 7. Backbone display of the LAO protein conformations generated in the calculations starting from the LAO closed (**a**) and the LAO open (**b**) crystal structures. The positions of N and C termini are indicated.

strands, recognizes potentially interacting domain interfaces, and, finally, performs the BPMC sampling calculations generating a set of conformations with ideal covalent geometry and low energies distributed within a narrow interval. The generated conformations should represent possible conformational states of a multidomain protein. The main criterion for success in the test calculations with the BJ and the LAO proteins was defined to be a structural overlap between the two sets of the low-energy conformations with ideal covalent geometry generated from known open and closed forms. It was shown that a wide variety of conformations can be calculated during a single run. It is particularly interesting that one of the generated LAO protein conformations was found close to the vicinity of the conformation of the other crystal form, and no information about the latter was used. The corresponding rearrangement of the starting structure involved complex concerted changes of the conformations of both linker strands resulting in overall 14.5 Å sRMSD displacement of the second domain.

Despite a rather broad distribution and a structural overlap of the generated sets for both test proteins, the conformational space near the closed crystal forms was not sufficiently sampled in the simulations started from the open forms. One of the possible explanations is an internally complex structure of the interdomain interfaces, which plays an important role in a sufficiently closed conformation. The interdomain interfaces include numerous side chains that require accurate positioning, which was not achieved for the time of the calculations. Other explanations are specific for each protein. In case of the BJ protein, the relatively long linker allows substantial freedom for the domain rearrangements. Longer simulations with a series of experimental or knowledge-based restraints could cover the accessible conformational space more extensively and, in principle, should allow an approach to the closed form. In the case of the LAO protein, the crystal closed form is known to exist in the complex of the protein with its peptide ligand, while its crystal uncomplexed form is in the open conformation. Perhaps the closed form conformation could be generated if the ligand molecule is considered in the simulation with the protein structure. In either case, the problem of finding a closed form conformation without imposing specific restraints (due to crystal packing, oligomerization, ligand binding, etc.) appears difficult.

Different domain configurations observed in a crystal are stabilized by additional interactions, for example, with other subunits or ligands. Therefore, for calculations assuming a solvent environment, either finding a single "alternative" conformation or

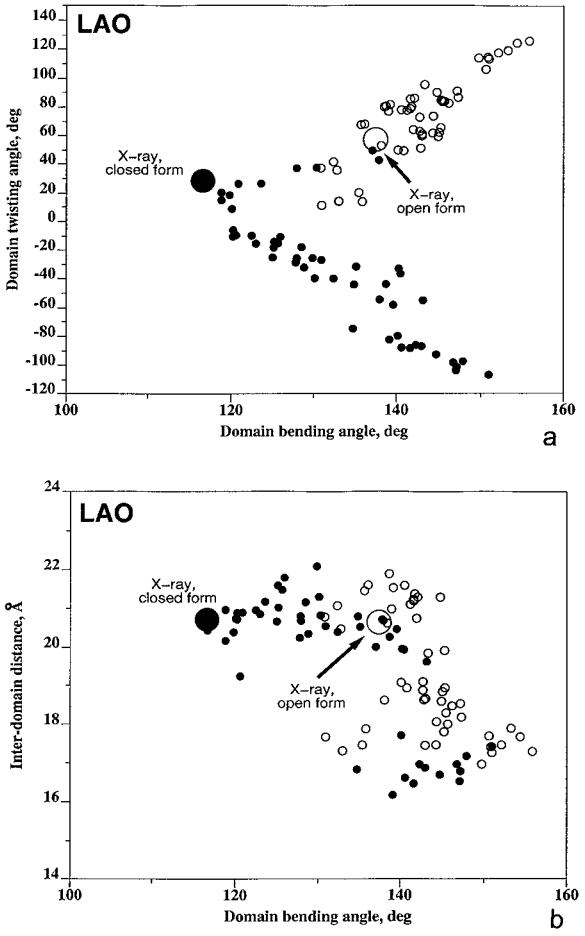

Fig. 8. Variations of the relative domain positions in the generated conformations of the LAO protein. Notations are the same as in Figure 6.

a unique pathway connecting known crystal conformational states is unrealistic. In these cases, an ensemble of states is a legitimate representation of conformational possibilities of a multidomain protein and could be a goal of the calculations.

The "discontinuous" BPMC-like sampling procedure cannot simulate domain rearrangement pathways in the way molecular dynamics does. Yet, an elementary conformational change during a BPMC run involves only several torsion angles, and the subsequent local minimization with respect to all free torsion angles cannot change them drastically. Large structural rearrangements require an accumulation of several more localized concerted conformational changes, which makes a series of the conformations accepted during the run look like a large time scale dynamic trajectory. Therefore, an "animation" of a BPMC trajectory may be helpful in analysis of the conformational degrees of freedom involved in the domain rearrangements. Conformations accepted during the BPMC calculations of the BJ and the LAO proteins were used to produce a molecular animation of the domain rearrangements which is available from World Wide Web site http://saturn.med.nyu.edu/beta/groups/Abagyan.html or on videotape by request to the authors.

The described approach appears to be rather efficient in "shrinking" the conformational space and reducing the whole problem of the domain rearrangements to the sampling of the most essential degrees of freedom. Conformational space of a polypeptide structure grows exponentially with the number of variables, and may become too large for reliable sampling for even a short peptide, while a typical multidomain protein is larger than 150–200 residues. Using torsion angles instead of Cartesian coordinates reduces the number of variables by a factor of seven, but still leaves very large numbers of 1130 and 1311 standard torsions for BJ and LAO proteins, respectively. However, further exclusion of non-essential intra-domain variables reduces the number of free torsions by a factor of 10 to 20 (67 and 66 for the two BJ runs, and 151 and 117 for the two LAO runs), permitting efficient calculation of the energy and its derivatives with respect to the free torsions. A complementary advantageous feature of the BPMC sampling protocol is its ability to make large nonlocal random moves in several torsion angles simultaneously with a high acceptance ratio. For example, in the two BJ runs the ratio was equal to 0.48 and 0.50, and even for the LAO protein calculations where loop closure conditions had to be satisfied, the acceptance ratio was as high as 0.42 (LAO closed run) and 0.46 (LAO open run).

BPMC sampling calculations were performed at an elevated temperature. Bruccoleri and Karplus showed[75] that a temperature of 1500 K was necessary to surmount local energy barriers in a molecular dynamics sampling of the conformational space of antigen-binding loop. After a set of preliminary calculations at different temperatures starting from 300 K, we have chosen the simulation temperature equal to 3000 K. The ideal covalent geometry (bond lengths, bond angles, and phase angles) is automatically preserved in the ICM model, and we were not concerned about "structure explosion" possible in unrestrained cartesian molecular dynamics. However, high temperature leads to an artificially increased scatter of the potential energy in an ensemble. In our case, energy ranges for the retained low-energy stack conformations were 47.3 and 31.0 kcal/mol in the two BJ runs, and 55.9 and 107.6 kcal/mol in the two LAO runs (Tables III and IV). In this regard, the method does not guarantee that exactly the same conformations can be realized at room temperature, and should be considered as a large-scale sampling technique aimed at the generation of a variety of low-energy candidate conformations. These conformations can further be refined with a more thorough local sampling procedures.

An important step intrinsic to the described method is the identification of the linker portion of the polypeptide chain, the problem closely related to the identification of protein domains.[60–68] The main requirement for such an algorithm was to provide a visually reasonable linker/domain assignment consistent with the results of other methods and minimizing the dependence of the assignment on the relative domain positions. In addition to Siddiqui and Barton's analysis of inter- and intradomain contacts,[69] we analyzed intralinker and linker-domain contacts. The optimized split function (Equations A2–A4 in the Appendix) identified almost the same linkers (with one residue accuracy) for remarkably different relative domain positions (Table I). Another valuable feature of the proposed split function is a weak dependence of the results on the contact cutoff distance: the values from 4 to 6 Å lead to an almost identical domain/linker partition. However, it is problematic to get complete independence of the domain interface side-chain torsions list from the relative domain positions. For example, using 5 Å sphere radius for identifying the interdomain interface (see the Methods section, step 2 of the protocol for domain rearrangement modeling) in both closed and open conformations of the LAO protein results in 101 and 67 free side-chain torsion angles, respectively. However, these torsion angle sets seem redundant to account for possible domain rearrangements, and, therefore, the above differences do not affect the results seriously.

We suppose that the combination of (1) sampling in a subspace of essential torsion angles, (2) application of the efficient global sampling techniques which can find low-energy conformations with minimal computational effort, (3) usage of the detailed energy function with solvation and entropy terms, and (4) application of the local deformation algorithm for a double-stranded linker is a promising way to explore the conformational space of large-scale domain rearrangements. It may be useful in a number of situations. First, a quantitative estimate of possible domain rearrangements may be important in protein design, such as in designing an intersubunit linker with given flexibility requirements (e.g., in antibody design). Second, knowledge about the sterically allowed range of domain movements may be useful for testing biological hypotheses about the principal possibility of interaction between a multidomain protein and a complex receptor with two or more specifically spaced binding sites. Third, the described methodology may be combined with a docking procedure, if ligand binding or protein-protein interactions induce the domain rearrangements.

ACKNOWLEDGMENTS

We are indebted to Maxim Totrov for numerous stimulating discussions and to the referees for helpful suggestions and comments.

APPENDIX: INTERDOMAIN LINKER IDENTIFICATION

We modified Siddiqui and Barton's method[69] for identification of interdomain linkers. These authors suggested that two amino acid residues are in contact if any two nonhydrogen atoms belonging to the residues are within a certain distance (5 Å). Our first modification was taking into account a strength (w) of the contact in the model of contiguous domains (Fig. 1a,b) which we defined as the number of pairs of nonhydrogen atoms belonging to the two given residues. Then, split value ρ can be attributed to each probe ith residue position dividing the chain into two pieces: residues 1 to $i-1$ (probe domain A) and i to N_{res} (probe domain B), $1 < i < N_{\text{res}}$:

$$\rho(i) = \frac{\sum_{AA} w_{AA}(i) \times \sum_{BB} w_{BB}(i)}{\sum_{AB} w_{AB}^2(i)} \qquad \text{(A1)}$$

where indices AA, BB, and AB correspond to summation over the lists of intra- and interdomain contacts, respectively, in the probe partition between the domains, and N_{res} is the number of residues in the structure. The assignment of the domains is naturally defined by a residue where the split value is maximal. If one of the domains includes two segments (Fig. 1b), then the split function should be searched for in two dimensions of all possible pairs of ith and jth residues, $1 < i < j < N_{\text{res}}$, separating the domains:

$$\rho(i, j) = \frac{\sum_{AA} w_{AA}(i, j) \times \sum_{BB} w_{BB}(i, j)}{\sum_{AB} w_{AB}^2(i, j)} \qquad \text{(A2)}$$

and the domain assignment should be 1 to $i-1$, $j+1$ to N_{res} (domain A) and i to j (domain B) corresponding to the maximal value of split function ρ.

The following modifications are necessary if one allows a linker to separate the domains. The whole polypeptide chain is assumed to be split between the domain parts and the linker part(s), and two cases should be considered (Fig. 1c,d). If both domains are single-segment, then domain positions are unambiguously defined by the two linker residues i_1 and i_2 (Fig. 1c). Analysis of a series of two-domain protein structures was performed. It was found that the better split function form differs from the simple one given by Equation (A1). We introduced a modified form of the split function combining all six types of intra- and interdomain, and intra-linker and domain-linker contacts to take into account the redistribution of the contacts between the domains and the linker as compared to the model of the contiguous domains (Fig. 1a):

$$\rho(i_1, i_2) = \left(\frac{\sum_{AA} w_{AA}(i_1, i_2) \times \sum_{BB} w_{BB}(i_1, i_2)}{\sum_{AB} w^2_{AB}(i_1, i_2)} \right)^2 \times \frac{\sum_{LL} w_{LL}(i_1, i_2)}{\sum_{AL} w_{AL}(i_1, i_2) \times \sum_{BL} w_{BL}(i_1, i_2)}. \quad \text{(A3)}$$

Subscripts AL, BL, LL refer to the domain-linker (AL, BL) and intralinker (LL) contacts, respectively. If one of the domain includes two segments (domain A, Fig. 1d), then the split value is a function of the positions of four residues i_1, i_2, j_1, j_2:

$$\rho(i_1, i_2, j_1, j_2) = \left(\frac{\sum_{AA} w_{AA}(i_1, j_2) \times \sum_{BB} w_{BB}(i_2, j_1)}{\sum_{AB} w_{AB}(i_1, i_2, j_1, j_2)} \right)^2 \times \frac{\sum_{LL} w_{LL}(i_1, i_2, j_1, j_2)}{\sum_{AL} w_{AL}(i_1, i_2, j_1, j_2) \times \sum_{AL} w_{BL}(i_1, i_2, j_1, j_2)}. \quad \text{(A4)}$$

The resulting set of two (Fig. 1c) or four (Fig. 1d) residues corresponding to the maximal split value defines our domain/linker assignment. For protein structures with a single-stranded linker the domain/linker assignment was performed according to Equation (A2). In case of a double-stranded linker, a two-step procedure was applied to avoid a time consuming search in four-dimensional space of residue positions i_1, i_2, j_1, j_2. First, the positions of residues i_0 and j_0 representing approximate positions of residues i_1, i_2 and j_1, j_2, respectively, were identified using the contiguous domain model (Fig. 1b) according to Equation (A2). Second, the search for the exact positions of residues i_1, i_2 and j_1, j_2 in four dimensions (Fig. 1d) was performed in a broad (15-residue) vicinity of residues i_0 and j_0 by exhaustive enumeration of all possible combinations according to Equation (A4).

REFERENCES

1. Janin, J., Wodak, S.J. Structural domains in proteins and their role in the dynamics of protein function. Prog. Biophys. Mol. Biol. 42:21–78, 1983.
2. Bennett, W.S., Huber, R. Structural and functional aspects of domain motions in proteins. Crit. Rev. Biochem. 15:291–384, 1984.
3. Huber, R. Flexibility and rigidity, requirements of the function of proteins and protein pigment complexes. Biochem. Soc. Trans. 15:1009–1020, 1987.
4. Schulz, G.E. Domain motions in proteins. Curr. Opin. Struct. Biol. 1:883–888, 1991.
5. Gerstein, M., Lesk, A.M., Chothia, C. Structural mechanisms for domain movements in proteins. Biochemistry 33:6739–7649, 1994.
6. Abola, E.E., Bernstein, E.C., Bryant, S.H., Koetzle, T.F., Weng, J. Protein data bank in crystallographic databases: information content, software systems, scientific applications. Data Commission Int Union Crystallogr. 107– 132, 1987.
7. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M. The protein data bank: A computer-based archival file for macromolecular structures. J. Mol. Biol. 112:535–542, 1977.
8. Schiffer, M., Ainsworth, C., Xu, Z.-B., Carperos, W., Olsen, K., Solomon, A., Stevens, F.J., Chang, C.-H. Structure of a second crystal form of Bence-Jones protein Loc: Strikingly different domain associations in two crystal forms of a single protein. Biochemistry 28:4066–4072, 1989.
9. Faber, H.R., Matthews, B.W. A mutant T4 lysozyme displays five different crystal conformations. Nature 348:263–266, 1990.
10. Dixon, M.M., Nicholson, H., Shewchuk, L., Baase, W.A., Matthews, B.W. Structure of a hinge-bending bacteriophage T4 lysozyme mutant, Ile3 → Prog. J. Mol. Biol. 227:917–933, 1992.
11. Holland, D.R., Tronrud, D.E., Pley, H.W., Flaherty, K.M., Stark, W., Jansonius, J.N., McKay, D.B., Matthews, B.W. Structural comparison suggests that thermolysin and related neural proteases undergo hinge-bending motion during catalysis. Biochemistry 31:11310–11316, 1992.
12. Sali, A., Veerapandian, B., Cooper, J.B., Moss, D.S., Hofmann, T., Blundell, T.L. Domain flexibility in aspartic proteinases. Proteins 12:158–170, 1992.
13. Quiocho, F.A. Atomic structures of periplasmic binding proteins and the high-affinity active transport systems in bacteria. Phil. Trans. R. Soc. Lond. [Biol.] 326:341–351, 1990.
14. Ravichandran, K.G., Boddupalli, S.S., Hasemann, C.A., Peterson, J.A., Deisenhofer, J. Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's. Science 261:731–736, 1993.
15. Vonrhein, C., Schlauderer, G.J., Schulz, G.E. Movie of the structural changes during a catalytic cycle of nucleoside monophosphate kinases. Structure 3:483–490, 1995.
16. Gerstein, M., Schulz, G., Chothia, C. Domain closure in adenylate kinase: Joints on either side of two helices close like neighboring fingers. J. Mol. Biol. 229:494–501, 1993.
17. Gerstein, M., Anderson, B.F., Norris, G.E., Baker, E.N., Lesk, A.M., Chothia, C. Domain closure in lactoferrin: Two hinges produce a see-saw motion between alternative close-packed interfaces. J. Mol. Biol. 234:357–372, 1994.
18. Lesk, A.M., Chothia, C. Mechanisms of domain closure in proteins. J. Mol. Biol. 174:175–191, 1984.
19. Boutonnett, N.S., Rooman, M.J., Wodak, S.J. Automatic analysis of protein conformational changes by multiple linkage clustering. J. Mol. Biol. 253:633–647, 1995.
20. McCammon, J.A., Gelin, B.R., Karplus, M., Wolynes, P.G. The hinge-bending mode in lysozyme. Nature 262:325–326, 1976.
21. Mao, B., McCammon, J.A. Structural study of hinge bending in L-arabinose-binding protein. J. Biol. Chem. 59:964–4970, 1984.
22. Colonna-Cesari, F., Perahia, D., Karplus, M., Eklund, H., Braden, C.I., Tapia, O. Interdomain motion in liver alcohol dehydrogenase: Structural and energetic analysis of the hinge bending mode. J. Biol. Chem. 261:15273–15280, 1986.
23. Arnold, G.E., Ornstein, R.L. A molecular dynamics simulation of bacteriophage T4 lysozyme. Protein Eng. 5:703–714, 1992.
24. Arnold, G.E., Manchester, J.I., Townsend, B.D., Ornstein, R.L. Investigation of domain motions in bacteriophage T4 lysozyme. J. Biomol. Struct. Dyn. 12:457–474, 1994.
25. Paulsen, M.D., Ornstein, R.L. Dramatic differences in the motions of the mouth of open and closed cytochrome P450BM-3 by molecular dynamics simulations. Proteins 21:237–243, 1995.
26. Miaskiewicz, K., Ornstein, R.L. DNA binding by TATA-box binding protein (TBP): A molecular dynamics computational study. J. Biomol. Struct. Dyn. 13:593–600, 1996.
27. Mehler, E.L., Pascual-Ahuir, J.L., Weinstein, H. Structural dynamics of calmodulin and troponin C. Protein Eng. 4:625–637, 1991.
28. Gulati, J., Akella, A.B., Su, H., Mehler, E.L., Weinstein, H. Functional role of arginine-11 in the N-terminal helix of skeletal troponin C: Combined mutagenesis and molecular dynamics investigation. Biochemistry 34:7348–7355, 1995.
29. Elamrani, S., Berry, M.B., Phillips Jr., G.N., McCammon,

J.A. Study of global motions in proteins by weighted masses molecular dynamics: Adenilate kinase as a test case. Proteins 25:79–88, 1996.
30. Kottlam, J., Case, D.A. Langevin modes of macromolecules: Applications to crambin and DNA haxamers. Biopolymers 29:1409–1421, 1990.
31. Mazur, A.K., Abagyan, R.A. New methodology for computer-aided modelling of biomolecular structure and dynamics. 1. Non-cyclic structures. J. Biomol. Struct. Dyn. 6:815–832, 1989.
32. Mazur, A.K., Dorofeyef, V.E., Abagyan, R.A. Derivation and testing of explicit equations of motion for polymers described by internal coordinates. J. Comp. Phys. 92:261–272, 1991.
33. Dorofeyef, V.E., Mazur, A.K. Investigation of conformational equilibrium of polypeptides by internal coordinate stochastic dynamics. J. Biomol. Struct. Dyn. 10:143–167, 1993.
34. Mathiowetz, A.M., Jain, A., Karasawa, N., Goddar III, W.A. Protein simulations using techniques suitable for very large systems: The cell multipole method for nonbond interactions and the Newton-Euler inverse mass operator method for internal coordinate dynamics. Proteins 20:227–247, 1994.
35. Head-Gordon, T., Brooks III, C.L. Virtual rigid body dynamics. Biopolymers 31:77–100, 1991.
36. Wade, R.C., Luty, B.A., Demchuk, E., Madura, J.D., Davis, M.E., Briggs, J.M., McCammon, J.A. Simulation of enzyme-substrate encounter with gated active sites. Nat. Struct. Biol. 1:65–69, 1994.
37. Noguti, T., Gõ, N. Efficient Monte Carlo method for simulation of fluctuating conformations of native proteins. Biopolymers 24:527–546, 1985.
38. Amadei, A., Linssen, A.B.M., Berendsen, H.J.C. Essential dynamics of proteins. Proteins 17:412–425, 1993.
39. van Aalten, D.M.F., Amadei, A., Linssen, A.B.M., Eijsink, V.G.H., Vriend, G., Berendsen, H.J.C. The essential dynamics of thermolysin: Comparison of simulations in vacuum and water. Proteins 22:43–54, 1993.
40. Abagyan, R.A., Totrov, M.M., Kuznetsov, D.A. ICM: A new method for protein modeling and design: Applications to docking and structure prediction from the disordered native conformation. J. Comp. Chem. 15:488–506, 1994.
41. Abagyan, R.A., Totrov, M.M. Biased probability Monte Carlo conformational search and electrostatic calculations for peptide and proteins. J. Mol. Biol. 235:983–1002, 1994.
42. Borchert, T.V., Abagyan, R.A., Kishan, K.V.R., Zeelen, J.P., Wierenga, R.K. The crystal structure of an engineered monomeric triosephosphate isomerase, monoTIM: The correct modeling of an eight-residue loop. Structure 1:205–213, 1993.
43. Eisenmenger, F., Argos, P., Abagyan, R.A. A method to configure protein side-chains from the main-chain trace in homology modeling. J. Mol. Biol. 231:849–860, 1993.
44. Totrov, M.M., Abagyan, R.A. Detailed ab initio prediction of lysozyme-antibody complex with 1.6 A accuracy. Nature Struct. Biol. 1:259–263, 1994.
45. Abagyan, R.A., Frishman, D., Argos, P. Recognition of distantly related proteins through energy calculations. Proteins 19:132–140, 1994.
46. Cardozo, T., Totrov, M.M., Abagyan, R.A. Homology modeling by the ICM method. Proteins 23:403–414, 1995.
47. Oh, B.-H., Pandit, J., Kang, C.-H., Nikaido, K., Gokcen, S., Ames, G.F.-L., Kim, S.-H. Three-dimensional structures of the periplasmic lysine/arginine/ornithine-binding protein with and without a ligand. J. Biol. Chem. 268:11348–11355, 1993.
48. Momany, F.A., McGuire, R.F., Burgess, A.W., Scheraga, H.A. Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. J. Phys. Chem. 79:2361–2381, 1975.
49. Nemethy, G., Pottle, M.S., Scheraga, H.A. Energy parameters in polypeptides. 9. Updating of geometric parameters, nonbonded interactions and hydrogen bond interactions for the naturally occurring amino acids. J. Phys. Chem. 87:1883–1887, 1983.
50. Nemethy, G., Gibson, K.D., Palmer, K.A., Yoon, C.N., Paterlini, G., Zagari, A., Rumsey, S., Scheraga, H.A. Energy parameters in polypeptides. 10. Improved geometric parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides. J. Phys. Chem. 96:6472–6484, 1992.
51. Wesson, L., Eisenberg, D. Atomic solvation parameters applied to molecular dynamics of proteins in solution. Protein Sci. 1:227–235, 1992.
52. Li, Z., Scheraga, H.A. Monte Carlo-minimization approach to the multiple-minima problem in protein folding. Proc. Natl. Acad. Sci. USA 84:6611–6615, 1987.
53. Metropolis, N.A., Rosenbluth, A.W., Rosenbluth, N.M., Teller, A.H., Teller, E. Equation of state calculation by fast computing machines. J. Chem. Phys. 21:1087–1092, 1953.
54. Williams, R.L., Vila, J., Perrot, G., Scheraga, H.A. Empirical solvation models in the context of conformational energy searches: Application to bovine pancreatic trypsin inhibitor. Proteins 14:110–119, 1992.
55. Caflisch, A., Niederer, P., Anliker, M. Monte Carlo minimization with thermalization for global optimization of polypeptide conformations in cartesian coordinate space. Proteins 14:110–119, 1992.
56. Zhorov, B.S., Ananthabarayanan, V.S. Similarity of Ca^{2+}-bound conformations of morphine and Met-enkephalin: a computational study. FEBS Lett. 354:131–134, 1994.
57. Kato, K., Lian, L.Y., Barsukov, I.L., Derrick, J.P., Kim, H., Tanaka, R., Yoshino, A., Shiraishi, M., Shimada, I., Arata, Y., Roberts, G.C.K. Model for the complex between protein G and an antibody Fc fragment in solution. Structure 3:79–85, 1995.
58. Abagyan, R.A., Argos, P. Optimal protocol and trajectory visualization for conformational searches of peptides and proteins. J. Mol. Biol. 225:519–532, 1992.
59. Lee, B., Richards, F.M. The interpretation of protein structures: Estimation of static accessibility. J. Mol. Biol. 55:379–400, 1971.
60. Crippen, G.M. The tree structural organization of proteins. J. Mol. Biol. 126:315–321, 1978.
61. Rashin, A.A. Location of domains in globular proteins. Nature 291:85–86, 1981.
62. Wodak, S.J., Janin, J. Location of structural domains in proteins. Biochemistry 20:6544–6552, 1981.
63. Lesk, A.M., Rose, G.D. Folding units in globular proteins. Proc. Natl. Acad. Sci. USA 78:4304–4308, 1981.
64. Zehfus, M.H., Rose, G.D. Compact units in proteins. Biochemistry 25:5759–5765, 1986.
65. Holm, L., Sander, C. Parser for protein folding units. Proteins 19:256–268, 1994.
66. Sowdhamini, R., Blundell, T.L. An automatic method involving cluster analysis of secondary structure for the identification of domains in proteins. Protein Sci. 4:506–520, 1995.
67. Swindells, M.B. A procedure for detecting structural domains in proteins. Protein Sci. 4:103–112, 1995.
68. Islam, S.A., Luo, J., Sternberg, M.J.E. Identification and analysis of domains in proteins. Protein Eng. 8:513–525, 1995.
69. Siddiqui, A.S., Barton, G.J. Continuous and discontinuous domains: An algorithm for the automatic generation of reliable protein domain definition. Protein Sci. 4:872–884, 1995.
70. Gõ, N., Scheraga, H.A. Ring closure and local conformational deformations of chain molecules. Macromolecules 3:178–187, 1970.
71. Abagyan, R.A., Mazur, A.K. New methodology for computer-aided modelling of biomolecular structure and dynamics. 2. Local deformations and cycles. J. Biomol. Struct. Dyn. 6:833–845, 1989.
72. Browner, M.F., Fauman, E.B., Fletterick, R.J. Tracking conformational states in allosteric transitions of phosphorylase. Biochemistry 31:11297–11304, 1992.
73. Zhang, X.-J., Wozniak, J.A., Matthews, B.W. Protein flexibility and adaptability seen in 25 crystal forms of T4 lysozyme. J. Mol. Biol. 250:527–552, 1995.
74. Chang, C.-H., Short, M.T., Westholm, F.A., Stevens, F.J., Wang, B.-C., Furey, W., Solomon, A., Schiffer, M. Novel arrangement of immunoglobulin variable domains: X-ray crystallographic analysis of the lambda-chain dimer Bence-Jones protein Loc. Biochemistry 24:4890–4898, 1985.
75. Bruccoleri, R.E., Karplus, M. Conformational sampling using high-temperature molecular dynamics. Biopolymers 29:1847–1862, 1990.