Eur Biophys J (1996) 24: 387–403

ARTICLE

Daniel Hoffmann · Ernst-Walter Knapp

Polypeptide folding with off-lattice Monte Carlo dynamics: the method

Received: 22 January 1996 / Accepted: 3 April 1996

Abstract The MC dynamics of an off-lattice all-atom protein backbone model with rigid amide planes are studied. The only degrees of freedom are the dihedral angle pairs of the C_α-atoms. Conformational changes are generated by Monte Carlo (MC) moves. The MC moves considered are single rotations (simple moves, SM's) giving rise to global conformational changes or, alternatively, cooperative rotations in a window of amide planes (window moves, WM's) generating local conformational changes in the window. Outside the window the protein conformation is kept invariant by constraints. These constraints produce a bias in the distribution of dihedral angles. The WM's are corrected for this bias by suitable Jacobians. The energy function used is derived from the CHARMM force field. In a first application to polyalanine it is demonstrated that WM's sample the conformational space more efficiently than SM's.

Key words Monte Carlo dynamics · Off-lattice moves · Local conformational changes · Protein long-time dynamics · Protein folding

Introduction

The relationship of structure, function and dynamics of proteins[1] (Dickerson and Geis 1969; Fersht 1984; Schulz and Schirmer 1978; Hol 1991; Levitt 1992; Honig 1993; Kuntz 1994; Eaton and Szabo 1991) and the protein folding problem (Kolata 1986; Baldwin 1990; Creighton and Kim 1991; Creighton 1992; Knapp and Sklenar 1993) are great challenges in theoretical molecular biophysics. Functionally important dynamics in proteins are often slow compared to picosecond and subpicosecond dynamics of vibrational degrees of freedom. The dynamics of protein folding with a typical time scale from microseconds to seconds is even slower and cannot be accessed by conventional methods of molecular dynamics. The conventional method of computer simulation of protein dynamics is based on Newton's equations of motion (Brooks et al. 1983; McCammon and Harvey 1987; van Gunsteren 1988; Karplus 1986) which are solved in Cartesian coordinates for all atoms. This is very time consuming so that computer simulations of dynamics far beyond the nanosecond time regime are not yet possible. The reasons for these limitations in time scale are as follows:

1. The number of non-bonded atom pair interactions is very large, so that most of the CPU time is spent on the evaluation of non-bonded interactions.
2. It is necessary to follow the fast intramolecular vibrations in all details, which requires an elementary step of propagation in time of typically 1 fs.

There are two important strategies (Karplus 1986; van Gunsteren 1991) to improve this situation:

1. Reduction of the number of non-bonded atom pair interactions.
2. Elimination of fast vibrations and thus increase of the elementary time step.

The replacement of small molecular groups such as methyl groups by compound atoms reduces the number of atom pair interactions. The concept of compound atoms can be extended further by replacing large molecular groups such as amino acid side chains by single structureless compound atoms as demonstrated in early (Levitt and Warshel 1975; Tanaka and Scheraga 1976; Kuntz et al. 1976; McCammon et al. 1980) and also more recent studies (Wilson and Doniach 1989; Honeycutt and Thirumalai 1990; Guo et al. 1992; Head-Gordon and Brooks 1991; Gerber 1992; Kang et al. 1993; Fukugita et al. 1993). This

[1] Abbreviations: CPU, Central Processing Unit; MC, Monte Carlo; MCD, Monte Carlo Dynamics; MD, Molecular Dynamics; RMS, Root-Mean-Square; RMSD, Root-Mean-Square-Deviation; SM, Simple Move; WM, Window Move

D. Hoffmann · E.-W. Knapp (✉)
Freie Universität Berlin, Fachbereich Chemie,
Institut für Kristallographie, Takustrasse 6,
D-14195 Berlin, Germany
(Fax (+49) 30838-3464 e-mail knapp@Chemie.fu-Berlin.de)

procedure leads not only to a reduction of the number of non-bonded atom pairs but also eliminates high frequency vibrations. Hence with the use of compound atoms the step size of time propagation can be increased. However, these savings go along with a loss of detail in the protein model, and therefore with a loss of spatial and energetic resolution.

Bond lengths and bond angles are stiff degrees of freedom with fast vibrations which allow only small atomic displacements. Even though these degrees of freedom may serve as the "lubricant" for large conformational changes of a protein (Karplus 1986) they are not interesting in their own right. By eliminating these degrees of freedom one can increase the elementary time step.

Holonomic constraints can be used to keep bond lengths constant. If the molecule is described in Cartesian coordinates, the equations defining the constraints can be solved iteratively by the SHAKE algorithm (Ryckaert et al. 1977). If bond angle constraints are also included, the CPU time required to solve the equations for the constraints outweights the gain from a larger step for the propagation in time (van Gunsteren and Berendsen 1977). Beside these more technical problems, holonomic constraints give rise to a bias in the statistical weights (van Gunsteren and Berendsen 1977; Fixman 1974; Fixman 1978; Helfand 1979). This is due to an uneven treatment of kinetic energy terms from constrainted and unconstrained degrees of freedom which can be overcome by considering the contributions of a suitable metric tensor (Fixman 1974). New molecular dynamics methods, based on implicit integrators, are being developed (Derreumaux and Schlick 1995). They promise an increase of the time step by an order of magnitude relative to conventional molecular dynamics.

In the present approach we use an all atom protein model where bond lengths and bond angles are fixed and the amide plane is kept planar, so that the only degrees of freedom in the protein backbone are the dihedral angles ϕ and ψ at the C_α-atoms. Problems with the metric tensor (Fixman 1974) can be avoided by considering the cooperative motion of extended parts of the molecule where the friction is large enough, so that the motion can be described as a diffusion process. This treatment is valid, if the time step suitable for the time propagation of these units is larger than the decay time of the velocity autocorrelation function, which is typically 0.1 ps (Nadler et al. 1987) for protein side chain atoms.

In the present work the diffusion process is simulated by a Metropolis algorithm (Metropolis et al. 1953). New conformations are obtained with a Monte Carlo (MC) method, which uses the so-called window moves (WM's). WM's generate local conformational changes by cooperative rotations in a small window of consecutive amide planes (Knapp and Irgens-Defregger 1991; Knapp and Irgens-Defregger 1992; Knapp 1992; Rey and Skolnick 1993; Kurochkina et al. 1994). A similar approach was used much earlier to prove that a tripeptide cannot be cyclic (Gō and Scheraga 1970). Later this approach was used in different applications such as, for example, generation and variation of loop conformations (Bruccoleri and Karplus 1985; Sklenar et al. 1986; Palmer and Scheraga 1992; Hubbard et al. 1994; Braun 1987), modeling of idealized protein structures and investigating their structural flexibility. A related MC method uses collective rotations of dihedral angles corresponding to normal modes of the protein (Noguti and Gō 1985). Another efficient MC strategy takes advantage of dynamically optimized MC moves (Bouzida et al. 1992). Recently MC moves involving rotations with flexible bond lengths and bond angles have been successfully applied (Sung 1994; Abagyan and Totrov 1994). A problem from robotics, namely the inverse kinematics for serial manipulators with six joints (Roth et al. 1973; Primrose 1986; Manseur and Doty 1989), is closely connected to the search for possible conformations in WM's. The problem in robotics is to orient and position a hand of a robot in a specific way. Quite recently a MC method with WM's was applied to a polymer model (Dodd et al. 1993). There it was found that due to the constraints which are necessary to keep the conformation of the polymer invariant outside the window, the distribution of dihedral angles is biased. In earlier applications of WM's (Knapp and Irgens-Defregger 1991, 1992; Knapp 1992) this bias had not been recognized. In the present treatment the bias is removed by using suitable Jacobians, i.e. in a way similar to that used in Dodd et al. (1993). In a first application to polyalanine the performance of WM's is contrasted to that of a simple MC move (simple move, SM), where the polypeptide conformation is altered by changes of a single dihedral angle. A SM leads to large displacements of atoms far away from the corresponding rotation axes. In globular proteins SM's therefore often result in structures with atomic clashes, which are energetically unacceptable.

Polymers have also been modeled on regular lattices. In these cases the monomers were represented by point-like particles jumping between sets of points on a regular grid (Lowry 1970; Doi and Edwards 1986; de Gennes 1971; Wall and Mandel 1975; Kremer and Binder 1988; Baumgärtner 1984; Gō 1983). With these lattice models structural detail is lost, but one can easily reach the time domain of protein folding (Skolnick and Kolinski 1990; Kolinski et al. 1991; Skolnick and Kolinski 1991; Rey and Skolnick 1991; Hinds and Levitt 1992; Šali et al. 1994). The protein model of the present approach considers all atoms, and its conformations are not restricted to a regular lattice.

The protein backbone model used here has a rigid bond geometry, which accounts for the bonded interactions, except the torsion potentials of the mobile torsional degrees of freedom. In the absence of non-bonded interactions it has been shown that the time decay of the end-to-end distance vector of the Rouse model (Verdier 1966; Verdier 1973; Orwoll and Stockmayer 1969) agrees with the Monte Carlo dynamics of the protein backbone model (Knapp 1992). From this comparison the time corresponding to a single scan of local MC moves (WM's) along the protein backbone chain is estimated to be about 15 ps (Knapp and Irgens-Defregger 1993). This time unit is four orders of magnitude larger than the step size of time propagation

used in conventional methods of dynamics simulation. On the other hand the smallest significant time unit for on-lattice Monte Carlo dynamics is of the order of ten to a hundred nanoseconds (Rey and Skolnick 1991). The present approach covers an intermediate time regime. Thus it can close the gap between the time regimes of conventional molecular dynamics (MD) simulations, which solve the classical equations of motion, and the time resolution of Monte Carlo dynamics (MCD) of simplified protein models on a lattice. Since an all atom model is used, a quantitative comparison with data from conventional MD simulations is facilitated considerably.

The paper is organized as follows. In the next section the protein backbone model and the Monte Carlo moves which generate the conformational changes in the protein backbone are introduced. Then the algorithm to generate the window moves (WM's) is described. One section is dedicated to the demonstration that WM's must be properly implemented to provide an unbiased distribution dihedral angle space. The energy function used for the MC simulation is derived from the energy function of the MD program CHARMM (Brooks et al. 1983). In the two final sections it is demonstrated that the MCD method folds polyalanine into conformations with a high α-helix content and the results are summarized. Technical details of the WM algorithm are explained in two appendices. An application of the method to the protein folding problem can be found elsewhere (Hoffmann and Knapp 1996).

Protein backbone model and Monte Carlo moves

The protein backbone is modeled as a chain of rigid amide planes (Knapp and Irgens-Defregger 1991; Knapp 1992; Rey and Skolnick 1993; Kurochkina et al. 1994; Knapp and Irgens-Defregger 1993). The backbone atoms of the same amide plane (C_α, C′, N, O, H) as well as the corresponding C_β-atom are placed according to the equilibrium bond geometry of the CHARMM22 (Brooks et al. 1983) parameter set. The dihedral angle ω, which describes the twist of an amide plane around the C′-N bond, is fixed at 180°. Thus the only degrees of freedom of the protein backbone are the dihedral angle ϕ and ψ at each C_α-atom.

New conformations are generated by Monte Carlo moves. The energy of the new conformation E_{new} is compared with the energy value of the former conformation E_{old} by using the Metropolis criterion (Metropolis et al. 1953)

$$p(\Delta E) = \min\left(1, \exp\left(-\frac{\Delta E}{k_B T}\right)\right), \tag{1}$$

where $\Delta E = E_{new} - E_{old}$. $p(\Delta E)$ is the probability to accept the new conformation. The decision whether to accept or reject the new conformation is accomplished by comparing $p(\Delta E)$ with a random number $r \in [0, 1]$. If $r > p(\Delta E)$, the new conformation is rejected and the old conformation is retained. If $r < p(\Delta E)$, the new conformation is accepted, i.e. the old conformation is replaced by the new conformation. As a result one obtains a Markov chain of conformations which represent a canonical ensemble at temperature T.

The probability, Eq. (1), can be introduced in a rate expression, if it is multiplied with a pre-exponential factor k_0 accounting for the energy barrier between state i (old) and state j (new). Downhill ($\Delta E < 0$), the reaction from state i to j proceeds with rate $k_{ij} = k_0$ and uphill ($\Delta E > 0$), it proceeds with rate $k_{ij} = k_0 \exp(-\Delta E/(k_B T))$. Assuming realistic moves leading from state i to j and a common value for the pre-exponential factor k_0 for all transitions between different states, the Markow chain represents a trajectory of a time evolution generated by Monte Carlo dynamics (MCD). The value of the pre-exponential factor k_0, which determines the time scale, is generally unknown. The time scale can be determined by comparing data from MCD computer simulation with results from analytical models (Knapp 1992), with data from key experiments such as formation of secondary structure and folding intermediates (Englander 1993; Radford et al. 1992; Miranker et al. 1993), or with benchmark simulations of conventional dynamics of model systems (Brooks 1989; DiCipua et al. 1990; Tirado-Rives and Jorgensen 1991; Daggett et al. 1991; Soman et al. 1991; Tobias et al. 1992; Daggett and Levitt 1992; Deloof et al. 1992; Pleiss and Jähnig 1992; Brooks and Nilsson 1993; Grønbech-Jensen and Doniach 1994).

Two types of MC moves are considered: simple moves (SM's) and window moves (WM's). The SM is a change of a single dihedral angle in the protein backbone. It generates a global conformational change, in that the positions of all atoms on one side of the rotation axis are changed relative to the atomic positions on the other side. The CPU-time is governed by the number of non-bonded energy terms, which are affected by a move. The CPU-time of a SM depends on the position of the C_α-atom to which the rotation axis belongs, within the polypeptide sequence. On average the CPU-time per SM is proportional to $n^2/6$, where n is the total number of amide planes of the protein molecule. For a SM whose rotation axis is close to the C- or N-terminus of the protein the amount of CPU time to evaluate the non-bonded interactions is smaller.

The WM consists of a cooperative change of several dihedral angles in a window of consecutive amide planes, that does not alter the protein backbone conformation outside the window. In contrast to SM's the WM's generate local conformational changes. The amount of CPU time for a single WM is proportional to $(\Delta m + 1)(n - \Delta m + 1)$ where n is the number of amide planes of the protein and Δm is the number of amide planes in the window. Note that for large proteins and small window sizes ($n > \Delta m$), the CPU time required for a WM is much smaller than for a SM.

For each SM the rotation axis, and for each WM the location of the window of amide planes is chosen randomly on the protein backbone. The values of the rotation angles are taken from an interval $[-\gamma, +\gamma]$ with fixed $\gamma \in [0°, 180°]$. A succession of MC moves, where each possible position

on the protein backbone is hit once on average is called a scan of MC moves (for short a "scan"). A scan represents the intrinsic time unit of MCD as does the step of time propagation for conventional MD computer simulations. With a sequence of several scans arbitrary conformational changes can be generated.

The window moves

A WM generates a local conformational change in a window of amide planes. For a window of Δm amide planes (corresponding to $\Delta m+1$ residues) $2(\Delta m+1)$ rotational degrees of freedom are involved. To keep the protein backbone structure outside the window unchanged, the rotation angle changes are correlated and have to fulfil six constraining conditions. Three of the conditions take care of the orientational invariance of the protein backbone structure outside the window. With a matrix representation of the $2(\Delta m+1)$ rotations in the window, two for each C_α-atom ($D_i^{(1)}$ and $D_i^{(2)}$), the condition of orientational invariance can be cast into the form of (Knapp and Irgens-Defregger 1991)

$$\prod_{i=1}^{\Delta m+1} D_i^{(1)} D_i^{(2)} = I, \tag{2}$$

where I is the (3×3)-matrix representing the identity transformation. The three other constraining conditions guarantee translational invariance which means that the distance vector connecting the C_α-atoms 1 and $\Delta m+1$ at the ends of the window remains invariant. With the vectors $\vec{r}_i, i=1, 2, \ldots, \Delta m$ connecting subsequent C_α-atoms in the window of amide planes, the condition for translational invariance reads (Knapp and Irgens-Defregger 1991)

$$\sum_{i=1}^{\Delta m}\left[\left(I-\prod_{k=1}^{i} D_k^{(1)} D_k^{(2)}\right)\vec{r}_i\right]=0. \tag{3}$$

Note that in the formulation of translational invariance the rotations $D_k^{(j)}$ are applied towards amide planes of rising index (towards the right end of the window). The rotations $D_{\Delta m+1}^{(1)}$ and $D_{\Delta m+2}^{(1)}$ at the right end apply to amide planes outside the window only. Therefore they need not to be considered in Eq. (3).

Since the number of rotational degrees of freedom is $2(\Delta m+1)$ for a window of Δm amide planes and since they are constrained by six conditions, there are $2(\Delta m-2)$ remaining degrees of freedom. Hence, the smallest window size with available rotational degrees of freedom, namely two, is a window of $\Delta m=3$ amide planes. To generate new conformations in a window of Δm amide planes, $2(\Delta m-2)$ rotations are chosen arbitrarily. The remaining six rotations are used to fulfil the constraining conditions, Eqs. (2) and (3). For windows located at the C- or N-terminus of the protein the rotations are directed towards this terminus. In this case no constraints are applied and the moves are called "end WM's".

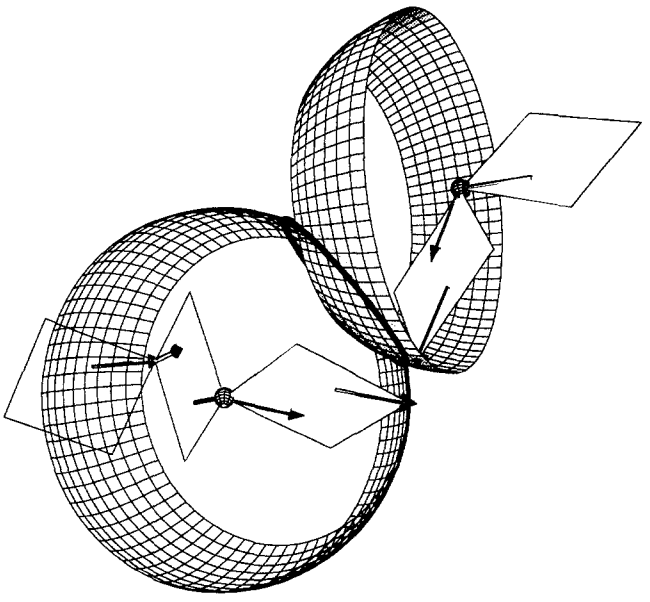

Fig. 1 Geometrical interpretation of the algorithm of window moves. The amide planes of the protein backbone model are shown as *rectangles*. The relevant rotation axes C_α-C′ and C_α-N are drawn as *arrows* directed towards the "cut". A configuration is shown where two pre-rotations have been applied towards the cut-C_α-atom. Thus the connectivity of the protein backbone at this atom is destroyed. The four rotations introduced by changes of the dihedral angles ϕ, ψ at the two joint C_α-atoms (*small spheres*) are used to restore the connectivity. By varying these dihedral angles the two halves of the cut-C_α-atom move on iris-shaped parts of spherical surfaces. Theses irises intersect on parts of the cut-circle (*bold line*). In a valid new conformation, both halves of the cut-C_α-atom are located at a point of the cut-circle, where the bond angle formed with the two neighbouring backbone atoms adopts its proper value

In practice the algorithm for the WM's proceeds as follows:

1. The available rotational degrees of freedom are eliminated by pre-rotations: Three of the $\Delta m+1$ C_α-atoms in a window of Δm amide planes are selected at random. The central of these three atoms is called "cut"-C_α-atom, the other to are called "left joint" and "right joint" C_α-atom, respectively (see Fig. 1). The four rotations at the left and right joint C_α-atoms are called "active" rotations. The two rotations at the cut-C_α-atoms are not explicity applied and are therefore called "implicit" rotations. At all other C_α-atoms in the window pre-rotations are performed, which means that there the dihedral angles ϕ and ψ are changed randomly. For a window of three subsequent amide planes there is only one C_α-atom where pre-rotations are performed. The active rotations are directed towards the cut-C_α-atom. They do not apply to amide planes beyond the cut-C_α-atom. As a consequence the connectivity of the protein backbone is destroyed at the cut-C_α-atom, which is split in a left and right cut-C_α-atom. Chain connectivity and geometry at the cut can possibly be repaired by using the remaining rotational degrees of freedom, i.e. the two active rotational degrees of freedom in the left and right joint, respectively. By introducing the cut, two of the six

unknown torsional degrees of freedom, the implicit rotations, have been eliminated.

2. Next two intersecting "irises" are introduced: By changing the dihedral angles ϕ_1 and ψ_1 in the left joint, the left cut-C_α-atom moves on the surface of a sphere. The same applies to the right cut-C_α-atom. If the two spherical surfaces intersect, the curve of intersection is part of a circle, the so-called cut-circle. The spatial positions of the cut-C_α-atom which restore the connectivity of the protein backbone must lie on this circle. As a consequence the manifold of where to search for new local conformations has become one-dimensional. The torsion axes $\vec{n}_1$ and $\vec{c}_1$ corresponding to the dihedral angle pair ϕ_1, ψ_1 are not orthogonal to each other. Hence the surfaces accessible to the left and right cut-C_α-atoms form "iris"-shaped sections of the corresponding spherical surfaces. The surfaces of the two irises intersect only in parts of the cut-circle. There the connectivity of the protein backbone can be restored.

3. Now the solutions on the relevant parts of the cut-circle must be generated: For this purpose the cut-circle is parameterized by a variable $\xi \in [-180°, +180°]$, the so-called cut-angle. The cut-angle is defined such that in the absence of pre-rotations the initial conformation belongs to the cut-angle $\xi = 0°$ (see appendix A for details). For a valid new conformation the backbone connectivity at the cut-C_α-atom is restored and the backbone bond angle at the cut-C_α-atom must assume its proper value γ_{bond}. This condition can be formulated in terms of a cut-function $f_c(\xi)$

$$f_c(\xi) = \vec{n}_c(\xi) \cdot \vec{c}_c(\xi) - \cos(\gamma_{bond}) \tag{4}$$

whose roots $f_c(\xi) = 0$ must be determined. In Eq. (4) $\vec{n}_c$ and $\vec{c}_c$ are unit vectors in direction of the bonds N-C_α, C_α-C′ at the cut-C_α-atom. Since $f_c(\xi)$ is a function in one variable only, all its roots can be found with standard numerical procedures (see appendix A). The cut-function possesses four distinct branches and up to four distinct definition regimes on the cut-circle, as explained in appendix A.

4. Once the values of the cut-angle ξ are determined as roots of the cut-function the valves of the dihedral angles of the new protein backbone conformation in the window are obtained by re-substitution of the found cut-angles ξ (see appendix A for details).

From the analogy of the WM's with the inverse kinematics of serial manipulators with six joints (Roth et al. 1973) it is known that the maximum number of roots of the cut-function is 16 (Primrose 1986; Manseur and Doty 1989). It is shown in the following, that this number can also be rigorously obtained by geometrical arguments. For the sake of simplicity the arguments use spheres instead of irises. The former is a special case of the latter.

For the argument, two facts concerning spheres are required:

(*i*) Variation of ϕ and ψ generates a double layered sphere. This can be understood most easily by comparison with the representation of a sphere in polar coordinates. The polar angles $\vartheta \in [-\pi/2; \pi/2[$ and $\varphi \in [-\pi; \pi[$ specify each point on a sphere uniquely and vice versa. The dihedral angles ϕ and ψ are both defined in $[-\pi; \pi[$ and therefore with this representation each point on a sphere is given by two pairs of ϕ and ψ, i.e. a sphere parametrized in ϕ and ψ is double layered. In other words, given e.g. a peptide with two amide planes, where the first amide plane is fixed in space and the second one movable by adjustments to ϕ and ψ at the central (i.e. second) C_α-atom, then for each possible position in space of the third C_α-atom, there are two conformations, given by two pairs of ϕ and ψ.

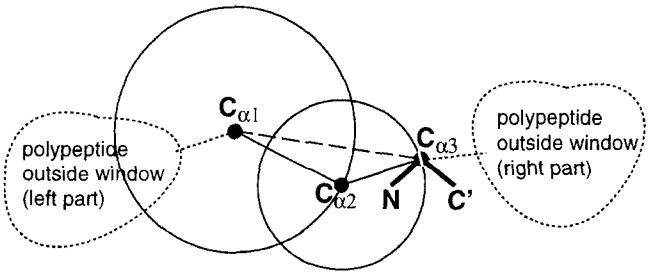

Fig. 2 Schematic drawing of the window used for the geometrical derivation of the maximum number of solutions. The first and second joint and the cut are placed in $C_{\alpha 1}$, $C_{\alpha 2}$ and $C_{\alpha 3}$, respectively. The two *circles* symbolize the spheres on which $C_{\alpha 2}$ and $C_{\alpha 3}$ move as the dihedral angles at $C_{\alpha 1}$ and $C_{\alpha 2}$, respectively, are varied. Valid conformations require the two halves of the cut-C_α-atom $C_{\alpha 3}$ to be united and the bond angle $\angle(N-C_{\alpha 3}-C')$ to assume its correct value. Rotations around the axis $C_{\alpha 1}$-$C_{\alpha 2}$ (*dashed line*) generate the one-dimensional manifolds discussed in the text

(*ii*) A unidirectional movement of a sphere causes space-fixed points to touch the spherical surface up to two times. This is obvious for translatory movements, but is also true for excentrical rotations of the sphere.

The set-up of the window used in this proof differs from that employed in the algorithm introduced above. The window consists of only two amide planes; this is the smallest and therefore simplest possible window with no degrees of freedom available for the pre-rotations (but the argument is valid for any window size). Here the joints are located in the first ($C_{\alpha 1}$) and second ($C_{\alpha 2}$) C_α-atom, both belonging to the left arm. The cut-C_α-atom ($C_{\alpha 3}$) is placed at the right end of the window (see Fig. 2). According to (*i*) the variation of the dihedral angles at the first C_α-atom yields a double layered sphere around this atom, on which the second C_α-atom is situated. The second C_α-atom is the center of another double layered sphere, on which the left half of the cut-C_α-atom is located. The right half of the cut-C_α-atom is fixed in space. Let us now recall the two conditions concerning the geometry at the cut-C_α-atom, which have to be fulfilled by valid conformations: (A) the two halves of this atom must be united, and (B) the bond angle N-$C_{\alpha 3}$-C′ at the cut-C_α-atom has to assume its proper value. Condition (A) is fulfilled for all combinations of dihedral angles at the first two C_α-atoms, that cause the sphere around the second C_α-atom to touch the spatially fixed right half of the cut-C_α-atom. Both spheres around the joints being double layered, each combination of dihedral angle fulfilling condition (A) is accompanied by a three

further combinations, i.e. in total there are four combinations of dihedral angles at $C_{\alpha 1}$ and $C_{\alpha 2}$ fulfilling condition (A). With (*ii*) the right half of the cut-C_α-atom, which is spatially fixed, touches the sphere around the second C_α-atom twice as it is moved around the first C_α-atom. This brings the number of combinations of dihedrals satisfying condition (A) from four to eight. A sphere around the second C_α-atom for which condition (A) can be fulfilled, i.e. a sphere on which the right half of the cut-C_α-atom ($C_{\alpha 3}$) is situated, can be rotated around an axis through $C_{\alpha 1}$ and $C_{\alpha 3}$ (dashed line in Fig. 2), without violating condition (A). This rotation generates a one-dimensional circular manifold from each of the eight combinations of dihedral angles. These circles are the baselines of eight cones around the rotation axis $C_{\alpha 1}$–$C_{\alpha 3}$, which are generated by the N-atom of the moving N-$C_{\alpha 3}$ bonds as the rotation is applied. The tips of the cones are located in the (re-united) cut-C_α-atom $C_{\alpha 3}$. On each of the one-dimensional manifolds condition (A) is satisfied. The remaining degree of freedom of these manifolds can be used to fulfil condition (B), namely the bond angle condition at the cut-C_α-atom. Condition (B) requires that the N-$C_{\alpha 3}$ bond is lying on a cone around the spatially fixed $C_{\alpha 3}$-C′ bond with the correct bond angle as semiangle. The solutions now correspond to the intersection of the eight cones with the cone around the $C_{\alpha 3}$-C′ bond. With the exception of singular points each intersection yields two solutions, which correspond to the crossing points of the baselines of the cones (appendix A, Eq. (18)). Thus in total there are a maximum of 16 solutions.

The cone around the $C_{\alpha 3}$-C′ bond does not necessarily intersect with each of the eight cones, so that the actual number of solutions can have any even value from zero to 16. In two singular cases this statement has to be modified. The first one is the case where the cone around the $C_{\alpha 3}$-C′ bond touches one of the eight cones without actual intersection; this leads to a twofold degeneracy of the corresponding solution. The second exception is the case where the cone around the $C_{\alpha 3}$-C′ and one of the other eight cones are identical, which leads to a one dimensional manifold of degenerate solutions, corresponding to a cut-function $f_c(\xi)$ which is zero for all values of ξ. The $C_{\alpha 3}$-C′ bond is then collinear with the axis through $C_{\alpha 1}$ and $C_{\alpha 3}$, and the conformation in the window can be freely moved as a rigid entity like a door around a hinge.

In practice the maximum number of real roots of the cut-function depends on details of the protein backbone model such as the geometry of the amide planes and the value of the bond angle at the cut (see Eq. (4)). The maximum number of solutions which has been observed so far for the present protein backbone model and windows of three amide planes is twelve. The actual number depends also on the start conformation and the pre-rotations in the window. The problem to find solutions for WM's can in principle be cast in a form, where the roots of a polynomial must be calculated (Manocha and Canny 1994). The polynomial is a function in tan (φ/2), where φ is the angle at a joint. Its degree is 16. All real roots of the polynomial are solutions of the problem of WM's. The real roots of this polynomial appear pairwise which means the number of solutions for WM's must be even. This is in accord with the conclusion drawn from the geometrical argument above and the fact that the cut-function consists of closed curves (see e.g. Figs. 3–5) and thus always has an even number of roots. An interesting consequence of this fact is, that an actual protein backbone conformation in a window has at least one other allowed conformational counterpart which can be reached without pre-rotations, i.e. for a window with an arbitrary number Δm of amide planes the values of the dihedral angles need to change only at three C_α-atoms (cut

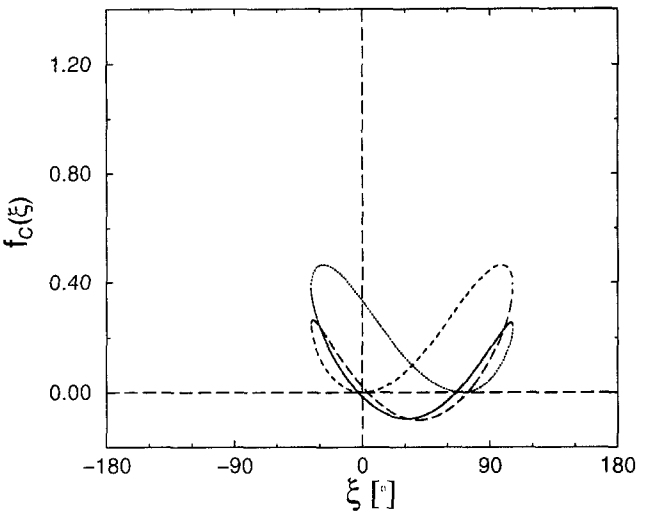

Fig. 3 The cut-function $f_c(\xi)$ for a window on a β-strand structure ($\phi_i = -120°$, $\psi_i = 120°$). The joints are located at the first (or N-terminal) and the third C_α-atom in a window of three amide planes. The pre-rotations, which are due to take place at the last C_α-atom in the window, are not carried out. Hence the root at $f_c(\xi = 0°)$ corresponds to the initial conformation

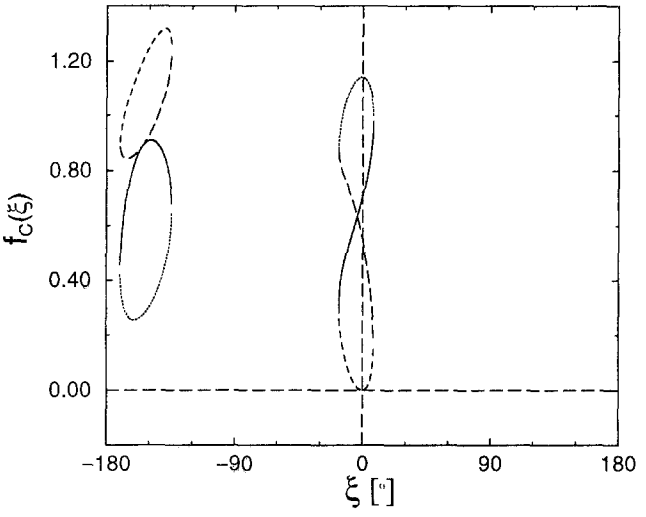

Fig. 4 The cut-function $f_c(\xi)$ for a window on an α-helical structure ($\phi_i = -57°$, $\psi_i = -47°$). As in Fig. 3 no pre-rotations are applied. $f_c(\xi)$ possesses only one accidental double root corresponding to the initial conformation

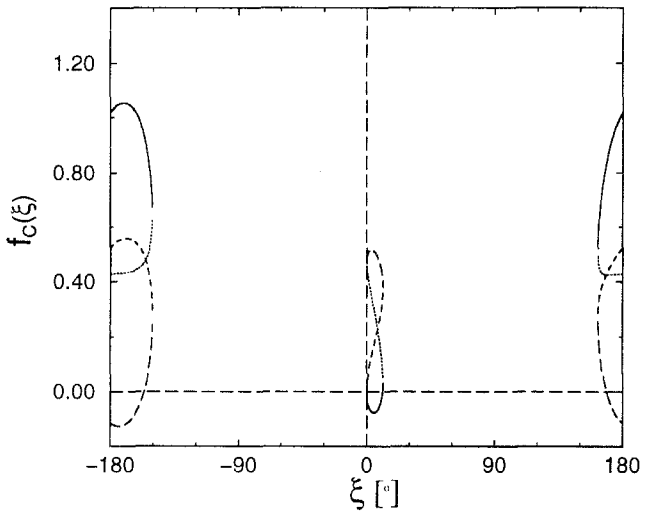

Fig. 5 The cut-function $f_c(\xi)$ for a window on a 3_{10}-helical structure ($\phi_i=-60°$, $\psi_i=0°$). As in Fig. 3 no pre-rotations are applied. $f_c(\xi)$ has four zeros. One is close to the root corresponding to the initial conformation. The other two roots are located on the opposite side of the cut-circle

and joints) to produce another conformation in the window.

The fact that the number of solutions is always even provides us with a method to assess the reliability of the root finding procedure. As outlined in appendix A, the root finding is carried out with a secant algorithm. With five secant trials per branch of $f_c(\xi)$ and per interval on which $f_c(\xi)$ is defined, one finds in 3.6% of all WM's an odd number of solutions. Assuming that only one solution was missed in these cases, and given that the mean number of possible solutions per WM (for windows comprising three amide planes) is 3.7, one arrives at less than 1% of missed solutions relative to the total number of solutions. The average time for the root finding with this accuracy is less than 10 ms per WM on a SGI Indigo computer with a R4000 CPU.

The cut-function $f_c(\xi)$, Eq. (4), is displayed for three typical situations in Figs. 3 through 5. In all three cases displayed, the pre-rotations vanish so that $\xi=0°$ is a solution of the cut-equation (Eq. (4)). The cut-function is displayed for the case where the axes of the (vanishing) pre-rotations are located at the C-terminal C_α-atom of the window. Hence the two arms connecting the left and right joint C_α-atom with the cut-C_α-atom are diagonal vectors in a single amide plane located between the cut- and the corresponding joint C_α-atom.

In Fig. 3 the conformation at $\xi=0°$ corresponds to a β-strand, i.e. $\phi_i=-120°$, $\psi_i=+120°$. Note that there are several solutions close to the initial conformation at $\xi=0°$. Such small variations in the cut-angle ξ imply that the corresponding conformations are close to the initial conformation. Close to $\xi=0°$ one branch of the cut-function has nearly zero inclination over a relatively large interval of the cut-angle. This is almost a double root of the cut-function which enlarges the flexibility of the β-strand conformation. This allows one to leave the initial conformation easily by WM's and renders the β-strand conformation relatively flexible. The situation is quite different, if the initial structure in the window is an α-helix (Fig. 4). Without pre-rotations the initial configuration nearly represents a double root. Those are the only solutions available. In this case the cut-function touches the zero line only in a very narrow regime of the cut-angle. This suggests that the α-helix is relatively inflexible compared to the β-strand. Under the given circumstances the α-helix is so to speak topologically trapped. The situation changes as other C_α-atoms are chosen for the pre-rotations or the pre-rotations are applied.

In Fig. 5 the cut-function is displayed for a 3_{10}-helix whose hydrogen bonds connect amide planes one unit closer in the sequence as for α-helices. The gross features of the cut-function resemble those of an α-helix. However, in contrast to an α-helix the cut-function of a 3_{10}-helix possesses four roots. Two of them belong to cut-angles whose values are close to $\xi=180°$, opposite to the initial conformation at $\xi=0°$. Here the double root of the α-helix is resolved in a pair of roots: one remains at $\xi=0°$ the other is in close proximity. The new conformation corresponding to the root with small cut-angle is close to the initial conformation.

Correction of the dihedral angle distribution

In the absence of non-bonded interactions the dihedral angles of the protein model should be uniformly distributed. This is also a necessary prerequisite for ergodicity in the case where the protein model is used with all non-bonded interactions. It is obvious that SM's produce a uniform distribution of dihedral angles. The same is certainly true for end WM's. For inner WM's this is a priori not clear. In fact it was shown recently (Dodd et al. 1993) that a similar type of MC-move does not generate evenly distributed dihedrals. As will now be shown, this applies to WM's, too. But it will also be outlined how to correct the WM algorithm, so that it generates a uniform distribution of dihedrals. The details of the correction procedure are described in appendix B.

The non-uniformity of the dihedral angle distribution produced by WM's can be quantified with the following computer experiment. The conformational space of a 4-mer, i.e. a polypeptide of four residues (five amide planes), whithout non-bonded interactions is sampled with WM's. The window size is set to three amide planes. In each scan we first apply a left end WM, then the only possible inner WM, and finally a right WM. In the end WM's dihedral angles are allowed to change randomly by any value out of $[-180°; 180°[$, this ensures that each inner WM is applied to a completely randomized conformation. The end WM's alone produce an even distribution of dihedral angles, any deviation from uniformity must therefore be caused by the inner WM's. In the inner WM's dihedrals may change by up to $\pm180°$ per move. The C_α-atom se-

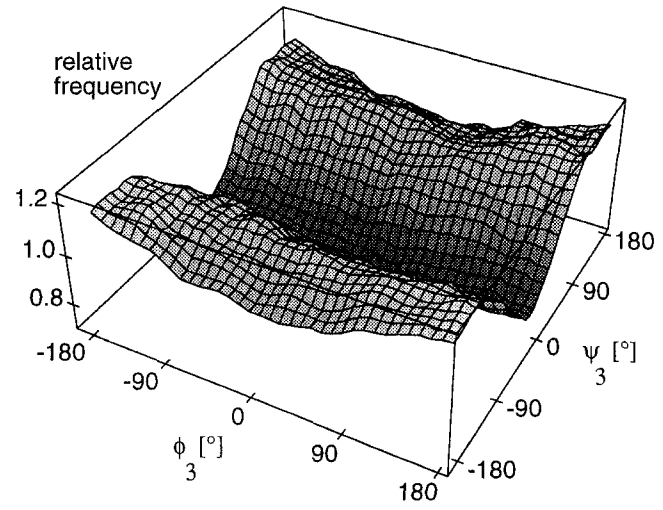

Fig. 6 Density of states of a peptide of five amide planes without non-bonded interactions, as generated by uncorrected WM's. The data were produced by starting 10^6 times from random conformations and applying a single WM without pre-rotations. Each WM leads to a new conformation, which is then stored. The distribution of dihedral angles is project on the dihedral angles ϕ_3, ψ_3 at the third C_α-atom, which in this simulation always plays the role of the cut-C_α-atom. The figure shows in the form of a two dimensional histogram the occupation numbers of 36^2 squares of 10° side length, relative to the average occupancy. The values are smoothed by a running average over each square and its eight neighbours

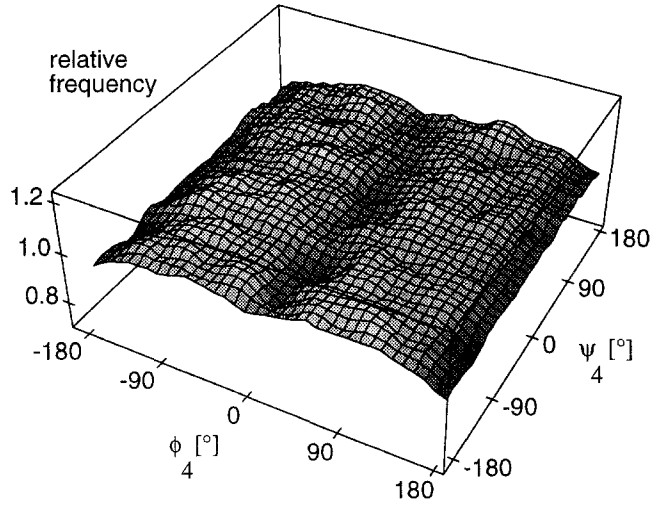

Fig. 7 Density of states of a peptide of five amide planes projected on the dihedral angles ϕ_4, ψ_4 at the fourth C_α-atom. This atom located in the simulation the right joint. For further details refer to Fig. 6

lected for the pre-rotation is always the second C_α-atom, but the pre-rotations are set to zero. The latter guarantees the presence of at least two possible conformations for each inner WM, namely the conformation existing before the WM is carried out, and a conjugated solution of the WM-algorithm (see previous section). Thus the probability of acceptance is unity for inner WM's, and the deviation of the dihedral angle distribution from uniformity is not blurred by the uniform contribution of the end WM's. The outcome of the experiment is shown in Figs. 6 and 7. It is found, that the conformations are non-uniformly distributed in four of the eight dihedrals in the three amide plane window: the ψ-angle at the left joint (ψ_l), the two dihedrals at the cut (ϕ_c, ψ_c), and the ϕ-angle at the right joint (ϕ_r). A histogram of the frequencies of occurrence 4-tupels (ψ_l, ϕ_c, ψ_c, ϕ_r) reveals that frequencies can differ by more than a factor of three, which corresponds to a difference in entropy of about k_B, or in free energy of about k_BT. The differences between free energies of various conformations due to this bias will increase in proportion to the length of the polypeptide and the bias will dominate the free energy landscape as the temperature rises.

Before discussing the reasons for the bias and the means to remove it, we now introduce a sensitive method to quantify the deviations from an even distribution. The conformational space spanned by the four non-uniformly distributed dihedrals ψ_l, ϕ_c, ξ_c, and ϕ_r is divided into $m = 9^4 = 6561$ four-dimensional cubes of 40° side length. The number of MC scans of the kind described above is $N = 10^6$. If the conformations are uniformly distributed in dihedral angle space, the probability $p_N(n)$ of one of the m cubes containing n conformations is given by the binomial distribution

$$p_N(n) = \binom{N}{n} q^n (1-q)^{N-n}, \tag{5}$$

where $q = 1/m$. p_N can be compared directly with the normalized frequency $P(n)$ from the simulation data, i.e. the number of cubes containing n conformations, divided by the total number m of cubes. The normalization of $P(n)$ reads

$$\sum_{n=0}^{N} P(n) = 1. \tag{6}$$

A direct comparison of $P(n)$ and $p_N(n)$ is problematic, because the number of cubes with given number of conformations is small, and thus $P(n)$ too noisy. We therefore average the values of $P(n)$ over all $n \in \{(j-1)\cdot\Delta n, \ldots, j\cdot\Delta n\}$ ($j = 1,2,\ldots,N/\Delta n$). Besides we scale $P(n)$ in a more convenient way by transforming its argument from n to $n/\langle n\rangle$; this makes the first moment of P independent of N and m by moving it from $\langle n\rangle = N/m$ to 1. The scaled function is normalized to unity by multiplying it by $\langle n\rangle$. In summary a normalized and averaged probability density $\bar{P}(n/\langle n\rangle)$ is obtained:

$$\bar{P}\left(n/\langle n\rangle\right) = \frac{\langle n\rangle}{\Delta n} \sum_{n'=(j-1)\cdot\Delta n}^{j\cdot\Delta n} P(n'), \tag{7}$$

$$\forall n \in \{(j-1)\cdot\Delta n, \ldots, j\cdot\Delta n\}$$
$$(j = 1,2,\ldots,N/\Delta n),$$

which, with $n_r \equiv n/\langle n\rangle$, fulfils

$$\int \bar{P}(n_r)\,dn_r = \int n_r \bar{P}(n_r)\,dn_r = 1. \tag{8}$$

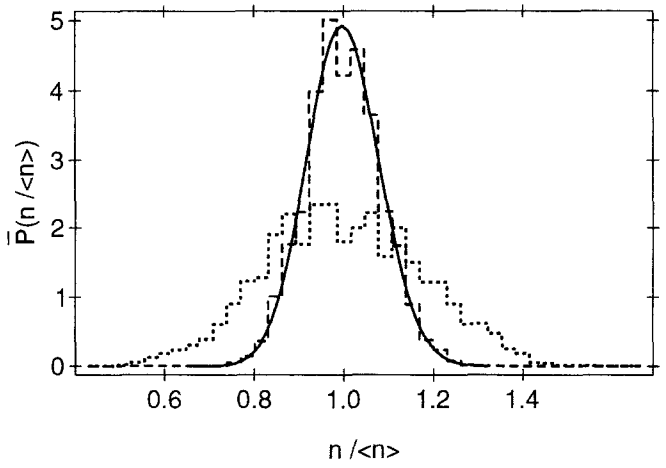

Fig. 8 Histogram function $\bar{P}$ (Eq. 7) produced by uncorrected (*dotted line*) and corrected (*dashed line*) WM's, compared to the theoretical curve for a uniform distribution (*solid line*)

The probability density $\bar{P}(n/\langle n\rangle)$ can now be compared with the correspondingly rescaled binomial distribution $\langle n\rangle\, p_N(n/\langle n\rangle)$, as shown in Fig. 8. The figure clearly demonstrates, that a priori the window algorithm yields a distribution of conformations in dihedral angle space, which markedly deviates from a uniform distribution. This biased distribution is contrasted with a distribution obtained from a simulation, which uses the window moves with a correction introduced in the remainder of this chapter and appendix B.

Some of the dihedrals are uniformly distributed, others are not. The explanation for the non-uniform distribution lies in the general fact that correlations between different interacting parts of a system can lead to structures in the distributions of quantities which characterize the system. Here the correlations are invoked by the constraints which keep the protein backbone conformations outside the window invariant and the characteristic quantities are the dihedral angles in the window. But, as already mentioned, not all of the dihedral angles are non-uniformly distributed. Those dihedrals which are changed by the pre-rotations are uniformly distributed because they are not subject to any constraint. The ϕ-angle at the left joint (ϕ_l) and the ψ-angle at the right joint (ψ_r) are also evenly distributed. This can be seen from the following argument. Let us assume, that the two joints are located at the first and last C_α-atoms of the window (but the argument is also valid in all other cases). The angles ϕ_l and ψ_r, with their axes located in amide planes outside the window, describe the orientation of inner parts of the window (which are constrained by Eq. (2) and Eq. (3)) relative to the parts of the polypeptide outside the window (which are not affected by the window constraints). Since the orientations of the latter part are uniformly distributed for each window conformation, ϕ_l and ψ_r are also uniformly distributed. This also means that two of the six constraints contained in vector equations Eq. (2) and Eq. (3) have no effect on the distribution of dihedrals. The remaining four constraints lead to four non-uniformly distributed dihedral angles, ψ_l, ϕ_c, ψ_c, and ϕ_r as already mentioned above, independently of window size and positions of joints and cut. The four constraints can be recast into

$$c_i\left(\psi_l^{(k)}, \phi_c^{(k)}, \psi_c^{(k)}, \phi_r^{(k)}\right) = const; \qquad (i = 1,2,3,4;\ k = 1,2,\ldots,s), \tag{9}$$

where the c_i represent the four constraint quantities describing the shape of the window (see appendix B, Eqs. (21) through (24)), and s is the number of solutions for the given set of c_i's. If we carry out four independent infinitesimal perturbations of the c_i's, the dihedral angles ψ_l, ϕ_c, ψ_c, and ϕ_r must also change. These changes of dihedrals span four-dimensional volume elements $d\psi_l^{(k)} \cdot d\phi_c^{(k)} \cdot d\psi_c^{(k)} \cdot d\phi_l^{(k)}$, i.e. each solution can be considered to occupy a certain volume in phase space which can be expressed as the Jacobian

$$J_k = \left| \frac{\partial\left(\psi_l^{(k)}, \phi_c^{(k)}, \psi_c^{(k)}, \phi_r^{(k)}\right)}{\partial(c_1, c_2, c_3, c_4)} \right|, \quad k = 1,2,\ldots,s. \tag{10}$$

It is clear that within a set of solutions each solution should be weighted according to its volume J_k to achieve a uniform distribution of conformations in dihedral angle space. In practice there are two groups of solutions in a WM, one group of N_p solutions accessible after the pre-rotations are carried out at the C_α-atoms chosen for this purpose, and a second group of N_o solutions accessible for vanishing pre-rotations. The weight of each of the two groups is proportional to the sum of the J_k of its respective members. Within each of the two groups the weight of a single solution is proportional to its J_k a proper weighting of solutions in one step can be acommplished by randomly selecting one out of all pre-rotated and not pre-rotated solutions with probability

$$p_i = J_i \left(\sum_{k=1}^{N_o+N_p} J_k \right)^{-1}, \quad i = 1,2,\ldots,N_o + N_p. \tag{11}$$

It is obvious that the weighting of conformations according to Eq. (11) fulfills the detailed balance criterion. Figs. 8 and 9 show that the WM's corrected by this weighting scheme generate a uniform distribution of dihedral angles.

As pointed out earlier, the non-uniform distribution of dihedral angles was recently described by Dodd et al. (1993) for a MC move similar to the WM introduced in this work, which was used to sample polymer conformations. Dodd et al. also propose a method to eliminate the bias which uses Jacobians in a weighting scheme. Their Jacobian is based on a 6×6-matrix, whereas the Jacobian in Eq. (10) is generated from a 4×4-matrix. In practice even more important is that Dodd et al. carry out the weighting with a rejection procedure in analogy to the Metropolis algorithm. The selection procedure based Eq. (11) is more efficient because the probability to select a new conformation is much higher than with a rejection procedure.

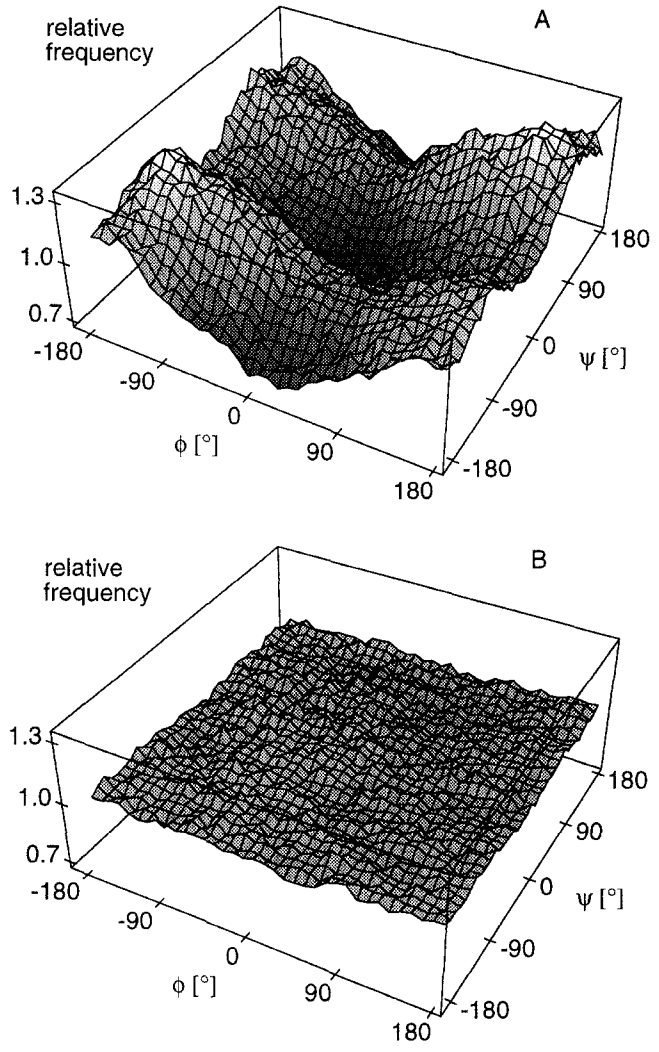

Fig. 9 ϕ, ψ distribution from simulations of a 18-mer with uncorrected (**A**) and corrected (**B**) WM's. The landscapes are obtained by averaging the ϕ, ψ distribution at C_α-atoms 4–15. The window size is three amide planes. The data are based on a simulation of $4 \cdot 10^5$ MC scans. The pre-rotations are always carried out at the second C_α-atom of a window. Only those WM's are accepted, in which no dihedral angle changes by more than 20°. For a description of the axes see also Fig. 6

The energy function

In order to facilitate the comparison of MC and MD simulation data, we use an energy function, which is derived from a force field applied in MD simulations, in a way similar to that proposed by Gerber (1992). Conventional MD energy functions cannot be used directly, because the relatively rigid protein backbone model employed in the present work would give rise to frequent atomic clashes, and thus would severely slow down the dynamics (van Gunsteren and Karplus 1981). In standard MD simulations with a fully flexible protein model these clashes are efficiently prevented by small amplitude motions of bond lengths and bond angles. In this sense Karplus has stated that high frequency small amplitude modes of motion serve as a "lubricant" for the slow, large amplitude dihedral modes of motion (Karplus 1986). The most severe atomic clashes can be attributed to atom pairs from sequentially adjacent amide planes. These can effectively be avoided by incorporating some of the flexibility in interaction terms between neighbouring amide planes. In the present protein model, atoms in sequentially adjacent amide planes can only be moved relative to each other by changing the two dihedral angles at the common C_α-atom. Therefore a discrimination of the bonded torsion interaction terms, the non-bonded Lennard-Jones and electrostatic interaction terms between these amide planes is no longer useful in the polypeptide model proposed in the corresponding section above. Instead, a generalized two-dimensional dihedral angle energy landscape is used. Furthermore the rigidity of the amide planes of the protein backbone makes it superfluous to consider interactions between atoms of the same amide plane. No cut-off scheme for non-bonded interactions is employed.

The dihedral angle energy landscape is derived from the CHARMM22 energy function by considering an alanine dipeptide of two amide planes, as was demonstrated in appendix 2 of Brooks et al. (1983). The C_α-atoms at the ends of this dipeptide are replaced by methylgroups to avoid the charged groups which normally terminate a polypeptide. To calculate the value of the dihedral angle energy landscape for a specific dihedral angle pair (ϕ, ψ) the following steps are performed:

1. The conformation of the dipeptide is generated for the dihedral angle pair (ϕ,ψ) with ripid amide plane geometry.
2. An energy minimization is carried out with the MD energy function, where the amide planes are fully flexible but the dihedral angle pair (ϕ,ψ) is fixed.
3. The full MD energy function is evaluated for the final conformation.

The energy landscape is tabulated for a two-dimensional grid with a resolution of 10° and can be evaluated at arbitrary angles by interpolation.

In summary the energy function used is the CHARMM22 energy function (Brooks et al. 1983), except for interactions between atoms in neighbouring peptide planes. The latter energy terms are calculated from the tabulated values of the torsion energy landscape at the given dihedral angle pairs (ϕ_i, ψ_i).

Application

In a first application we compare SM's and WM's with respect to their abilities to find equilibrated structures of polypeptides. As a model system we have chosen 18-alanine, which is still quite simple and small but nevertheless can already develop globular-like conformations with a high secondary structure content. More detailed analyses of trajectories and results for hetero-polypeptides are presented elsewhere (Hoffmann and Knapp 1996).

Three trajectories were produced with each of the two types of moves: the trajectories generated with SM's are called S(i) and those generated with WM's are called W(i), (i=1,2,3). All simulations started from the maximum extended all-trans conformation. The simulation data within each of the two groups of trajectories S(i) and W(i) differ only in the seeds with which the random number generator was initialized. The random number generator used is the subroutine ran2 from Press et al. (1992). The simulations were carried out for the isolated polypeptide, i.e. in vacuum, at a temperature of 400 K. The termini of the polypeptide were blocked with uncharged methyl groups to avoid the strong electrostatic attraction of the charged or polar amino- and carboxyl-ends, which usually terminate polypeptides.

The dihedral angles are allowed to change by at most ±20° per WM, which guarantees a gradual evolution of conformations in time. The maximum change of dihedral angles per SM is restricted to ±10°. Here the dihedral changes are smaller to avoid atomic clashes, which occur more frequently for SM's because of the global character of these moves. Thus the acceptance probability in the Metropolis algorithm lies for SM's at about 0.34 to 0.39 (in contrast to an acceptance probability of 0.53 to 0.60 for WM's with dihedral changes of a maximum of ±20°).

Each of the trajectories W(i) comprises 5×10^5 scans with a window of three amide planes. The SM trajectories are shorter, containing only $2.5\cdot10^5$ scans. The reason for the difference is, that in a polypeptide of N monomers the average number of non-bonded energy terms to be evaluated (and thus the CPU-time per per scan) in a SM scan is proportional to $N^3/3+O(N^2)$, whereas in a WM scan it is proportional to $3N^2+O(N)$. The CPU-time needed for the generation of the moves is negligible compared to the time spend on the evaluation of the energy function. This difference in the amount of CPU-time for SM and WM scans accounts for the factor of two in the length of the trajectories. It allows one to compare SM and WM trajectories that needed about the same amount of CPU-time for their generation.

Figure 10 shows conformations which have been obtained by energy minimization of the conformations of lowest energy in each trajectory. For each trajectory the minimization was carried out by 5000 scans of WM's at a temperature of 0 K and a maximum dihedral angle change per WM of 5°. The minimization leads to a drop in energy of 3–4 kcal/mol for each of the W(i) and 5–6 kcal/mol for each of the S(i) conformations. Without minimization the ranking of minimum energies is given by $E_{min}(W(i))<E_{min}(S(i))$, i=1,2,3 and $E_{min}(S(i))<E_{min}(W(i+1))$, i=1,2. After the minimization the ranking remains the same, except for the S(3) and W(3) which have interchanged their positions. The conformations generated by WM's have a tendency to develop a higher content of α_R-helix than the conformations produced by SM's. The conformations of lowest energy in trajectories W(1), W(2) and W(3) are a compact rudimentary helical hairpin, an almost perfect α_R-helix, and a piece of α_R-helix followed by a kink and a left handed helical turn at the C-terminus, respectively. Viewed as a whole the conformations of lowest energy of the SM trajectories seem to have a lower helix content and shorter stretches of α_R-helix. All three structures from the SM trajectories are globular. One feature in particular is shared by these globular structures obtained by SM's: the two termini are linked by hydrogen bonds, which make the conformations quasi-cyclic. The global character of conformational changes generated by SM's requires the termini to be freely movable without having a overcome a larger energetic barrier such as, for example, the breaking of hydrogen bonds. Therefore SM trajectories are often trapped in these quasi-cyclic conformations. WM trajectories may also show quasi-cyclic conformations, but since WM's do not require freely movable chain ends, these conformations do not inhibit further dynamics.

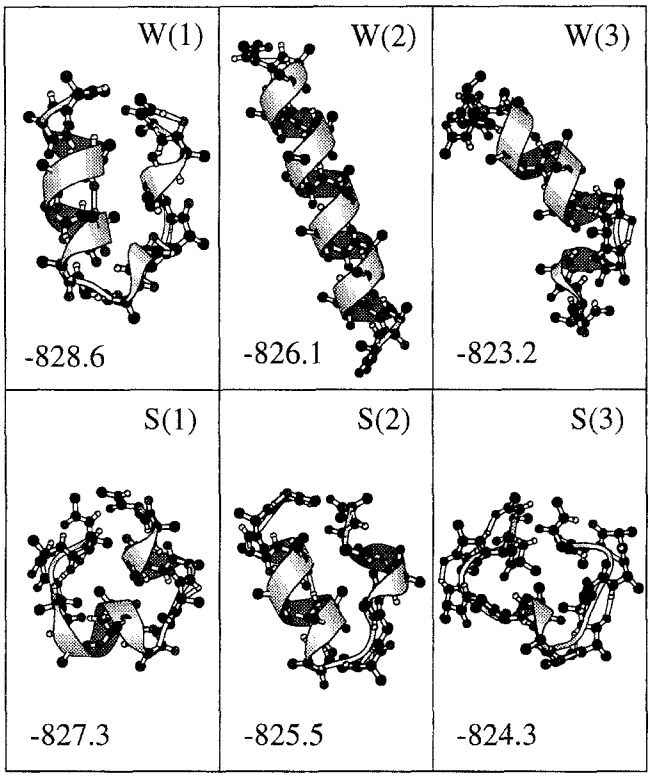

Fig. 10 Energy minimized conformations from the conformation of lowest energy in each of the trajectories W(i), S(i), i=1, 2, 3, of 18-alanine. Conformations were extracted from the trajectories and energy minimized by 5000 WM scans, with a maximum dihedral angle change per WM of 5°. The numbers in the figure indicate energies in kcal/mol after the minimization. The plots where produced with MOLSCRIPT (Kraulis 1991). Ribbons are helical turns recognized by MOLSCRIPT (at least three α_R-helical C_α atoms in a row). The N-terminus lies in the upper left corner in all plots

A more quantitative comparison of the minimum energy structures is shown in Fig. 11. Two quantities measuring differences between structures are depicted in this figure. The root mean square deviation (RMSD) of the C_α-atom coordinates of the structures after pairwise optimal superposition (Kabsch 1976) (C_α-RMSD), and the RMSD of dihedral angle vectors (ϕ_1, ψ_1, ..., ϕ_{18}, ψ_{18}) (dihedral

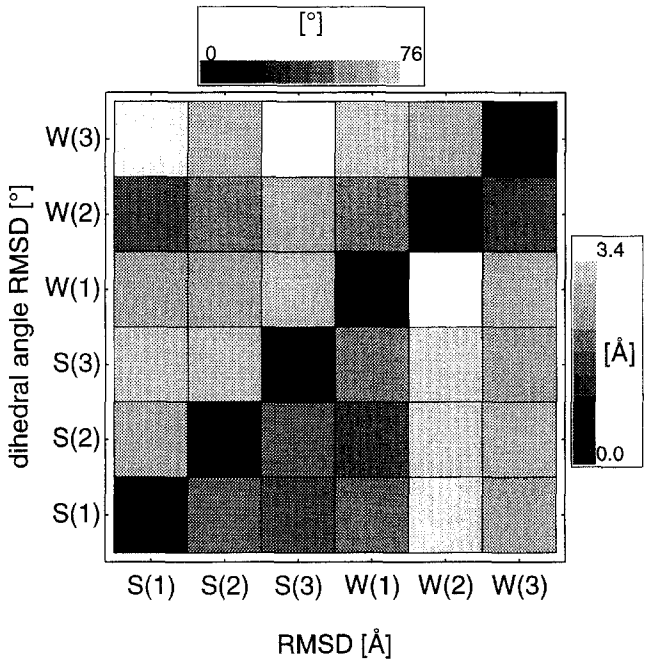

Fig. 11 Pairwise comparisons of conformations from Fig. 10 using spatial and dihedral angle RMSD's. The spatial RMSD's (*lower right triangle*) are calculated for the C_α-atom coordinates after optimal superposition with an algorithm of Kabsch (1976). The dihedral angle RMSD's (*upper left triangle*) are evaluated with respect to the backbone dihedrals ϕ_i and ψ_i, ($i = 1, 2, \ldots, 18$). The gray scales were obtained by dividing the differences between the maximum and minimum of all pairwise RMSD's and dihedral RMSD's, respectively, into ten intervals of equal size

Table 1 Averages of energy, distance of first and last C_α-atom and radius of gyration with respect to the C_α-atoms. The data are mean values for the last tenth of each trajectory W(i), S(i), $i = 1,2,3$. The numbers in parentheses are the standard deviations. In the last column the prevailing character of the structures in these parts of the trajectories is indicated ($\alpha = \alpha_R$-Helix, t = reverse turn)

Trajectory	Energy [kcal/mol]	Distance $C_\alpha(1)$-$C_\alpha(18)$ [Å]	C_α-radius of gyration [Å]	Structural motif
W(1)	−812.5 (3.9)	6.4 (0.5)	5.9 (0.1)	$\alpha\alpha$
W(2)	−810.0 (4.2)	27.0 (0.8)	8.4 (0.1)	α
W(3)	−803.4 (6.4)	19.8 (2.1)	7.5 (0.3)	α+coil
S(1)	−810.3 (3.3)	5.6 (0.3)	5.7 (0.2)	α+coil
S(2)	−807.3 (4.1)	5.8 (0.8)	6.0 (0.1)	α+coil
S(3)	−803.3 (4.3)	5.4 (0.4)	6.0 (0.1)	coil

RMSD). The latter quantity is a measure of the average local similarity of corresponding residues, whereas the former measures the global similarity of the folds. Figure 11 supports the qualitative observations made above. A surprising finding is the small dihedral RMSD of the structures from W(2) and S(1); the content of helical ϕ and ψ angles is obviously larger in S(1) than expected from Fig. 10. The reason for this discrepancy is that the conformation from S(1) not only has the two helical stretches indicated by the ribbons in Fig. 10 but also contains some singular residues in the α_R-helix region. Figure 11 supports the finding that the minimum energy conformations of the S(i) trajectories are globular, since the pairwise C_α-RMSD's are all small (dark gray squares). The local structures given by the dihedral angles are obviously more diverse, as indicated by the higher dihedral RMSD's. The minimum energy structure in W(1) is the most globular structure of the three conformations obtained from the WM trajectories. Its C_α-RMSD to the S(i) conformations lies within the distribution of pairwise C_α-RMSD's of S(i) conformations. The same is also true for the dihedral RMSD's. The minimum energy conformation of W(2) is an α_R-helix and is therefore quite extended. Hence the C_α-RMSD to all other structures, except perhaps that from W(3), is quite large (3 Å or more). Note however that for longer polypeptides or proteins a C_α-RMSD of 3 Å would not be considered large, whereas for smaller polypeptides this value is typical for dissimilar structures. The minimum energy conformation of W(3) is rather elongated, which is reflected by the relatively low C_α-RMSD to the helix of W(2). The W(3) structure is remarkable in that it has on the average the largest dihedral RMSD to all other minimum energy conformations. This is due to its kink and its α_L-helical turn at the C-terminus.

Table 1 shows that the observations made for the minimum energy conformations are also representative of averages over the last 10% of each trajectory. The table shows that the ranking of average energies $\bar{E}$ is the same as for the energies of the non-minimized conformations of lowest energy mentioned above. W(3) is remarkable again because it shows the largest fluctuations of energy, end-end-distance and radius of gyration. This trajectory is obviously not yet fully equilibrated. The small distances of the terminal C_α-atoms in trajectories S(i) illustrate again that the conformations are quasi-cyclical. None of the WM trajectories develops quasi-cyclic conformations. In W(1) conformations are as compact as in the S(i) trajectories, as can be seen from the small radius of gyration. Nevertheless the end-end-distance is significantly larger than in the S(i), which indicates that even the relatively compact structure from W(1) does not become quasi-cyclical. The fluctuations of the tabulated quantities suggest that all trajectories, except W(3) are equilibrated.

In summary WM's tend to generate conformations with lower energies than SM's. SM's clearly prefer quasi-cyclical and compact structures. WM's sample both open structures with high helix content, and compact globular conformations.

Conclusions

A MC method for the time evolution of an off-lattice all-atom protein backbone model has been developed. The small amplitude, fast modes of motion (bond angle bending and bond stretching) are avoided by considering amide planes, bond lengths and bond angles to be rigid. A MD energy function adapted to these constraints is used.

Two types of MC moves have been employed. The simple move (SM) is a change applied to an individual backbone dihedral. This move is easy to implement but leads to large atomic displacements far away from the corresponding rotation axis. As a consequence the acceptance of SM's is small for conformations of globular polypeptides. SM's have the tendency to lead too fast to a condensed protein backbone structure in which the molecule then becomes trapped. SM's have problems to generate longer α-helical stretches.

More efficient off-lattice moves, which are closer to realistic motions occurring in the conventional MD simulations, are the window moves (WM's). These moves are cooperative changes of dihedral angles in a window of amide planes, which do not change the conformation of the polypeptide outside the window. WM's mimic the local moves often used for polymer models defined on regular lattices. However, the WM's occur, so to say, on a irregular and dynamic lattice, which is very flexible and can therefore model protein conformations with high accuracy. The implementation of the local off-lattice moves is more complicated than that of the on-lattice moves.

The distribution of dihedral angles obtained with a series of WM's is uneven if no specific precautions are taken. This is due to the constraints which must be fulfilled to obtain a local move from a combination of dihedral angle changes. The constraints lead to differences in the phase space volume available for different window geometries and backbone conformations in the window. In the present treatment this bias is corrected by using suitable Jacobians.

Quite generally off-lattice MC simulations using more complicated moves, which are subject to constraints, will suffer from the same problem. On a simple lattice, where all moves are topologically equivalent no such problem exists. Complex lattice models are used to model the protein backbone structure in more detail (Skolnick et al. 1993). In these cases the moves may correspond to different topologies. Then care must be taken to avoid a bias in the distribution of conformations for on-lattice moves, too.

The off-lattice MC methods has been applied to 18-alanine to generate three trajectories for SM's and WM's. The structure with the lowest energy value obtained by WM's is a compact conformation with a high α-helix content. The structures obtained with SM's are all compact but have only shorter stretches of α-helix. The superiority of WM's over SM's, i.e. their higher acceptance probability in globular conformations and efficiency in sampling the conformational space, as well as their ability to generate more realistic movements, will become more obvious when larger polypeptides are considered (see for instance Ref. (Hoffmann and Knapp 1996)).

Acknowledgements The authors gratefully acknowledge the help of Fredo Sartori with the force field. The CHARMM source code was provided by Prof. Dr. Martin Karplus and Molecular Simulations Inc. This work is supported by the European Union contract ERBCHRXCT930112, the Deutsche Forschungsgemeinschaft, SFB312, project D7 and project Kn329/1 as well as the Fonds of the Deutsche Chemische Industrie.

A Generation of new local conformations

This section describes the implementation of the window algorithm in greater detail. Figure 12 displays a possible situation for a model window of three amide planes in order to illustrate the terminology. The algorithm presented works for all window sizes of two or more amide planes.

First the pre-rotations are carried out (see above section about window moves). They destroy the connectivity of the polypeptide at the "cut", where the C_α-atom is split into two parts, the left and right cut-C_α-atom. The connectivity of the polypeptide and the integrity of the cut-C_α-atom are restored along the intersection of the two irises (Fig. 1). A necessary condition for an intersection of the two irises is the intersection of the corresponding spherical surfaces, i.e. the sum of their radii must be larger than the distance between the centers of the two spheres:

$$R_l + R_r > D. \tag{12}$$

R_l is the length of vector $\vec{R}_l$ from left joint to the left cut-C_α, R_r is the length of vector $\vec{R}_r$ from right joint to the right cut-C_α, and D is the length of vector D from left joint to right joint (l,r and c index vectors originating in or pointing to the left joint, right joint and cut-C_α-atom, respectively).

If inequality 12 holds, the spherical surfaces intersect in the "cut-circle". The cut-circle is located in the plane perpendicular to $\vec{D}$, its radius is $R_c = \sqrt{R_l^2 - D_l^2}$ and its center measured from the left joint is given by the vector

$$\vec{D}_l = \frac{D^2 + R_l^2 - R_r^2}{2D^2}\vec{D}. \tag{13}$$

The cut-circle, seen from the left joint, can be parametrized as function of an angle ξ:

$$\vec{R}_l(\xi) = \vec{D}_l + R_c(\vec{x}_0 \cos\xi + \vec{y}_0 \sin\xi), \tag{14}$$

where $\vec{x}_0$ and $\vec{y}_0$ are two orthonormal vectors in the plane of the cut-circle. Here $\vec{x}_0$ and $\vec{y}_0$ are chosen so that for vanishing pre-rotations the starting conformation is given by $\xi = 0°$:

$$\vec{x}_0 = \frac{(\vec{R}_l + \vec{R}_r) + ((\vec{R}_l + \vec{R}_r)\vec{d})\vec{d}}{\left|(\vec{R}_l + \vec{R}_r) + ((\vec{R}_l + \vec{R}_r)\vec{d})\vec{d}\right|}, \tag{15}$$

$$\vec{y}_0 = \vec{x}_0 \times \vec{d}, \tag{16}$$

where

$$\vec{d} = \vec{D}/\mathrm{D}. \tag{17}$$

$\vec{R}_l$ and $\vec{R}_r$ are here the vectors from the joint to the tip of left and right arm, respectively, immediately after the pre-rotations. The function $f_c(\xi)$ (Eq. (4)) is defined only on those parts of the cut-circle, which are common to both irises. The intervals of ξ where the function $d_c(\xi)$ is defined, are therefore limited by ξ-values which belong to points where iris edges and cut-circle cross. The inner and

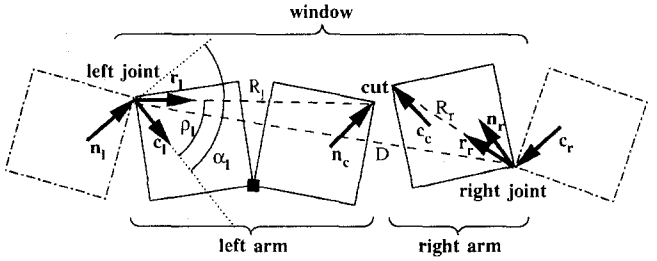

Fig. 12 Schematic view of a window of the protein backbone. The figure shows the situation after the pre-rotations were applied at the C_α-atom (*filled square*). The window contains three amide planes (depicted in *solid lines*). Only these amide planes are affected by the window more. Vectors $\vec{n}$ and $\vec{c}$ are parallel to the N-C_α and C′-C_α bonds, respectively. The indices *l*, *r* and *c* refer to the C_α-atoms at the left joint, cut-position and right-joint, respectively

outer edges of the left iris as well as the cut-circle can be interpreted as edges of cones, with their tips at the left joint. The semiangles of the cones are $(\max(\rho_l,\alpha_l)-\min(\rho_l,\alpha_l))$, $(\rho_l+\alpha_l)$ and $\arccos(D_l/R_l)$, where ρ_l is the angle between $\vec{r}_l$ and $\vec{c}_l$, and α_l the angle between $\vec{n}_l$ and $\vec{c}_l$. The cone axes are $\vec{n}_l$, $\vec{n}_l$ and $\vec{D}$, respectively. An analogous interpretation is possible for the right iris and the cut-circle. Given the general case of two cones with their tips at the same point, unit length cone axes $\vec{a}$, $\vec{b}$ and semiangles α and β, usually one of the following situations is encountered: either the two cones intersect at the tip only, which is the case for $\alpha+\beta<\arccos(\vec{a}\cdot\vec{b})$, or they intersect along the following two vectors $\vec{v}_+$, $\vec{v}_-$

$$\vec{v}_\pm = s_+\frac{\vec{a}+\vec{b}}{|\vec{a}+\vec{b}|}+s_-\frac{\vec{a}+\vec{b}}{|\vec{a}+\vec{b}|}\pm\sqrt{1+s_+^2-s_-^2}\,\frac{\vec{a}\times\vec{b}}{|\vec{a}\times\vec{b}|} \qquad (18)$$

where

$$s_\pm = \frac{\cos\alpha\pm\cos\beta}{|\vec{a}\times\vec{b}|}.$$

With the last relation all vectors $\vec{v}_\pm$ directed to points where iris edges and cut-circle cross can be calculated. Thus two sets of intersection points are obtained, one set for the intersection of the cut-circle with the edge of each of the two irises. The corresponding ξ values are limiting two sets of ξ-intervals. The function $f_c(\xi)$ is defined only on the intervals given by the intersection of both sets of intervals. Incidentally the argumentation in the geometrical terms of the two irises and the cut-circle provides the maximum possible number of definition intervals of $f_c(\xi)$. Since the cut-circle and each of the two irises intersect in a maximum of two separated arcs, the maximum number of arcs common to the cut-circle and both irises is four.

The search for new local conformations is now basically a search for the roots of $f_c(\xi)$ on these intervals. Since $f_c(\xi)$ is defined as $f_c(\xi)=\vec{n}_c(\xi)=\vec{n}_c(\xi)\cdot\vec{c}_c(\xi)-\cos\gamma_{bond}$, where γ_{bond} is the $NC_\alpha C'$-bond angle of the protein backbone model, one has to express the vectors $\vec{n}_c$ and $\vec{c}_c$ in terms of the angle ξ. The ξ-dependence is here derived for the vector $\vec{n}_c(\xi)$, which is associated with the left arm; the evaluation of $\vec{c}_c(\xi)$ at the right arm is analog. After the pre-rotations there are no remaining degrees of freedom within the left arm, so that the left arm can only be moved as a rigid entity by the rotations at the left joint C_α-atom. The following three vectors are fixed with respect to the left arm and in general also linearly independent: 1. the unit-length vector $\vec{r}_l(\xi)$ in direction of $\vec{R}_l(\xi)$ (Eq. (14)), 2. the unit-length bond vector $\vec{c}_l(\xi)$, which can be obtained from Eq. (18) by setting $\vec{a}$, $\vec{b}$, α and β equal to $\vec{n}_l$, $\vec{r}_l$, α_l and ρ_l, respectively, and 3. $(\vec{r}_l(\xi)\times\vec{c}_l(\xi))$. Hence the components γ, δ and ε of $\vec{n}_c(\xi)$ in the directions of $\vec{r}_l(\xi)$, $\vec{c}_l(\xi)$ and $\vec{r}_l(\xi)\times\vec{c}_l(\xi)$ are independent of ξ:

$$\vec{n}_c(\xi)=\gamma\vec{r}_l(\xi)+\delta\vec{c}_l(\xi)+\varepsilon\vec{r}_l(\xi)\times\vec{c}_l(\xi). \qquad (19)$$

Since Eq. (18) yields two vectors $\vec{c}_{l\pm}(\xi)$, according to Eq. (19) there are also two vectors $\vec{n}_{c\pm}(\xi)$ for each ξ. For the same reasons there exist two vectors $\vec{c}_{c\pm}(\xi)$ for each value of ξ on the right side of the cut. Therefore the function $f_c(\xi)$ has four branches, one for each air of $\vec{n}_{c\pm}(\xi)$ and $\vec{c}_{c\pm}(\xi)$:

$$f_{c,nm}(\xi)=\vec{n}_{c,n}(\xi)\cdot\vec{c}_{c,m}(\xi)-\cos\gamma_{bond}, \qquad (20)$$

where

$n,m=+or-$

At interval limits the last term in Eq. (18) vanishes and therefore the branches become pairwise identical, smoothly joining in vertical tangents.

The roots of $f_c(\xi)$ are searched for with the secant algorithm (Press et al. 1992). For each branch and each interval where $f_c(\xi)$ is defined, several starting positions for the secant method are tried. Since $f_c(\xi)$ is smooth and only defined on certain intervals of ξ the iteration scheme behaves in one of the following two ways: either the point where the secant crosses the abscissa leaves the interval after a few iteration steps or it converges quickly to a root $\xi_{i,nm}$, where *nm* indicate the branch of $f_c(\xi)$ (Eq. (20)). All roots ξ_i are collected in an array because they are used in the correction procedure, which is carried out at this stage of the algorithm (appendix B). If solution ξ_i is chosen in the correction procedure, the corresponding new conformation is calculated. This means that one has to calculate dihedral angles and atomic positions (which were not needed in the root search) by resubstitution. First $\vec{r}_l(\xi_{i,nm})=\vec{R}_l(\xi_{i,nm})/|\vec{R}_l(\xi_{i,nm})|$ is obtained from Eq. (14) and then $\vec{c}_l(\xi_{i,nm})$ through Eq. (18) by equating $\vec{v}_n$, $\vec{a}$, $\vec{b}$, α, β with $\vec{c}_l(\xi_{i,nm})$, $\vec{n}_l$, $\vec{r}_l(\xi_{i,nm})$, γ_{bond}, ρ_l, respectively. $\vec{c}_l(\xi_i)$ yields the position of the C′-atom of the first amide plane of the left arm and thus the value of the dihedral angle ϕ_l at the left joint. The calculation of ψ_l requires the position of the N-atom in this amide plane. The position of this atom can be expressed in terms of $\vec{r}_l(\xi_i)$ and $\vec{c}_l(\xi_i)$ in analogy to Eq. (19). Hence ψ_l can be calculated, too. Now all dihedral angles left of the cut are known and therefore all atomic positions on this side can be constructed. The same procedure is also applied to the right arm, yielding dihedral angles and atomic positions of the right side.

The purely geometrical part of the WM, i.e. the generation of a new local conformation of the protein backbone which was demonstrated above, is then followed by the evaluation of the conformational energy (see section on energy function) and the application of the Metropolis criterion.

B Correction Jacobians

The correction requires two sets of solutions produced according to the algorithm described in appendix A: 1. the set of solutions which is obtained for vanishing pe-rotations about the selected pre-rotation axes, and 2. the set of solutions calculated for non-vanishing pre-rotations about the same axes. The reduced set of constraints described in the above section on the correction of the dihedral angle distribution is the same for all solutions within each of the two sets. The two sets are indicated by the index $\alpha = 1, 2$. The changes of dihedral angles at the A C_α-atoms in the window $\phi_i^{(k)}$, $\psi_i^{(k)}$, $i = 1,\ldots,A$; $k = 1,\ldots,s$ must fulfil the four equations

$$D^{(\alpha)} = const \tag{21}$$

$$\beta_l^{(\alpha)} = const \tag{22}$$

$$\beta_r^{(\alpha)} = const \tag{23}$$

$$\alpha^{(\alpha)} = const \tag{24}$$

where (Fig. 12) $D = |\vec{D}|$ is the window diameter, $\beta_l = \angle(\vec{n}_1, \vec{D})$ and $\beta_r = \angle(\vec{c}_r, \vec{D})$ are angles, and $\delta = \angle(\vec{n}_l, \vec{D}, \vec{c}_r)$ is a dihedral angle. Note however that independently of the window size only four of the dihedral angles ϕ_i, ψ_i are subject to Eqs. (21) through (24); the pre-rotations are *per definitionem* unconstrained and the two dihedral angles belonging to torsion axes $\vec{n}_l$, $\vec{c}_r$ can be changed freely without conflicting with Eqs. (21) through (24). The constrained angles are the right dihedral of the left joint, the left dihedral of the right joint and the two dihedrals at the cut, which always belong to axes $\vec{c}_l$, $\vec{n}_c$, $\vec{c}_c$, $\vec{n}_r$, respectively. Conveniently the set of four constrained dihedrals is named ψ_l, ϕ_c, ψ_c, ϕ_r. Infinitesimal perturbations applied to D, β_l, β_r and δ require infinitesimal adjustments $d\psi_l^{(k)}$, $d\phi_c^{(k)}$, $d\psi_c^{(k)}$, $d\phi_r^{(k)}$. In other words the volume elements $dD^{(\alpha)} \cdot d\beta_l^{(\alpha)} \cdot d\beta_r^{(\alpha)} \cdot d\delta^{(\alpha)}$ transforms into s different volume elements $d\psi_l^{(k\alpha)} \cdot d\phi_c^{(k\alpha)} \cdot d\psi_c^{(k\alpha)} \cdot d\phi_r^{(k\alpha)}$. Since the aim is a uniform distribution of conformations in the space of the constrained dihedrals, the solutions have to be weighted according to these volume elements. In mathematical terms, the infinitesimal volume elements are proportional to the Jacobian determinants

$$J^{(k\alpha)} = \left| \frac{\partial\left(\psi_l^{(k\alpha)}, \phi_c^{(k\alpha)}, \psi_c^{(k\alpha)}, \phi_r^{(k\alpha)}\right)}{\partial\left(D^{(\alpha)}, \beta_l^{(\alpha)}, \beta_r^{(\alpha)}, \delta^{(\alpha)}\right)} \right|. \tag{25}$$

The use of these Jacobians as weights for the solutions allows an exact correction of the biased distribution caused by the constraints. In the remainder of this section, analytical expressions for the Jacobians are derived.

The rotation of a vector $\vec{v}$ about an axis $\vec{a}$ by an infinitesimal angle $d\varepsilon$ yields a rotated vector $\vec{v}'$,

$$\vec{v}' = \vec{v} + d\varepsilon\, \vec{a} \times \vec{v}. \tag{26}$$

Equation (26) is implicitely used in the following equations. By appling dihedral angle changes $d\psi_l^{(k\alpha)}$, $d\phi_c^{(k\alpha)}$, $d\psi_c^{(k\alpha)}$, $d\phi_r^{(k\alpha)}$, $k = 1,\ldots,s$ towards the right side of the backbone and observing which constraints (Eqs. (21) to (24)) are perturbed by which of the changes one arrives at the equation

$$\begin{pmatrix} 0 & T_{12}^{(k\alpha)} & T_{13}^{(k\alpha)} & 0 \\ T_{21}^{(k\alpha)} & T_{22}^{(k\alpha)} & T_{23}^{(k\alpha)} & 0 \\ 0 & T_{32}^{(k\alpha)} & T_{33}^{(k\alpha)} & T_{34}^{(k\alpha)} \\ T_{41}^{(k\alpha)} & T_{42}^{(k\alpha)} & T_{43}^{(k\alpha)} & T_{44}^{(k\alpha)} \end{pmatrix} \begin{pmatrix} d\psi_l^{(k\alpha)} \\ d\phi_c^{(k\alpha)} \\ d\psi_c^{(k\alpha)} \\ d\phi_r^{(k\alpha)} \end{pmatrix} = D^{(\alpha)} \times \tag{27}$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ S_{21}^{(\alpha)} & S_{22}^{(\alpha)} & 0 & 0 \\ S_{31}^{(\alpha)} & 0 & S_{33}^{(\alpha)} & 0 \\ S_{41}^{(\alpha)} & 0 & 0 & S_{44}^{(\alpha)} \end{pmatrix} \begin{pmatrix} dD^{(\alpha)} \\ d\beta_l^{(\alpha)} \\ d\beta_r^{(\alpha)} \\ d\delta^{(\alpha)} \end{pmatrix},$$

where the matrix elements are

$$T_{12}^{(k)} = -\vec{R}_l\,(\vec{n}_c \times \vec{R}_r);\; T_{13}^{(k)} = \vec{R}_l\,(\vec{c}_c \times \vec{R}_r);$$

$$T_{21}^{(k)} = (\vec{n}_l \times \vec{D})(\vec{n}_l \times (\vec{c}_l \times \vec{D}));\; T_{22}^{(k)} = -(\vec{n}_l \times \vec{D})(\vec{n}_c \times \vec{R}_r));$$

$$T_{23}^{(k)} = (\vec{n}_l \times \vec{D})(\vec{n}_l \times (\vec{c}_c \times \vec{R}_r));$$

$$T_{32}^{(k)} = -(\vec{c}_r \times \vec{D})\,(\vec{c}_r \times (\vec{n}_c \times \vec{R}_l));$$

$$T_{33}^{(k)} = (\vec{c}_r \times \vec{D})\,(\vec{c}_r \times (\vec{c}_c \times \vec{R}_l));$$

$$T_{34}^{(k)} = (\vec{c}_r \times \vec{D})\,(\vec{c}_r \times (\vec{n}_r \times \vec{D}));$$

$$T_{41}^{(k)} = \Big[(\vec{n}_l \times (\vec{c}_l \times \vec{D})) \times \vec{p} - \vec{q} \times ((\vec{c}_l \times \vec{D}) \times \vec{c}_r) - \vec{q} \times (\vec{D} \times (\vec{c}_l \times \vec{c}_r))\Big] \cdot \vec{D};$$

$$T_{42}^{(k)} = \Big[-(\vec{n}_l \times (\vec{n}_c \times \vec{R}_r)) \times \vec{p} + \vec{q} \times ((\vec{n}_l \times \vec{R}_r) \times \vec{c}_r) - \vec{q} \times (\vec{D} \times (\vec{n}_c \times \vec{c}_r))\Big] \cdot \vec{D} - (\vec{q} \times \vec{p})(\vec{n}_c \times \vec{R}_r));$$

$$T_{43}^{(k)} = \Big[(\vec{n}_l \times (\vec{c}_c \times \vec{R}_r)) \times \vec{p} - \vec{q} \times ((\vec{c}_c \times \vec{R}_r) \times \vec{c}_r) + \vec{q} \times (\vec{D} \times (\vec{c}_c \times \vec{c}_r))\Big] \cdot \vec{D} + (\vec{q} \times \vec{p})(\vec{c}_c \times \vec{R}_r));$$

$$T_{44}^{(k)} = \Big[\vec{q} \times (\vec{D} \times (\vec{n}_r \times \vec{c}_r))\Big] \cdot \vec{D};$$

$$(\vec{p} \equiv \vec{c}_r \times \vec{D};\; \vec{q} \equiv \vec{n}_l \times \vec{D})$$

and

$$S_{21} = \sin^2 \beta_l;\; S_{31} = \sin^2 \beta_r;\; S_{41} = 3D \sin\delta;$$

$$S_{22} = \frac{D}{2} \sin(2\beta_l);\; S_{33} = \frac{D}{2} \sin(2\beta_r);\; S_{44} = D^2 \cos$$

For the sake of clarity on the right hand sides indices $k = 1,\ldots,s$ and $\alpha = 1,2$ have been omitted. If Eq. (27) is written in shorthand form as:

$$\mathbf{T}^{(k\alpha)} \left(d\psi_l^{(k\alpha)}\, d\phi_c^{(k\alpha)}\, d\psi_c^{(k\alpha)}\, d\phi_r^{(k\alpha)}\right)^t = D^{(\alpha)}\, \mathbf{S}^{(\alpha)} \left(dD^{(\alpha)}\, d\beta_l^{(\alpha)}\, d\beta_r^{(\alpha)}\, d\delta^{(\alpha)}\right)^t;$$

then the required Jacobians $J^{(k\alpha)}$ for solutions $k = 1,\ldots,s$ are

$$J^{(k\alpha)} = D^{(\alpha)} \det\left((\mathbf{T}^{(k\alpha)})^{-1}\mathbf{S}^{(\alpha)}\right) \tag{28}$$

$$= \frac{D^{(\alpha)^5} \sin(2\beta_l^{(\alpha)} \sin(2\beta_r^{(\alpha)}) \cos\delta^{(\alpha)}}{4 \det(\mathbf{T}^{(k\alpha)})}.$$

The $J^{(k\alpha))}$ are then used in the weighting scheme described in section on the correction of the distribution of dihedrals. Note that all quantities required to calculate the element of Eq. (27) are by products of the root finding procedure. The actual dihedral angles or atomic positions are not needed.

References

Abagyan R, Totrov M (1994) Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. J Mol Biol 235:983–1002

Baldwin RL (1990) Pieces of the folding puzzle. Nature 346:409–410

Baumgärtner A (1984) Simulation of polymer motion. Ann Rev Phys Chem 35:419–435

Bouzida D, Kumar S, Swendsen RH (1992) Efficient Monte Carlo methods for the computer simulation of biological molecules. Phys Rev A 45:8894–8901

Braun W (1987) Local deformation studies on chain molecules: differential conditions for changes of dihedral angles. Biopolymers 26:1691–1704

Brooks CL, Nilsson L (1993) Promotion of helix formation in peptides dissolved in alcohol and water-alcohol mixtures. J Am Chem Soc 115:11034–11035

Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculation. J Comp Chem 4:187–217

Brooks BR (1989) Molecular dynamcis for problems in structural biology. Chemica Scripta 29A:165–169

Bruccoleri RE, Karplus M (1985) Chain closure with bond angle variations. Macromolecules 18:2767–2773

Creighton TE, Kim PS (1991) Folding and binding. Curr Op Struct Biol 1:3–4

Creighton TE (1992) Folding and binding. Curr Op Struct Biol 2:1–5

Daggett V, Levitt M (1992) Molecular dynamics stimulation of helix denaturation. J Mol Biol 223:1121–1138

Daggett V, Kollman PA, Kuntz ID (1991) A molecular dynamics simulation of polyalanine: an analysis of equilibrium motions and helix-coil transitions. Biopolymers 31:1115–1132

de Gennes PG (1971) Reptation of a polymer chain in the presence of fixed obstacles. J Chem Phys 55:572–579

Deloof H, Nilsson L, Rigler R (1992) Molecular dynamics simulation of galanin in aqueous and nonaqueous solution. J Am Chem Soc 114:4028–4035

Derreumaux P, Schlick T (1995) Long timestep dynamics of peptides by the dynamics driver aproach. Proteins Struct Func Genet 21:283–302

DiCipua FM, Swaminathan S, Beveridge D (1990) Theoretical evidence for destabilization of an α-helix by water insertion: molecular dynamics of hydrated decaalanine. J Am Chem Soc 112:6768–6771

Dickerson RE, Geis J (1969) The structure and action of proteins. Harper, New York

Dodd LR, Boone TD, Theodorou DN (1993) A concerted rotation algorithm for atomistic Monte Carlo simulation of polymer melts and glasses. Molec Phys 78:961–996

Doi M, Edwards SF (1986) The theory of polymer dynamics. Clarendon Press, Oxford

Eaton WA, Szabo A (1991) Protein dynamics. Chem Phys 158:191–531

Englander SW (1993) In pursuit of protein folding. Science 262:848–849

Fersht A (1984) Enzyme structure and mechanism. Freeman, New York

Fixman M (1974) Classical statistical mechanics of constraints: a theorem and application to polymers. Proc Natl Acad Sci USA 71:3050–3053

Fixman M (1978) Simulation of polymer dynamics. I. General theory. J Chem Phys 69:1527–1545

Fukugita M, Lancaster D, Mitchard MG (1993) Kinematics and thermodynamics of a folding heteropolymer. Proc Natl Acad Sci USA 90:6365–6368

Gerber PR (1992) Peptide mechanics: a force field for peptides working with entire residues as smallest units. Biopolymers 32:1003–1017

Gō N, Scheraga HA (1970) Ring closure and local conformational deformations of chain molecules. Macromolecules 3:178–187

Gō N, (1983) Theoretical studies of protein folding. Ann Rev Biophys Bioeng 12:183–210

Grønbech-Jensen N, Doniach S (1994) Long-time overdamped langevin dynamics of molecular chains. J Comp Chem 15:997–1012

Guo Z, Thirumalai D, Honeycutt JD (1992) Folding kinetics of proteins: A model study. J Chem Phys 97:525–535

Head-Gordon T, Brooks CL (1991) Virtual rigid body dynamics. Biopolymers 31:77–100

Helfand E (1979) Flexible vs rigid constraints in statistical mechanics. J Chem Phys 71:5000–5007

Hinds D, Levitt M (1992) A lattice model for protein structure prediction at low resolution. Proc Natl Acad Sci USA 89:2536–2540

Hoffmann D, Knapp EW (1996) Protein dynamics with off-lattice Monte Carlo moves. Phys Rev E 53:4221–4224

Hol WGJ (1991) Proteins. Curr Op Struct Biol 1:875–876

Honeycutt JD, Thirumalai D (1990) Metastability of the folded states of globular proteins. Proc Natl Acad Sci US 87:3526–3529

Honig B (1993) Theory and simulation. Curr Op Struct Biol 3:223–224

Hubbard SJ, Eisenmenger F, Thornton JM (1994) Modeling studies of the change in conformation required for cleavage of limited proteolytic sites. Protein Science 3:757–768

Kabsch W (1976) A solution for the best rotation to relate two sets of vectors. Acta Cryst A32:922–923

Kang HS, Kurochkina NA, Lee B (1993) Estimation and use of protein backbone angle probabilities. J Mol Biol 229:448–460

Karplus M (1986) Internal dynamics of proteins. Methods Enzymol 131:283–307

Knapp EW, Sklenar H (1993) Protein folding – solution in sight? J Comp Aid Molec Design 7:365

Knapp EW, Irgens-Defregger A (1991) Long time dynamcis of proteins: an off-lattice Monte Carlo method. In: Harms U (ed) Supercomputer and chemistry vol 2. Springer, Berlin Heidelberg New York, pp 83–106

Knapp EW, Irgens-Defregger A (1992) Off-lattice Monte Carlo method with constraints: long-time dynamics of a protein model without non-bonded interactions. J Comp Chem 13:1–11

Knapp EW, Irgens-Defregger A (1993) Off-lattice Monte Carlo method with constraints: long-time dynamics of a protein model without nonbonded interactions. J Comp Chem 14:19–29

Knapp EW (1992) Long time dynamics of a polymer with rigid monomer units relating to a protein model: comparison with the rouse model. J Comp Chem 13:793–798

Kolata G (1986) Trying to crack the second half of the genetic code. Science 233:1037–1039

Kolinski A, Milik M, Skolnick J (1991) Static and dynamic properties of a new lattice model of polypeptide chains. J Chem Phys 94:3978–3985

Kraulis P (1991) Molscript: a program to produce both detailed and schematic plots of protein structures. J Appl Cryst 24:946–950

Kremer K, Binder K (1988) Monte Carlo simulations of lattice models for macromolecules. Comp Phys Rep 7:259–312

Kuntz ID, Crippen GM, Kollman PA, Kimelman D (1976) Calculation of protein tertiary structure. J Mol Biol 106:983–994

Kuntz ID (1994) Theory and simulation. Curr Op Struct Biol 4:231–233

Kurochkina NA, Kang HS, Lee B (1994) Experiences with dihedral angle space Monte Carlo search for small protein structures. In: Doniach S (ed) Statistical Mechanics, protein structure and protein substrate interaction. Plenum, New York

Levitt M, Warshel A (1975) Computer simulation of protein folding. Nature 253:694–698

Levitt M (1992) Theory and simulation. Curr Op Struct Biol 2:191–192

Lowry GG (1970) Markow chains and Monte Carlo calculations in polymer science. Dekker, New York

Manocha D, Canny JF (1994) Efficient inverse kinematics for general 6r manipulators. IEEE transactions on robotics and automation, 10(5):648

Manseur R, Doty KL (1989) A robot manipulator with 16 real inverse kinematic solution sets. Int J Rob Res 8:75

McCammon JA, Harvey SC (1987) Dynamics of proteins and nucleic acids. Cambridge University Press, Cambridge

McCammon JA, Northrup SH, Karplus M, Levy R (1980) Helix-coil transitions in a simple polypeptide model. Biopolymers 19:2033

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. J Chem Phys 21:1087–1092

Miranker A, Robinson CV, Radford SE, Aplin RT, Dobson CM (1993) Detection of transient protein folding populations by mass spectrometry. Science 262:896ff

Nadler W, Brünger AT, Schulten K, Karplus M (1987) Molecular and stochastic dynamics of proteins. Proc Natl Acad Sci USA 84:7933–7937

Noguti T, Gō N, (1985) Efficient Monte Carlo for simulation of fluctuating conformations of native proteins. Biopolymers 24:527–546

Orwoll RA, Stockmayer WH (1969) Stochastic models for chain dynamics. Adv Chem Phys 15:305–324

Palmer KA, Scheraga HA (1992) Standard-geometry chains fitted to x-ray derived structures: validation of the rigid-geometry approximation. II. Systematic searches for short loops in proteins: applications to bovine pancreatic ribonuclease A and human lysozyme. J Comp Chem 13:329–350

Pleiss J, Jähnig F (1992) Conformational transition of an α-helix studied by molecular dynamics. Eur Biophys J 21:63–70

Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1992) Numerical recipes in C (2nd edn). Cambridge University Press, Cambridge

Primrose EJF (1986) On the input-output equation of the general 7r-mechanism. Mech Mach Theory 21:509–510

Radford SE, Dobson CM, Evans PA (1992) The folding of hen lysozyme involves partially structured intermediates and multiple pathways. Nature 358:302–307

Rey A, Skolnick J (1991) Comparison of lattice Monte Carlo dynamics and Brownian dynamics folding pathways of α-helical hairpins. Chem Phys 158:199–219

Rey A, Skolnick J (1993) Computer modeling and folding of four-helix bundles. Proteins Struct Function Gen 16:8–28

Roth B, Rastegar J, Scheinman V (1973) On the design of computer controlled manipulations. On the theory and practice of robots and manipulators. CISM IFToMM Symposium 1:93–113

Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23:327–341

Schulz GE, Schirmer RH (1978) Principles of protein structure. Springer, Berlin Heidelberg New York

Sklenar H, Lavery R, Pullman B (1986) The flexibility of the nucleic acids: (I) "SIR", a novel approach to the variation of polymer geometry in constrained systems. J Biomol Struct Dyn 3:967–987

Skolnick J, Kolinski A (1990) Simulations of the folding of a globular protein. Science 250:1121–1125

Skolnick J, Kolinski A (1991) Dynamic Monte Carlo simulations of a new lattice model of globular protein folding, structure and dynamics. J Mol Biol 221:499–531

Skolnick J, Kolinski A, Brooks CL, Godzik A, Rey A (1993) A method for predicting protein structure from sequence. Curr Biol 3:414–423

Soman KV, Karimi A, Chase DA (1991) Unfolding of an α-helix in water. Biopolymers 31:1361–1361

Sung S-S (1994) Helix folding simulations with various initial conformations. Biophys J 66:1796–1803

Tanaka S, Scheraga HA (1976) Medium- and long-range interaction parameters between amino acids for predicting three-dimensional structures of proteins. Macromolecules 9:945–950

Tirado-Rives J, Jorgensen WL (1991) Molecular dynamics simulations of the unfolding of an α-helical analogue of ribonuclease A S-peptide in water. Biochemistry 30:3864–3871

Tobias TJ, Sneddon SF, Brooks CL (1992) Stability of a model β-sheet in water. J Mol Biol 227:1244–1252

Šali A, Shakhnovich E, Karplus M (1994) How does a protein fold? Nature 369:248–251

van Gunsteren WF, Berendsen HJC (1977) Algorithms for macromolecular dynamics and constraint dynamics. Mol Phys 34:1311–1327

van Gunsteren WF, Karplus M (1981) Effects of constraints, solvent and crystal environment on protein dynamics. Nature 293:677–678

van Gunsteren WF (1988) The role of computer simulation techniques in protein engienering. Protein Eng 2:5–13

van Gunsteren WF (1991) Computer simulation of biomolecular systems: Overview over time-saving techniques. In: Lavery R, Rivail JL, Smith J (eds) Advances in biomolecular simulations, vol 239 of American Institute of Physics Conference Proceedings. American Institute of Physics. New York, pp 131–146

Verdier PH (1966) Monte Carlo studies of lattice-model polymer chains. I. Correlation functions in the statistical-bead model. J Chem Phys 45:2118–2126

Verdier PH (1973) Monte Carlo studies of lattice-model polymer chains. III. Relaxation of Rouse coordinates. J Chem Phys 59:6119–6127

Wall FT, Mandel F (1975) Macromolecular dimensions obtained by an efficient Monte Carlo method without sample attrition. J Chem Phys 63:4592–4595

Wilson C, Doniach S (1989) A computer model to dynamically simulate protein folding: Studies with crambin. Proteins Struct Function, Gen 6:193–209