

Coarse-Grained Simulations of Conformational Dynamics of Proteins: Application to Apomyoglobin

T. Haliloglu^{1,2} and I. Bahar^{1,2*}

¹Polymer Research Center and Chemical Engineering Department, Bogazici University, Istanbul, Turkey

²TUBITAK Advanced Polymeric Materials Research Center, Bebek, Istanbul, Turkey

ABSTRACT A coarse-grained dynamic Monte Carlo method is proposed for investigating the conformational dynamics of proteins. Each residue is represented by two interaction sites, one at the α -carbon, and the other on the amino acid sidechain. Geometry and energy parameters extracted from databank structures are used. The method is applied to the crystal structure of apomyoglobin (apo-Mb). Equilibrium and dynamic properties of apo-Mb are characterized within computation times one order of magnitude shorter than conventional molecular dynamics (MD) simulations. The calculated rms fluctuations in α -carbons are in good agreement with crystallographic temperature factors. Regions exhibiting enhanced conformational mobilities are identified. Among the loops connecting the eight helices A to H, the loop CD undergoes the fastest motions, leading to partial unwinding of helix D. Helix G is the most stable helix on the basis of the kinetic stability of dihedral angles, followed by the respective helices A, E, H, and B. These results, in agreement with H/D exchange and two-dimensional NMR experiments, as well as with MD simulations, lend support to the use of the proposed approach as an efficient, yet physically plausible, means of characterizing protein conformational dynamics. *Proteins* 31:271–281, 1998. © 1998 Wiley-Liss, Inc.

Key words: low resolution models; knowledge-based potentials; unfolding kinetics; helix unwinding; cooperative motions; dynamic Monte Carlo; correlations between atomic fluctuations; virtual bond rotations

INTRODUCTION

It is widely recognized that low-resolution models and methods presently constitute the only means of exploring motions of the order of nanoseconds to milliseconds in proteins on a computer. Motions of this time scale are generally *collective* conformational changes. These include: (1) the correlated fluctuations between sequentially distant but topologically coupled residues, (2) the spatial reorganiza-

tion and/or structural transition of secondary structure units, (3) the cooperative changes in tertiary contacts, and (4) the larger-scale motions such as domain movements, all of which contribute to the eventual disruption of the folded structure. Conventional molecular dynamics (MD) simulations with full atomic coordinates and explicit hydration are feasible either for short polypeptides, or for the near-native fluctuations at early stages of unfolding, at the cost of excessive computational times. It is clear that the reduction in variables, upon resort to a low-resolution model, would permit us to explore much more efficiently the collective motions.

However, there is a loss in accuracy brought about by adoption of low-resolution approaches, which has somehow discouraged the use of coarse-grained simulations for investigating the collective dynamics of proteins. The accuracy loss is twofold. First, the chain geometry is distorted to some extent compared to real atomic models. This discrepancy may be more pronounced at regular structures such as α -helices and β -strands. Second, the interaction potentials between the unified sites of the low-resolution model are not precise enough to permit an accurate evaluation of the conformational energetics.

With the increasing number of X-ray or NMR-elucidated protein structures, however, there has been an explosion in the number of studies aiming at extracting structure-derived potential functions and geometry parameters,¹ which may be utilized to model and simulate proteins on a coarse-grained scale. After the original single-site-per-residue model and potentials,^{2,3} more elaborate potentials, including the distance^{4,5} and direction^{6–8} dependence of nonbonded interactions, as well as the torsional preferences of backbone bonds,^{9,10} have been developed.

In general, the low-resolution models and parameters have been used for exploring the equilibrium properties of proteins. For example, they have proven useful in discriminating between correct and incor-

Contract grant sponsor: Bogazici University Research Funds Project; Contract grant number: 97A050.

*Correspondence to: I. Bahar, Polymer Research Center and Chemical Engineering Department, Bogazici University, Bebek 80815, Istanbul, Turkey. E-mail: bahar@prc.bme.boun.edu.tr

Received 27 June 1997; Accepted 20 October 1997

rect folds in threading experiments,^{11–13} in predicting the effects of amino acid substitutions on stability,¹⁴ and in recognizing peptide binding sites on proteins.¹⁵ On the other hand, a smaller number of studies have aimed at exploring the dynamics of proteins using low-resolution approaches, after the pioneering work of Levitt and Warshel.¹⁶ Most of the low-resolution investigations of protein dynamics using knowledge-based potentials have been in the form of on-lattice simulations,^{17–23} as recently reviewed.²⁴ Current studies indicate, however, the utility and comparable efficiency of off-lattice simulations.^{24–28}

In the present study, a recently developed coarse-grained model and its energy parameters^{5,10} are used in an efficient off-lattice Monte Carlo/Metropolis simulation for exploring the collective dynamics of apomyoglobin (apo-Mb). Myoglobin (Mb) has proven to be a tractable protein for studying folding intermediates or molten globular forms. This is a single-domain protein consisting of eight α -helices A–H. Removing the heme group reduces its helical content from 78% to 55% at neutral pH.²⁹ Upon lowering the pH, apo-Mb unfolds in two stages: first a compact intermediate (I) with $\sim 35\%$ helix content forms, and then a more fully denatured state with little residual helix results.³⁰ Structural studies of apo-Mb with different ligands than heme have been reported, the heme pocket being an excellent binding site for porphyrin or tetrapyrrole ligands. Here, the coordinates of apo-Mb will be taken from a biliverdin–apo-Mb complex structure determined to 1.4-Å resolution.³¹ The stability and kinetics of apo-Mb have been thoroughly examined in experiments^{30,32–36} and MD simulations.^{37–39} Comparison with previous studies will permit us to judge the applicability and effectiveness of the present low-resolution approach.

Our recent calculations,^{40,41} based on a highly simplified Gaussian network model of interresidue interactions in the folded state, showed that a satisfactory agreement with experiments is obtainable, provided that the collective behavior or the coupling of all residues in contact (within 7-Å distance) is accounted for in the analytical treatment. This suggests that the adoption of detailed models and methods may be unnecessary for unraveling certain characteristics of proteins. Besides, a recent analysis²⁶ demonstrates that beyond a certain point, added complexity brings little improvement to accuracy.

The present coarse-grained approach is significantly more detailed than the recent Gaussian network model with a single-parameter harmonic potential,^{40,41} in that residue-specific long-range and short-range potentials are used in a two-sites-per-residue virtual bond model. However, it is certainly much simpler and faster compared to full atomic models. Thus, an attempt is made for assessing how far one can proceed in characterizing the equilibrium and dynamic properties of a protein by analyzing the

coarse-grained trajectories generated with a low-resolution model. It will be shown that: (1) the rms fluctuations in backbone atoms observed in simulations are in excellent agreement with those deduced from X-ray experiments, and (2) the early unwinding helices or helical segments are consistent with results from H/D exchange and NMR experiments, as well as MD simulations. Additionally, new information on the structural characteristics of the investigated protein will be obtained, such as the residue pairs and the helix pairs that exhibit the strongest cross-correlations, and the fact that secondary structural element, helix G in the case of apo-Mb, plays a critical role in maintaining the integrity of the tertiary structure. Finally, information on the mechanism of polypeptide bond rotations cooperatively accommodating local structural perturbations and thus preserving the regularity of the structure in compact globular proteins will be extracted.

MODEL AND METHOD

The atomic coordinates of apo-Mb for simulations were obtained from the X-ray structure of biliverdin–apo-Mb complex³¹ deposited in the Brookhaven Protein Data Bank (PDB).⁴² The coordinates of the biliverdin molecule and crystallographic waters were deleted. The biliverdin chromophore is reported to bind within the heme pocket of apo-Mb in a conformation mimicking a heme molecule and thus stabilizing the protein.³¹ The simulation of the unliganded form is thus expected to yield information on the early unfolding kinetics of apo-Mb. The ribbon representation of the structure is shown in Figure 1. This structure has the same helical content (78%) as the native Mb.

The atomic representation is reduced to the following simplified model in present simulations. Each residue i is represented by two sites, its C^α atom and a sidechain interaction center, S_i , selected on the basis of the specific properties of the particular amino acid.⁸ The conformation of the backbone for a protein of N residues is thus defined by $3N - 6$ variables: $N - 1$ virtual bond lengths, l_i , connecting the α -carbons $i - 1$ and i ; $N - 2$ virtual bond angles τ_i at the i^{th} α -carbon; and $N - 3$ dihedral angles ϕ_i describing the torsional rotation of the bond l_i . A schematic representation of the model is given in Figure 2. A given backbone conformation is defined by the set of generalized variables $\{\Phi\} = \{l_2, l_3, \dots, l_N, \theta_2, \theta_3, \dots, \theta_{N-1}, \phi_3, \phi_4, \dots, \phi_{N-1}\}$. The conformation of the sidechains with respect to the backbone is defined in terms of the variables l_i^s , θ_i^s , and ϕ_i^s , which refer to the respective virtual bond length, bond angle, and dihedral angle of the sidechain. ϕ_i^s is the torsional angle defined by the three consecutive bonds l_{i-1} , l_i , and l_i^s .

The energy, $E[\Phi]$, of conformation, $\{\Phi\}$, is found from the additive contribution of two types of interactions: (1) the long-range (LR) interactions between

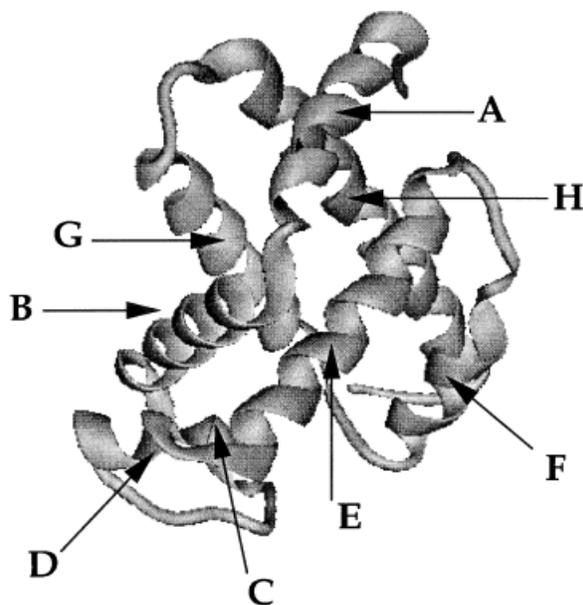


Fig. 1. Ribbon diagram of apomyoglobin (apo-Mb) in a biliverdin–apo-Mb complex determined to a resolution of 1.5 Å (PDB code: 1bvc).³¹ This is a single-domain α -protein composed of eight helices (A to H). Residues in helices are: A, Ser3–Glu18; B, Asp20–Ser35; C, His36–Lys42; D, Thr51–Ala57; E, Ser58–Lys77; F, Leu86–Thr95; G, Pro100–Arg118; and H, Gly124–Leu149.

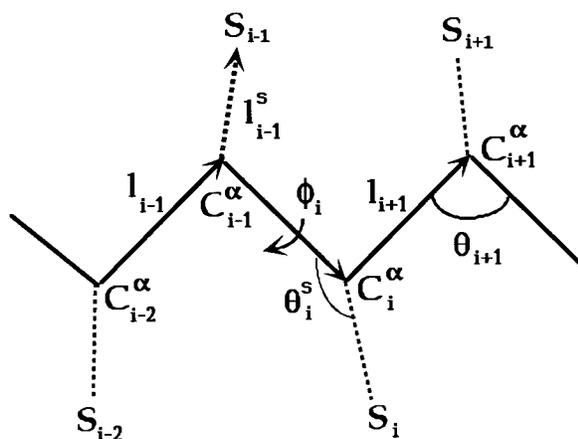


Fig. 2. Schematic representation of the virtual bond model. A protein segment between backbone units C_{i-2}^{α} and C_{i+1}^{α} is shown. Sidechain attached to the i^{th} α -carbon is designated as S_i . l_i is the i^{th} virtual bond extending from C_{i-1}^{α} to C_i^{α} . ϕ_i is the rotational angle of the bond l_i , defined by the respective locations of the four backbone units C_{i-2}^{α} , C_{i-1}^{α} , C_i^{α} , and C_{i+1}^{α} . θ_i is the bond angle between bonds l_i and l_{i+1} .

residue pairs distant along the chain sequence but close in space, which include sidechain–sidechain (S–S), sidechain–backbone (S–B), and backbone–backbone (B–B) interactions between S or B sites separated by at least five intervening virtual bonds, and (2) short-range (SR) interactions, associated with the nearest covalently bonded units. Accord-

ingly, $E[\Phi]$ is expressed as

$$E[\Phi] = E_{\text{LR}}[\Phi] + E_{\text{SR}}[\Phi] \quad (1)$$

where $E_{\text{LR}}[\Phi]$ is the long-range potential given by⁵

$$E_{\text{LR}}[\Phi] = \sum_{i=1}^{N-3} \sum_{j=i+3}^N W_{\text{SS}}(r_{ij}) + \sum_{i=1}^{N-4} \sum_{j=i+4}^N W_{\text{SB}}(r_{ij}) + \sum_{i=1}^{N-5} \sum_{j=i+5}^N W_{\text{BB}}(r_{ij}). \quad (2)$$

Here r_{ij} is the distance between sites i and j in conformation $\{\Phi\}$; $W_{\text{SS}}(r_{ij})$ is the potential between sidechains i and j ; $W_{\text{SB}}(r_{ij})$ is the potential between the S and B sites of the i^{th} and j^{th} residues; and $W_{\text{BB}}(r_{ij})$ is the one between the two backbone units C_i^{α} and C_j^{α} . The residue-specific parameters for all distance-dependent pairwise potentials at 2.0-Å intervals are extracted⁵ from protein databank structures. The reader is referred to Reference 5 for details of parameterization of these empirical long-range potentials.

The short-range conformational energy, $E_{\text{SR}}[\Phi]$, on the other hand, is evaluated from¹⁰

$$E_{\text{SR}}[\Phi] = \sum_{i=2}^N E(l_i) + \sum_{i=2}^{N-1} E(\theta_i) + \sum_{i=2}^{N-1} [E(\phi_i^-)/2 + E(\phi_{i-1}^+)/2 + \Delta E(\phi_i^-, \phi_i^+)] + \sum_{i=2}^{N-1} [\Delta E(\theta_i, \phi_i^-) + \Delta E(\theta_i, \phi_i^+)] \quad (3)$$

where the first summation is the potential associated with the stretching of virtual $C^{\alpha} - C^{\alpha}$ bonds, approximated by a stiff harmonic potential with a force constant of $10 \text{ RT}/\text{\AA}^2$; the second summation refers to the distortion of bond angles; the third represents the contribution of bond torsions; ϕ_i^- and ϕ_i^+ refer to the rotational angles of the virtual bonds preceding and succeeding the i^{th} α -carbon, respectively. The last term in this summation, and the two terms in the last summation, account for the pairwise couplings of bond torsion and/or bond angle distortion degrees of freedom. In our recent analysis¹⁰ of databank structures, the importance of these pairwise couplings are emphasized with application to several proteins. The contributions to $E_{\text{SR}}[\Phi]$ from these couplings are shown to be quantitatively of the same order of magnitude as those associated with independent torsion and bond angle distortion energies, which are represented by the second summation and the first two terms of the third summation in Equation (3). The complete set of LR and SR potentials for all residue types is available on the Internet (<http://www.boun.edu.tr/research/polymer>).

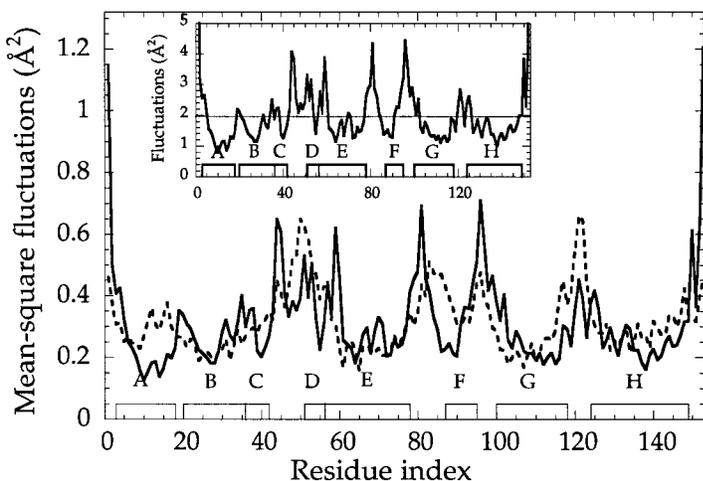


Fig. 3. Mean-square fluctuations $\langle \Delta R_i^2 \rangle$ in α -carbon positions as a function of residue numbers for apo-Mb. The solid curve is obtained from simulations, and the dashed curve represents the results from crystallographic temperature factors measurements.³¹ The curves are normalized so that the area under the curves is equal to one. The inset represents the unnormalized results from simulations. The dashed line drawn at $\langle \Delta R_i^2 \rangle = 1.98$ Å provides an estimate of conformationally labile (upper part), or nonlabile (lower part) regions.

The low-resolution structure is allowed to move by an off-lattice MC method. Accordingly, a randomly chosen site, either an α -carbon or a sidechain site, is subjected to a differential perturbation, using a uniformly distributed random number generator. The strength of the perturbation, Δx , is controlled by the formula $\Delta x = k(2r - 1)$, where r is the random variable $0 \leq r \leq 1$, and k is the damping factor taken as 0.8 Å. One MC time step is composed of 153 moves; this may be viewed as the average time required for all N residues of apo-Mb to have a chance to move. The acceptance of a move is based on the following criterion⁴³: the new conformation $[\Phi]$ is accepted if the condition $\exp[-(E[\Phi] - E[\Phi]_0)/cRT] \leq r$ is satisfied, where $[\Phi]_0$ represents the original state, and c is the Metropolis scaling parameter taken as $c = 1$ in our simulations. An acceptance rate of about 60% was observed on the average. Clearly, the fraction of accepted moves, and consequently the size of the motions, may be easily controlled by modifying the parameter c . Eight independent trajectories of 2,000 MC steps each were generated, and the coordinates of all sites were recorded every three MC steps for subsequent analysis. The length of MC trajectories were observed to lead to an rms deviation of the final structure from crystal structure coordinates of 3.5 ± 0.5 Å. The errors refer to the differences between the eight runs. Also, all runs exhibit a convergence to a consistent $E[\Phi]_0/NRT$ fluctuating in the range 9.4 ± 0.07 , where N is the number of residues.

RESULTS

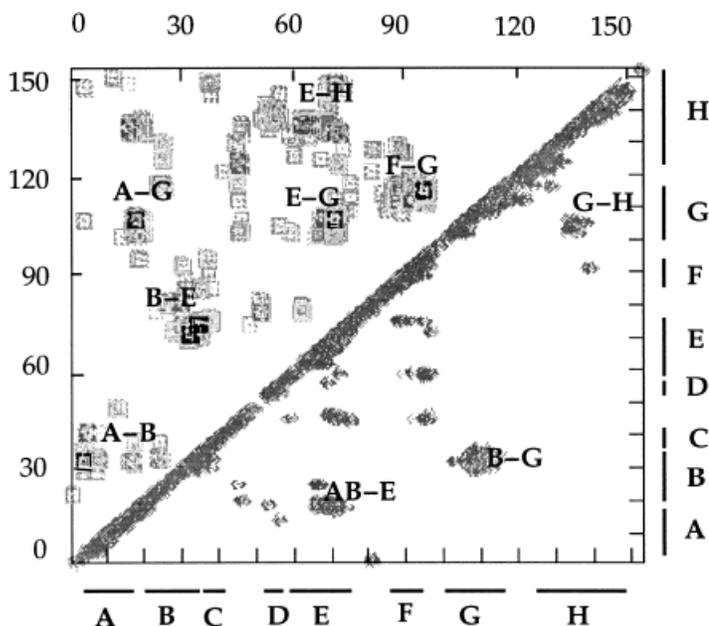
Mean-Square Fluctuations in Atomic Coordinates

The mean-square fluctuations, $\langle \Delta R_i^2 \rangle$, in the position vectors R_i of the backbone sites of apo-Mb are displayed in Figure 3 as a function of residue numbers. The solid curve represents the result from

simulations, while the dashed curve displays the fluctuations evaluated from the crystallographic temperature factors $B_i = 8\pi^2 \langle \Delta R_i^2 \rangle / 3$. Both curves are presented in normalized form, i.e., the area enclosed by each curves is unity. The inset displays the unnormalized results from simulations. The locations of the helices of apo-Mb are indicated by the horizontal segments along the abscissa. A satisfactory agreement between simulations and experiments is observed. Most of the experimental peaks and minima are correctly reproduced in simulations. Perhaps the most pronounced departure between the solid and dashed curves is near helix A. Here, simulations yield relatively smaller fluctuations. However, this lower mobility of helix A is consistent with observations from time-resolved experiments and full atomic MD simulations, as will be shown below.

An average value of 1.4 Å was observed for the rms fluctuations in backbone coordinates throughout runs of 2,000 MC steps (see Fig. 3 inset). The segments that are subject to larger-amplitude fluctuations from mean coordinates may be viewed as conformationally labile or flexible units, in parallel with the interpretation of Brooks.³⁷ As illustrated in the inset of Figure 3, the helices A, G, and B are the most nonlabile structural units with this qualitative classification, followed by helices H and E. The helices C, D, and F exhibit a dual character, possessing both higher- and lower-amplitude fluctuation regions with respect to the mean value. The loop regions between helices, on the other hand, are distinguished by their high mobilities. In particular, the loops CD, EF, and FG exhibit substantially large (≥ 4.0 Å) atomic displacements. The loops EF and FG appear more flexible than in experiment, while the loop GH is less flexible. Such slight differences could be attributed to the low resolution of the present simulations and possible inaccuracies in experimental temperature

Fig. 4. Correlation map for the fluctuations of α -carbons in apo-Mb. The axes represent the residue indices $1 \leq i \leq N$ and $1 \leq j \leq N$, where $N = 153$. Contours of highest correlations C_{ij} are displayed. C_{ij} is found from $C_{ij} = \langle \Delta R_i \cdot \Delta R_j \rangle / (\langle \Delta R_i^2 \rangle \langle \Delta R_j^2 \rangle)^{1/2}$. The diagonal and lower triangular parts display the regions exhibiting positively correlated fluctuations, while the upper triangular portion refers to negatively correlated (or anticorrelated) pairs of residues. The secondary structure elements participating in the strongest correlated or anticorrelated motions are indicated.



factors. It is interesting to note that the low mobility of the loop GH, compared to experiments, was also observed in full atomic simulations of Brooks.³⁷ This observation was explained by the fact that the loop GH packs against the relatively nonlabile A helix. Helices A, G, and H indeed form a relatively nonlabile subdomain, as will be elaborated below.

Cross-Correlations Between Residue Motions

Equilibrium cross-correlations between the motions of residues were evaluated using the relation $C_{ij} = \langle \Delta R_i \cdot \Delta R_j \rangle / (\langle \Delta R_i^2 \rangle \langle \Delta R_j^2 \rangle)^{1/2}$. The division by the mean-square fluctuations provides an estimate of the correlation between the directions of the motion undergone by sites i and j , eliminating the effect of the amplitude of motion. Thus, C_{ij} values lie in the range $-1 \leq C_{ij} \leq 1$; the upper and lower limits refer to fully correlated and fully anticorrelated pairs, respectively.

The results are depicted in Figure 4. The ordinate and abscissa represent the residue indices i and j . The helical regions are indicated along the two axes. The upper triangular part of the correlation map displays the residue pairs exhibiting negative correlations in the range $-1 \leq C_{ij} \leq -0.3$, i.e., the elements undergoing coupled but opposite sense fluctuations, whereas the diagonal and lower triangular parts describe the positively correlated pairs, mainly $0.3 \leq C_{ij} \leq 1.0$. For clarity, pairs exhibiting weak correlations ($-0.3 \leq C_{ij} \leq 0.3$) are not displayed on the map. The names of the helices or loops exhibiting relatively strong correlations ($0.45 \leq C_{ij} \leq 1$) are written on the map.

Examination of the correlation map reveals the following features. First of all, the helix G is involved in strongly correlated and anticorrelated motions with most of structural units. It is positively correlated with helices B and H, and negatively correlated with helices A, E, and F. This suggests the existence of two blocks undergoing opposite-sense fluctuations with respect to each other, mainly the block of helices B-G-H subject to concerted, same-direction fluctuations, and the helices A-E-F, which form another group, similarly exhibiting positive correlations. The existence of strong anticorrelations between helix pairs E-H, B-E, and A-B observed in the upper triangular part of the map supports this model. A positive correlation is found between the loop AB and the helix E, which indicates that the motion of the loop AB as well is coupled to that of the group A-E-F. The helix G, by participating in five strongly correlated pairs out of nine displayed in Figure 4, appears to form the core of the cooperative motions of the protein.

We note that helices C, D, and F do not appear in the correlation map; these are subject to relatively uncorrelated motions with other segments. In previous MD simulations, a major change in the tertiary structure of apo-Mb during unfolding was pointed out to be the disappearance of contacts between the unfolded D and F helices, and the rest of the molecule.³⁸ The absence of correlations between these helices and any other part of the molecule presently observed is consistent with the loss of contacts observed in these MD simulations.³⁸ Furthermore, an increase in contacts between the CD loop and

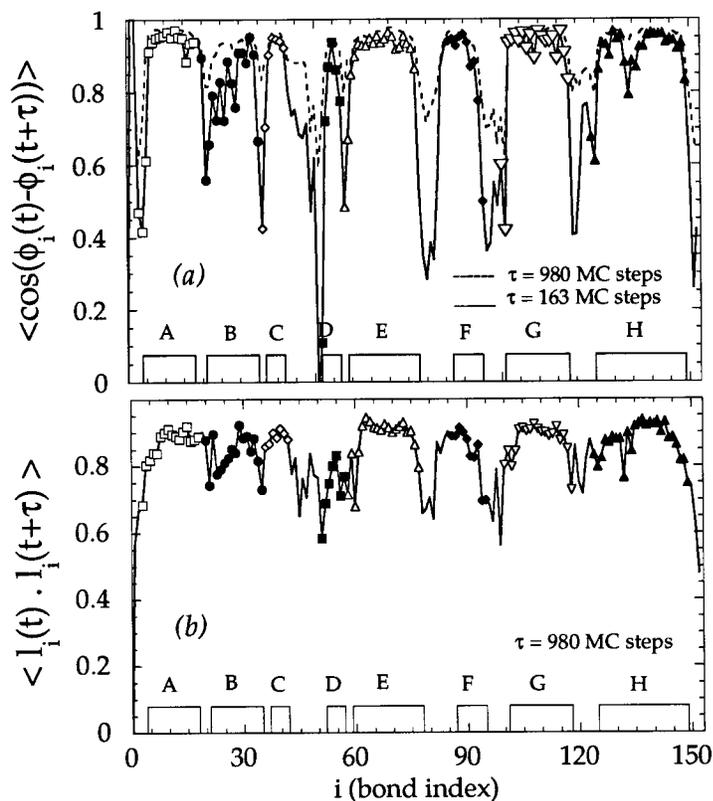


Fig. 5. **a**: Time-delayed autocorrelations for virtual bond rotations, $\langle \cos[\phi_i(t + \tau) - \phi_i(t)] \rangle$ evaluated for all $C^\alpha-C^\alpha$ virtual bonds of apo-Mb at $\tau = 163$ (dashed) and 980 (solid) MC steps. Results are shown as a function of residue index. For clarity, bonds belonging to distinct helices are displayed by different symbols. **b**: Time-delayed autocorrelation function $\langle \hat{l}_i(t) \cdot \hat{l}_i(t + \tau) \rangle$ associated with the reorientation of virtual bonds of apo-Mb evaluated at $\tau = 980$ MC steps as a function of bond index. $\hat{l}_i(t)$ designates the unit vector along the virtual bond $C_{i-1}^\alpha - C_i^\alpha$. Note the weaker loss in orientational correlation of backbone bonds in part B compared to that of dihedral angles in part A, which shows that bond torsional motions generally compensate each other so as to localize the conformational change and minimize the displacement of backbone atoms.

both helix G and loop FG was reported. These groups exhibit anticorrelated motions in our simulations, as may be observed from the upper diagonal portion of Figure 4.

Time Decay of Conformational Correlations

Time-delayed correlations between the conformational states of residues will be evaluated in terms of the time evolution of rotational states of $C^\alpha-C^\alpha$ virtual bonds. Torsional autocorrelation function $G_i(\tau) = \langle \cos[\phi_i(t + \tau) - \phi_i(t)] \rangle$ are evaluated for this purpose. Here, the angle brackets represent the average change in the dihedral angle ϕ_i of bond i within a time interval τ , taken over all time origins t , including all eight runs. $G_i(\tau)$ decays from unity, at $\tau = 0$, to zero, at long times.

Figure 5a illustrates torsional autocorrelations $G_i(\tau)$ of all bonds for $\tau = 163$ and 980 MC steps. Dotted and solid curves refer to the two respective time intervals. The two curves are similar from qualitative point of view; however, the losses of conformational correlations become more pronounced at longer times. Residues in different helices are indicated by different symbols for clarity. From this figure, it is possible to identify the regions that have different conformational mobilities. We can distinguish helices A, E, and G, for example, which have highly ordered, kinetically stable residues insofar as the bond torsional motions are concerned, while

helices B, F, and H possess both stable residues and relatively more mobile residues. Helices C and D exhibit the highest unwinding tendency.

A more detailed examination shows that motions in helix F increase steadily from the N-terminal end into the FG loop. It is this section that moves into the vacant heme cavity. Also, the relatively more mobile regions starting from the C-terminal end of the B helix through the N-terminal end of the E helix are generally those regions adjacent to the heme pocket, showing substantial motions as a result of the absence of the prosthetic group. Among the loops between helices, the CD loop exhibits the highest mobility. The virtual bond between Thr51 and Lys50 in the loop exhibits a complete loss of torsional correlation at $\tau = 980$ MC steps. The CD loop is followed by the EF and FG loops, which are also distinguished by torsional correlations of $G_i(\tau) < 0.4$. These observations are closely related with those made in recent experiments, as will be considered in the Discussion.

Figure 5b displays the losses in the bond orientations at $\tau = 980$ MC steps, evaluated from the bond autocorrelation function, $M_i(\tau) = \langle \hat{l}_i(t) \cdot \hat{l}_i(t + \tau) \rangle$, where $\hat{l}_i(t)$ is the unit vector along the instantaneous (time t) position of bond i . The angle brackets again refer to the ensemble average over all initial times t , and over all independent runs. Similar dynamic characteristics are observed in both Figure 5a and b.

TABLE I. Pairs of Virtual Bonds* Exhibiting Strongly Correlated Rotations

Anticorrelated pairs [†]	Correlated pairs [‡]
Bond pairs (bond location [§])	Bond pairs (bond location [§])
14–17 (A12–A15)	16–24 (A14–B5)
21–103 (B2–G4)	21–22 (B2–B3)
47–76 (CD5–E19)	21–23 (B2–B4)
55–56 (D5–D6)	21–107 (B2–G8)
66–68 (E10–E11)	21–151 (B2–C terminus)
103–107 (G3–G8)	39–47 (C3–CD5)
104–105 (G5–G6)	50–51 (CD8–CD9)
103–107 (G5–G8)	59–72 (E2–E15)
107–108 (G8–G9)	81–145 (EF4–H22)
111–112 (G12–G13)	83–146 (EF6–H23)
112–113 (G13–G14)	105–Ile107 (G6–G8)
131–132 (H8–H9)	112–114 (G13–G15)
	118–120 (G19–GH2)
	120–122 (GH2–GH4)
	144–146 (H21–H23)

*Bond *i* refers to virtual bond between α -carbons of residues *i* – 1 and *i*.

[†] $C_{ij}(\tau = 980 \text{ steps}) \leq -0.4$.

[‡] $C_{ij}(\tau = 980 \text{ steps}) > 0.4$.

[§]Bond location lists the bond index within indicated helices and loops.

This is expected since bond reorientations in Figure 5b are mainly induced by the torsional motions in 5a. However, a major difference is that the time decays of bond autocorrelation functions are substantially slower than those of dihedral angles, except for the chain termini. Thus, the time loss in torsional correlations is not directly reflected on bond reorientations, but substantially reduced. This property, also observed in simulations of polymers in dense media,⁴⁴ reveals the cooperative nature of bond rotations. The net effect is to localize the motion along the backbone, and consequently preserve most of tertiary contacts. A systematic analysis of cross-correlations between bond rotations, and the types and extent of coupled, compensating dihedral angle motions, will be presented next.

Cross-Correlations Between Bond Rotations

The time-delayed cross-correlations between the rotations of the virtual bonds *i* and *j* have been evaluated from $C_{ij}(\tau) = \langle \Delta\phi_i(\tau) \Delta\phi_j(\tau) \rangle / (\langle \Delta\phi_i(\tau)^2 \rangle^{1/2} \langle \Delta\phi_j(\tau)^2 \rangle^{1/2})$, where $\Delta\phi_i(\tau) = \phi_i(t + \tau) - \phi_i(t)$ is the change in the rotational angle of the virtual bond *i* occurring within a time period τ . The bond pairs whose cross-correlations are greater than 0.4 or less than -0.4 at $\tau = 980$ MC steps are listed in Table I.

Several bonds belonging to helix G appear in Table I. The high cooperativity between the rotations of the C^α – C^α bonds of helix G reveals, in fact, the strong tendency of this helix to preserve its secondary structure and tertiary contacts. As pointed out earlier, helix G is in a relatively tightly packed environ-

ment, its fluctuations being strongly correlated with those of several nonbonded units (see Fig. 4). Under these circumstances, the only means of accommodating bond rotations for residues in helix G appears to effectuate coupled, compensating rotations by near neighbor bonds. Figure 6 depicts, for example, the history of the rotational angles of some bonds in G helix. Figure 6A displays the coupled rotations of bonds 103 and 107. These are the virtual bonds between residues Lys102 and Tyr103, and between Phe106 and Ile107, respectively. A strong anticorrelation between the torsional motions of these bonds is observed. Positive departures from the helical dihedral angle (ca. 50° in the virtual bond representation) undergone by bond 103, for example, are accompanied by negative rotations of bond 107, and vice versa. It is interesting to notice that these coupled rotations involve the cooperative displacements of atoms/groups separated by up to eight virtual or 24 real bonds along the chain.

Figure 6B and C illustrate the coupled rotations of the bond pairs (105, 107) and (107, 108), respectively. The former pair undergoes positively correlated rotations, i.e., the two bonds exhibit a tendency to rotate simultaneously in the same sense. Bonds 107 and 108, on the other hand, show anticorrelated torsional motions, similar to the pair in Figure 6A.

In the interest of elucidating the generality of the types of correlations observed between near neighboring bond rotations along the chain, a systematic analysis was performed as a function of the separation (*j* – *i*) of rotating bonds along the chain. The results are displayed in Figure 7. The strongest correlation is found between the rotations of second neighbors, as observed from the peak at *k* = 2. The tendency of second-neighbor virtual bonds to undergo positively correlated rotational motions was already indicated in Figure 6B. On the other hand, the first, third, and fourth neighbors along the chain exhibit a tendency toward counterrotations, in conformity with the behavior seen in Figure 6A and C. Correlations between bond rotations vanish beyond the fourth neighbors. This does not prevent, however, the occurrence of strongly coupled rotations undergone by farther neighbors. Such coupled transitions do, in fact, occur due to tertiary interactions, some examples of which may be found in Table I.

DISCUSSION AND CONCLUSION

Advantages and Limitations of the Low-Resolution Approach

The two major advantages of the present low-resolution approach are its simplicity and computational efficiency. A very simple model, mainly a virtual bond approximation consisting of two sites per residue, is adopted, along with knowledge-based parameters specifically developed for the particular

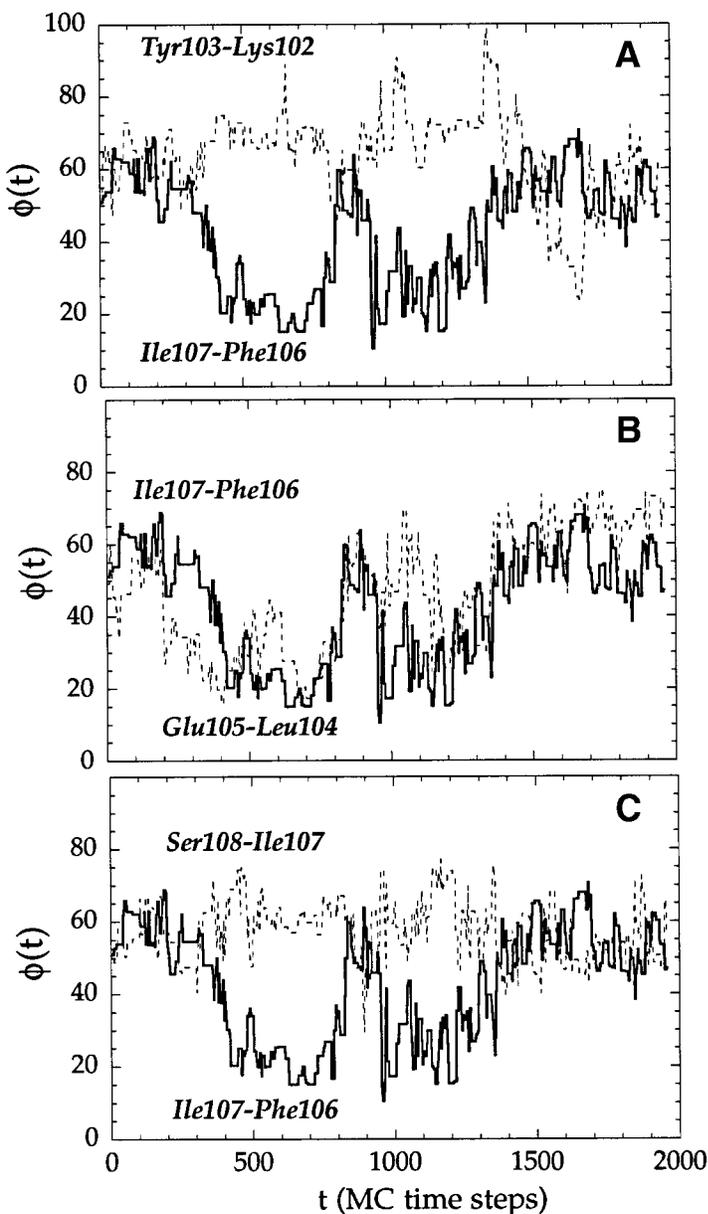


Fig. 6. Time evolution of bond rotations shown for different pairs of bonds. **A** displays the dihedral angles of bonds 103 (Lys102-Tyr103) and 107 (Phe106-Ile107), a pair of bonds exhibiting strongly anticorrelated torsional motions. Note the approximately symmetrical departure from the starting value ($\sim 50^\circ$) corresponding to helical state in the virtual bond model. **B** illustrates the behavior of bonds, 105 (Leu104-Glu105) and 107 (Phe106-Ile107), undergoing positively correlated rotations. Finally, a pair of anticorrelated bonds, 107 (Phe106-Ile107) and 108 (Ile107-Ser108), is displayed in **C**.

low-resolution model. The simulation algorithm is simple as well: MC moves in the form of cyclic perturbations over all coordinates, subject to a Metropolis criterion ensuring the generation of trajectories in the neighborhood of the native state. A series of independent runs, each requiring no more than 2-hr CPU time on a Silicon Graphics R4400 Challenge Workstation, were performed, which were verified to yield reproducible, statistically reliable data.

Several insightful results were obtained on the mechanisms of cooperative conformational motions, as illustrated in Figures 6 and 7. Also, conformational characteristics of apo-Mb in close accord with those extracted from previous MD simulations were obtainable. In MD simulations of isolated apo-Mb helices, the intrinsic stabilities were observed to be

in the order of $A \approx G \approx H > B > E > F$.³⁹ Previous MD simulations³⁷ showed that the helices A, B, E, G, and H undergo the least amount of motion, while helices C, D, and F move substantially. The more detailed MD simulations of apo-Mb by Tirrardo-Rives and Jorgensen³⁸ also showed that large segments of A, E, G, and H helices and one or two turns of helix B were quite stable. In that study, helix C was reported to fluctuate between folded and unfolded states, while complete decay of D and F helices and a break toward the C-terminal end of helix H were observed. Our simulations indicate similar features, in general, which is quite satisfactory in view of the incomparable efficiency and simplicity of the present approach with respect to full atomic MD simulations.

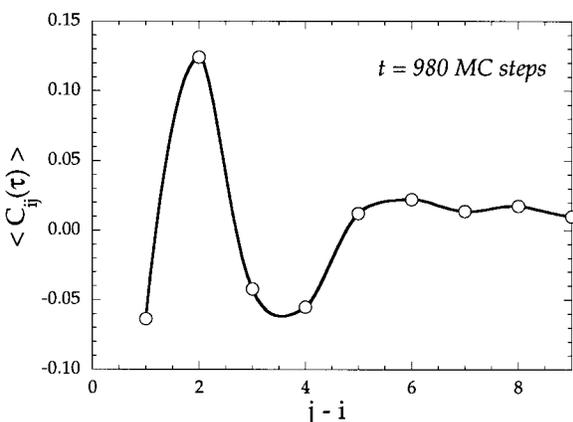


Fig. 7. Dependence of time-delayed cross-correlations $\langle C_{ij}(\tau) \rangle$ between bond torsional motions on the separation $k = j - i$ between the bonds along the polypeptide backbone. Results are displayed for $\tau = 980$ MC time steps. Second neighbor are observed to undergo corotations, while the first, third, and fourth neighbors exhibit a tendency for counterrotations. No net correlations are discernible beyond the fifth neighbor.

The major limitations are the following. Because this is a coarse-grained model, structural details at the atomic level cannot be determined. For example, specific interactions with the proximal or distal histidines at the binding site of apo-Mb, cannot be examined. We acquire, instead, a global picture of the collective dynamics of the molecule. Regions having a stronger tendency to unfold are determined, as well as those distinguished by an enhanced stability. Here we use the term *stability* in the kinetic sense, to indicate a stronger resistance to unwinding, for example, in the case of α -helices. This tendency is evaluated from the time decay of autocorrelations associated with bond torsional and orientational motions.

Another limitation of the approach is that the corresponding energy parameters being extracted from databank structures are representative of the energetics of compact, globular structures only, and may fail to account for the interactions in the more expanded, or unfolded, configurations. The solvent molecules are not explicitly considered either, but in an indirect way through solvent-mediated potentials of mean force. The satisfactory results presently obtained support the use of the approach to simulating the dynamics near native fold. Extension of the method to the complete unfolding might require the use of potentials accounting for the changes in environment, as we have discussed in a recent review.¹ Also, a more detailed description of amino acid sidechains may be adopted, as well as explicit consideration of hydration. Furthermore, directional dependences of interresidue interactions could also be taken into consideration. Recently, we analyzed the orientational preferences between sidechain interactions using a knowledge-based approach.⁸ The resulting potentials could also be included in the energy field adopted in simulations.

Fluctuations in C^α Atoms: Comparison With Crystallographic Temperature Factors and MD Simulations

The rms deviations in C^α coordinates from their mean values, averaged over all residues, were 1.4 Å during 2,000 MC step runs. This value is higher than that (~ 0.6 Å) indicated by experiments, and those (0.6–1.2 Å) observed in MD simulations of apo-Mb.^{37,38} This difference may be attributed to the adoption of a low-resolution model in the present study, in which the potential energy surface is smoother. Additionally, the amplitudes of motions are readily controlled by changing the strength of the Metropolis criterion. Therefore, rather than the absolute values of fluctuations, their variation with residues along the protein is of interest. Moreover, Figure 3 demonstrates that a satisfactory agreement between theory and experiments is obtained in this respect, notwithstanding the simplicity of the model.

For comparative purposes, we also considered the crystallographic B factors of sperm whale Mb.⁴⁵ A close similarity was observed between the temperature factors of Mb and those of apo-Mb in the complex formed with biliverdin, confirming that biliverdin plays a role comparable to that of heme in Mb. The most significant differences were the lower-amplitude fluctuations of Mb in the neighborhood of helix D, and the presence of a peak near loop AB in Mb, which is almost absent in apo-Mb. Interestingly, our simulations yielded residue-specific fluctuations comparable to both experimental results. This leads to the important conclusion that the fluctuations are predominantly determined by the collective dynamics of all residues in the protein, being little affected by the contribution of the prosthetic group. The collective motions, in turn, are adequately described by a low-resolution model, such as the present one.

Our previous study⁴⁰ further showed that the details of the crystallographic temperature factors may be reasonably well approximated even by simple analytical treatments devoid of residue specificity. A single-parameter harmonic potential was adopted therein between all residue pairs in close contact ($r_{ij} \leq 7.0$ Å). Thus, a global consideration of the three-dimensional pattern of all interresidue contacts ensures the correct prediction of the variations in fluctuation amplitudes of individual residues.

Comparison With Time-Resolved Experiments: Evidence for Stable Subdomain A-B-G-H

Partially folded structures of apo-Mb have been characterized by several time-resolved techniques, such as H/D exchange labeling experiments coupled with two-dimensional NMR,³² or stopped-flow circular dichroism.³³ Measurements for the partly unfolded intermediate I during the *unfolding* of apo-Mb indicated that helices A, G, and H are the most stable elements (exhibiting the highest protection from

exchange during unfolding), whereas B and E are relatively unstable.³² Exchange rates for helices C, D, and F could not be determined due to lack of available probes. On the other hand, H/D exchange pulse labeling and stopped-flow circular dichroism (CD) experiments³³ were used to identify the structure of the earliest intermediates formed during *refolding*: helices A, G, and H, and part of helix B were found to fold first, followed by the remainder of helix B, later on by the helices C and E, and finally by the CD loop. Thus, there is a close correspondence between the early folding regions³³ and those exhibiting the strongest resistance to unfolding.³² The formation of a structured A-G-H bundle at early stage (under 20 μ s) of folding is also indicated by the temperature-jump-induced refolding experiments probed by Trp fluorescence.³⁵

The potentially yielding or nonyielding regions of the protein, in the kinetic sense, are identified from the correlation analysis of our coarse-grained trajectories. The element that plays the most important role in mediating the collective motions of the protein is found to be helix G. The kinetic stability of helix G is manifested in two ways. First, the bond rotations within helix G occur in a highly cooperative way so as to compensate any occasional disruption of the helix structure, as illustrated in Figure 6. Second, in addition to strong intrahelical correlations, helix G exhibits substantial interhelical correlations. Its motions are strongly correlated with those of helices H and B, and anticorrelated with those of A, E, and F, as illustrated in Figure 4. In the other extreme case of helices exhibiting uncorrelated, almost random, motions, we distinguish helix D, followed by the C-terminal part of helix F, helix C, and a few residues at the N-terminus of helix E (Fig. 5). The high mobility of helices D and C, as opposed to the high stability of helix G, and the strong coupling between helices G, H, and A are in close agreement with the results from time-resolved experiments. Our simulations indicate that helix B is also coupled to the group A-G-H, supporting the model proposed by Jennings et al.³³

Further Comparison With NMR Experiments and Proposed Mechanism of Concerted Motions Near Folded State

In addition to the block A-G-H-B, we found that a large part of helix E is also highly stable, apart from the unwinding of a few residues at its N-terminus (Fig. 5). Helix E is strongly anticorrelated with the group G-H-B, as evidenced by the peaks in the upper diagonal portion of our fluctuation correlation map (Fig. 4). The high stability of helix E is not in accord with the models inferred from time-resolved experiments. However, it conforms with the one- and two-dimensional proton NMR measurements of Cocco and Lecomte.³⁴ In these experiments, the helices E, G, and H were pointed out to be largely intact in

apo-Mb, as well as the terminal helices A-C. Unfolding or large-amplitude fluctuations were proposed to be found mostly in the D helix, the beginning of E helix, the EF loop, the F helix, and the C terminus, which are in accord with the regions indicated by our simulations to undergo large-amplitude fluctuations (Fig. 3).

In a more recent study, Lecomte et al.³⁶ performed a detailed study of the A-B-G-H interface of apo-Mb, which confirmed that this subdomain is indeed folded into a structure superimposable (with rms deviation of 2.0 Å) on that of holo-Mb. However, a 10-residue segment of helix E, excluding the N-terminal part, was also shown therein to take part in the same compact subdomain, being stabilized by A-E and B-E interhelical contacts. This is consistent with the correlation map (Fig. 4) obtained in our simulations. In this map, a strong coupling, in the form of anticorrelation, is observed between helices B and E. On the other hand, a close examination of the lower portion of the map, in which the pairs of strongly coupled elements are displayed, indicates that the C-terminal part of helix A is strongly correlated with helix E. This lends support to the consideration of the helix E, or a large segment of it, as part of the early forming subdomain of apo-Mb.

In summary, the present coarse-grained simulations suggest that the protein may be divided into two groups of helices on the basis of the mechanism of concerted conformational motions: helices G-B-H and A-E-F. The group G-H-B undergoes highly correlated, synchronous motions, which are coupled but in opposite sense to those of the group A-E-F. Helix G appears to lead to first group, while A and E dominate the motion of the second group. In the other extreme case, the helix that exhibits the fastest unwinding is the D helix. Stable regions, such as helix G, are distinguished by cooperative bond rotations, extending up to more than 20 (real) bonds along the peptide backbone.

REFERENCES

1. Jernigan, R.L., Bahar, I. Structure-derived potentials and protein simulations. *Curr. Opin. Struct. Biol.* 6: 195-209, 1996.
2. Tanaka, S., Scheraga, H.A. Medium- and long-range interaction parameters between amino acids for predicting three-dimensional structures of proteins. *Macromolecules* 9:945-950, 1976.
3. Miyazawa, S., Jernigan, R.L. Estimation of effective inter-residue contact energies from protein crystal structures: Quasichemical approximation. *Macromolecules* 18:534-552, 1985.
4. Sippl, M.J. Knowledge-based potentials for proteins. *Curr. Opin. Struct. Biol.* 5:229-235, 1995.
5. Bahar, I., Jernigan, R.L. Interresidue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation. *J. Mol. Biol.* 266:195-214, 1997.
6. Singh, J., Thornton, J.M. SIRIUS: An automated method of analysis of preferred packing arrangements between protein groups. *J. Mol. Biol.* 211:595-615, 1990.
7. Vriend, G., Sander, C. Quality control of protein models:

- Directional atomic contact analysis. *J. Appl. Cryst.* 26:47–60, 1993.
8. Bahar, I., Jernigan, R.L. Coordination geometry of non-bonded residues in globular proteins. *Folding Design* 1:357–370, 1996.
 9. DeWitte, R.S., Shakhnovich, E.I. Pseudodihedrals: Simplified protein backbone representation with knowledge-based energy. *Protein Sci.* 2:1570–1581, 1994.
 10. Bahar, I., Kaplan, M., Jernigan, R.L. Short-range conformational energies, secondary structure propensities, and recognition of correct sequence-structure matches. *Proteins* 29:292–308, 1997.
 11. Hendlich, M., Lackner, P., Weitckus, S., et al. Identification of native protein folds amongst a large number of incorrect models: The calculation of low-energy conformations from potentials of mean force. *J. Mol. Biol.* 216:167–180, 1990.
 12. Kocher, J.-P.A., Rooman, M.J., Wodak, S. Factors influencing the ability of knowledge-based potentials to identify native sequence-structure matches. *J. Mol. Biol.* 235:1598–1613, 1994.
 13. Wang, Y., Zhang, H., Li, W., Scott, R.A. Discriminating compact nonnative structures from the native structure of globular proteins. *Proc. Natl. Acad. Sci. U.S.A.* 92:709–713, 1995.
 14. Miyazawa, S., Jernigan, R.L. Protein stability for single substitution mutants and the extent of local compactness in the denatured state. *Protein Eng.* 7:1209–1220, 1994.
 15. Altuvia, Y., Schueler, O., Margalit, H. Ranking potential binding peptides to MHC molecules by a computational threading approach. *J. Mol. Biol.* 249:244–250, 1995.
 16. Levitt, M., Warshel, A. Computer simulation of protein folding. *Nature* 253:694–698, 1975.
 17. Ueda, Y., Taketomi, H., Go, N. Studies on protein folding, unfolding and fluctuations by computer simulations. II. A three-dimensional lattice model of lysozyme. *Biopolymers* 17:1531–1548, 1978.
 18. Skolnick, J., Kolinski, A. Dynamic Monte Carlo simulation of globular protein folding/unfolding pathway. I. Six member, greek key β barrel proteins. *J. Mol. Biol.* 212:787–817, 1990.
 19. Covell, D.G. Lattice model simulation of polypeptide chain folding. *J. Mol. Biol.* 235:1032–1043, 1994.
 20. Kolinski, A., Skolnick, J. Monte Carlo simulation of protein folding. I. Lattice model and interaction scheme. *Proteins* 18:338–352, 1994.
 21. Sali, A., Shakhnovich, E., Karplus, M. Kinetics of protein folding: A lattice model study of the requirements for folding to the native state. *J. Mol. Biol.* 235:1614–1636, 1994.
 22. Vieth, M., Kolinski, A., Brooks, I.C.L., Skolnick, J. Prediction of the folding pathways and structure of GCN4 leucine zipper. *J. Mol. Biol.* 237:361–367, 1994.
 23. Elofsson, A., Le Grand, S.M., Eisenberg, D. Local moves: An efficient algorithm for simulation of protein folding. *Proteins* 23:73–82, 1995.
 24. Karplus, M., Sali, A. Theoretical studies of protein folding and unfolding. *Curr. Opin. Struct. Biol.* 5:58–73, 1995.
 25. Sun, S. Reduced representation model of protein structure prediction: Statistical potential and genetic algorithms. *Protein Sci.* 2:762–785, 1993.
 26. Park, B.H., Levitt, M. The complexity and accuracy of discrete state models of protein structure. *J. Mol. Biol.* 249:493–507, 1995.
 27. Sun, S., Thomas, P.D., Dill, K.A. A simple protein folding algorithm using a binary code and secondary structure constraints. *Protein Eng.* 8:769–778, 1995.
 28. Park, B.H., Levitt, M. Energy functions that discriminate X-ray and near-native folds from well-constructed decoys. *J. Mol. Biol.* 258:367–392, 1996.
 29. Griko, Y.V., Privalov, P.L., Venyaminov, S.Y., Kutysenko, V.P. Thermodynamic study of the apomyoglobin structure. *J. Mol. Biol.* 202:127–138, 1988.
 30. Barrick, D., Baldwin, R.L. Three-state analysis of sperm whale apomyoglobin folding. *Biochemistry* 32:3790–3796, 1993.
 31. Wagner, U.G., Moeiler, N., Schmitzberger, W., Falk, H., Kratky, C. Structure determination of the biliverdin apomyoglobin complex: Crystal structure analysis of two crystal forms at 1.4- and 1.5-Å resolution. *J. Mol. Biol.* 247:326–337, 1995.
 32. Hughson, F.M., Wright, P.E., Baldwin, R.L. Structural characterization of a partly folded apomyoglobin intermediate. *Science* 249:1544–1548, 1990.
 33. Jennings, P.A., Wright, P.E. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* 262:892–895, 1993.
 34. Cocco, M.J., Lecomte, J.T.J. The native state of apomyoglobin described by proton NMR spectroscopy: Interaction with the paramagnetic probe HyTEMPO and the fluorescent dye ANS. *Protein Sci.* 3:267–281, 1994.
 35. Ballew, R.M., Sabelko, J., Gruebele, M. Direct observation of fast protein folding: The initial collapse of apomyoglobin. *Proc. Natl. Acad. Sci. U.S.A.* 93:5759–5764, 1996.
 36. Lecomte, J.T.J., Kao, Y., Cocco, M.J. The native state of apomyoglobin described by proton NMR spectroscopy: The A–B–G–H interface of wild-type sperm whale apomyoglobin. *Proteins* 25:267–285, 1996.
 37. Brooks III, C.L. Characterization of 'native' apomyoglobin by molecular dynamics simulation. *J. Mol. Biol.* 227:375–380, 1992.
 38. Tirado-Rives, J., Jorgensen, W.L. Molecular dynamics of the unfolding of apomyoglobin in water. *Biochemistry* 32:4175–4184, 1993.
 39. Hirst, J.D., Brooks III, C.L. Molecular dynamics simulations of isolated helices of myoglobin. *Biochemistry* 34:7614–7621, 1995.
 40. Bahar, I., Atilgan, A.R., Erman, B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding Design* 2:173–181, 1997.
 41. Haliloglu, T., Bahar, I., Erman, B. Gaussian dynamics of folded proteins. *Phys. Rev. Lett.* 79:3090–3093, 1997.
 42. Bernstein, F., Koetzle, T., Williams, G., et al. The protein databank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112:535–542, 1977.
 43. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* 21:1087–1092, 1953.
 44. Haliloglu, T., Bahar, I., Erman, B., Kim, E.-G., Mattice, W.L. A dynamic rotational isomeric state approach for extension of the time scale of the local dynamics observed in fully atomistic molecular dynamics simulations. *J. Chem. Phys.* 104:4828–4834, 1996.
 45. Kuriyan, J., Wilz, S., Karplus, M., Petsko, G.A. X-ray structure and refinement of carbon-monooxy (Fell)-myoglobin at 1.5-Å resolution. *J. Mol. Biol.* 192:133–154, 1986.