ELSEVIER

Chemical Physics 261 (2000) 25–37

Chemical Physics

www.elsevier.nl/locate/chemphys

Harmonicity in slow protein dynamics

Konrad Hinsen [a], Andrei-Jose Petrescu [b,c], Serge Dellerue [c], Marie-Claire Bellissent-Funel [c], Gerald R. Kneller [a,*]

[a] *Centre de Biophysique Moléculaire (CNRS UPR 4301), Rue Charles Sadron, F-45071 Orléans Cedex 2, France*
[b] *Institute of Biochemistry, Splaiul Independentei 296, 77700 Bucharest 17, Romania*
[c] *Laboratoire Léon Brillouin (CEA-CNRS), CEA-Saclay, F-91191 Gif-sur-Yvette Cedex, France*

Received 11 January 2000

Abstract

The slow dynamics of proteins around its native folded state is usually described by diffusion in a strongly anharmonic potential. In this paper, we try to understand the form and origin of the anharmonicities, with the principal aim of gaining a better understanding of the principal motion types, but also in order to develop more efficient numerical methods for simulating neutron scattering spectra of large proteins. First, we decompose a molecular dynamics (MD) trajectory of 1.5 ns for a C-phycocyanin dimer surrounded by a layer of water into three contributions that we expect to be independent: the global motion of the residues, the rigid-body motion of the sidechains relative to the backbone, and the internal deformations of the sidechains. We show that they are indeed almost independent by verifying the factorization of the incoherent intermediate scattering function. Then, we show that the global residue motions, which include all large-scale backbone motions, can be reproduced by a simple harmonic model which contains two contributions: a short-time vibrational term, described by a standard normal mode calculation in a local minimum, and a long-time diffusive term, described by Brownian motion in an effective harmonic potential. The potential and the friction constants were fitted to the MD data. The major anharmonic contribution to the incoherent intermediate scattering function comes from the rigid-body diffusion of the sidechains. This model can be used to calculate scattering functions for large proteins and for long-time scales very efficiently, and thus provides a useful complement to MD simulations, which are best suited for detailed studies on smaller systems or for shorter time scales.

Keywords: Inelastic neutron scattering; Protein dynamics; Molecular dynamics; Brownian dynamics; Normal modes

1. Introduction

One of the most important applications of neutron scattering in the biological sciences has been the study of protein dynamics. In this field, computer simulations have always been essential for interpreting the experimental spectra due to the complexity of the systems [1,2]. The prevalent simulation technique has been molecular dynamics (MD) simulation, which calculates the detailed motion of all atoms. From this information, all observable quantities can be calculated in principle, although the length of simulation runs that can be performed in practice is often a limiting factor. Few, if any, MD trajectories ever produced for proteins are long enough to permit a

* Corresponding author. Fax: +2-38-63-1517.
E-mail address: kneller@cnrs-orleans.fr (G.R. Kneller).

0301-0104/00/$ - see front matter
PII: S0301-0104(00)00222-6

conformational sampling that includes the low-frequency large-amplitude motions. As an alternative technique, normal mode analysis has been used to reproduce vibrational spectra. This technique has the advantage of permitting the analytical calculation of time-dependent quantities on all time scales, without the sampling problems of MD, but it is inherently unable to describe the diffusive dynamics that dominates protein motion on long time scales and at physiological temperatures. Another problem is the limitation to a single local minimum of the potential energy surface, which cannot describe the amplitudes of large-scale motions correctly.

Several attempts have been made to introduce friction and thereby diffusive motions into normal mode models. Smith et al. [3] used a model in which they assigned a friction term to each normal mode independently, i.e. they neglected the coupling between modes that results from the random forces. They used an empirical model for the frequency-dependent friction constant. This approach has also been used by Hayward et al. [4]. Kottalam and Case [5] and later Ansari [6] have applied the Langevin mode formalism developed by Lamm and Szabo [7] on the basis of a theory by Wang and Uhlenbeck [8] to proteins, using an atomic model and force field with friction constants calculated according to an empirical formula that was developed by Pastor and Karplus [9] for liquid alkanes. The validity of this model for proteins is questionable, however, and as in the work of Smith et al., the limitation to a local minimum of an atomic force field remains. Friction constants have also been obtained from MD trajectories (e.g. Ref. [10]), but this approach has not become widely used.

The aim of this article is to gain a better understanding of the long-time diffusive dynamics of proteins, as measured by neutron scattering, and to examine whether it is possible to describe it by a much simpler model than explicit Newtonian dynamics on the atomic scale. Our starting point (Section 2) is a long MD trajectory for a mid-sized protein, a C-phycocyanin dimer. In Section 3, we decompose this trajectory into three terms which we expect to be approximatively independent: one term describes the large-scale motions of the backbone, the second one, the rotational diffusion of the sidechains relative to the backbone, and the third term describes the internal deformations of the sidechains.

In a compactly folded protein with a single well-defined equilibrium structure, the overall shape of the potential energy surface must be that of a single well. It is well known that the stable equilibrium configurations of proteins actually consists of multiple conformational substates [11–13], but this "fine structure" of the potential energy surface can be regarded as a modulation of a smooth single-well potential that defines the protein's general folding topology. Motion in such a system can be described qualitatively as vibrational motion within a local potential energy minimum on a short-time scale, and as diffusion on a smooth effective potential energy surface on a long-time scale. In Section 4, we will attempt to describe the large-scale dynamics of a protein by precisely such a model, assuming both the local minima and the overall potential well to be harmonic.

2. Molecular dynamics simulation

The MD trajectory was prepared and used to compare the simulation data with neutron scattering experiments performed on a hydrated powder of C-phycocyanin [14–16]. Protein powders obtained by lyophylisation and re-hydration are disordered systems that differ structurally both from solutions and from crystalline powders. A realistic description of such a system would have been an arrangement of four to eight molecules in the simulation box, with the corresponding amount of hydration water. However, this would have increased the number of atoms beyond what can be reasonably treated by MD simulations. It was therefore decided to use the simplest model that still retains the essential properties of the system: a protein dimer surrounded by a small amount of water such that the water/protein weight ratio is the same as in the powders used in experiments [15,16].

The preparation of the hydrated C-phycocyanin model and the generation of the MD trajectory were performed using the CHARMM program [17]

with version 22 of the all-atom potential function [18]. The parametrization for the chromophores, for which there are no parameters in this potential function, will be described elsewhere.

The starting point of the simulation was a crystallographic structure by Duerring et al. [19], PDB code 1CPC, from which an all-atom representation of the C-phycocyanin dimer was constructed. This model consists of the two peptide chains, three phycocyanobilins, and 243 water molecules. A box with water molecules at normal density of dimensions $9 \times 6 \times 4.5$ nm^3 was equilibrated using CHARMM. Water molecules with the oxygen within 0.26 nm from any heavy atom of the protein and at a distance larger than 0.47 nm were eliminated, as were five water molecules that would have been placed in pockets inside the protein where no water is found in the X-ray structure. In this way, a shell corresponding to a hydration level of $\approx$0.6 g g^{-1} and corresponding approximately to two hydration shells was generated around the protein. In total, there are 1086 water molecules in our system, which together with the protein make up for 8417 atoms.

The simulation was performed in the microcanonical (NVE) ensemble. The cutoff distance for the nonbonded list was set to 1.4 nm instead of the more usual value of 1.2 nm. The cutoff distances for the smoothing function were 0.9 m, for the start and 1.3 nm for the end. These values were found to increase the accuracy of the simulation without greatly affecting the computational effort per simulation step. To allow a simulation time step of 2 fs, bonds between hydrogens and heavy atoms were kept rigid.

After an initial energy minimization (500 steps of steepest descent, followed by 5000 steps with the second-order adopted basis Newton–Raphson (ABNR) algorithm), the system was brought to the simulation temperature of 300 K in steps of 5 K over 3000 steps. In the subsequent equilibration phase, the particle velocities were rescaled to bring the temperature back to 300 K whenever it deviated by more than 10 K from this value. This was continued until such rescaling steps were no longer necessary, which was after 5000 steps. Finally, a production run was started in which the system configuration was stored in a trajectory files every 50 steps over a total simulated time of 1.6 ns. The root-mean-square (RMS) deviation from the crystal structure during the dynamics was less than 0.15 nm.

3. Trajectory analysis

The first step in the analysis was the elimination of global translation and rotation from the trajectory, followed by the construction of three new trajectories of which each one contains only one specific part of the total motion. The Molecular Modeling Toolkit [20] was used for these operations.

To represent the collective motions of the backbone, we chose the C_α atoms; other choices (e.g., the center of mass of each residue) are possible, but make little difference. For the calculation of the scattering functions, each C_α atom was assigned the scattering length of the whole residue that it represents.

The second trajectory contains the rigid-body motion of the sidechains. However, contrary to an earlier study on rigid sidechain motions [21,22], we include only the motion relative to the backbone because the global motion of the whole residue is already taken into account in the first trajectory. Since the sidechains are connected to the backbone by a chemical bond, relative translation is negligible, such that only rotational motion must be kept. In the sidechain rotation trajectory, the backbone atoms keep their initial positions at all times. For the sidechains, a rigid-body superposition fit [23] onto the initial conformation is performed at each time step. The result of this fit is the linear transformation that transforms initial conformation of the sidechain to the conformation at time t with the smallest possible error; this transformation is described by a displacement vector $\mathbf{d}_{sc}(t)$ and a rotation matrix $\mathbf{R}_{sc}(t)$. The same superposition fit is performed for a five-residue segment of the backbone whose center is the residue under consideration; the result of this fit is described by $\mathbf{d}_{bb}(t)$ and $\mathbf{R}_{bb}(t)$. The displacement vectors are not used because only the rotational motion is of interest. The rotation of the sidechain relative to the backbone is described by the

rotation matrix $\mathbf{R}_{\mathrm{rel}}(t) = \mathbf{R}_{\mathrm{bb}}^{-1}(t) \cdot \mathbf{R}_{\mathrm{sc}}(t)$. This rotation is applied to all atoms of the sidechain in the initial configuration, with the position of the C_α atom as reference point. The resulting trajectory can contain unphysical configurations in which some interatomic distances are very small. However, this is not a problem as long as only single-atom quantities, such as incoherent scattering functions, are calculated.

The third trajectory contains the internal deformation of the sidechains. As for the sidechain rotation trajectory, the backbone atoms keep their initial positions at all times. The sidechain atom positions are determined by rotating the sidechains around the C_α positions with the rotation matrix $\mathbf{R}_{\mathrm{sc}}^{-1}(t)$. Again there is a possibility of generating unphysical configurations, limiting the use of these trajectories to the calculation of single-atom quantities.

The second step of the analysis is the calculation of the incoherent intermediate scattering function $F_{\mathrm{inc}}(q,t)$ for the full trajectory and for the three extracted trajectories, using a recent reimplementation [24] of the NMOLDYN package [25]. For each value of q, the scattering functions were calculated as an average over 30 q-vectors chosen randomly from a uniform direction distribution. We call the four results $F_{\mathrm{inc}}^{\mathrm{all}}$, $F_{\mathrm{inc}}^{\mathrm{C}_\alpha}$, $F_{\mathrm{inc}}^{\mathrm{rot}}$, and $F_{\mathrm{inc}}^{\mathrm{def}}$. If we assume that the three extracted trajectories to be approximately uncorrelated and their combination to represent all possible motions, then the product

$$F_{\mathrm{inc}}^{\mathrm{prod}}(q,t) = F_{\mathrm{inc}}^{\mathrm{C}_\alpha}(q,t) F_{\mathrm{inc}}^{\mathrm{rot}}(q,t) F_{\mathrm{inc}}^{\mathrm{def}}(q,t) \qquad (1)$$

should be a good approximation for $F_{\mathrm{inc}}^{\mathrm{all}}(q,t)$.

This approximation is verified in Fig. 1. The most important conclusion from this comparison is that the C_α and sidechain rotation contributions together provide an excellent approximation at low q, i.e. for large-scale motions, as well as for short times at all q. The C_α and the sidechain rotation contributions are of similar magnitude. The sidechain deformation contribution, however, is much smaller. At short times, it reduces the agreement with the full trajectory result, which means that the assumption of negligible correlation is not valid. For long times, when all correlations become small and F_{inc} measures essentially the position fluctuations of the atoms, the sidechain deformation term improves the agreement by providing the missing fluctuation contribution.

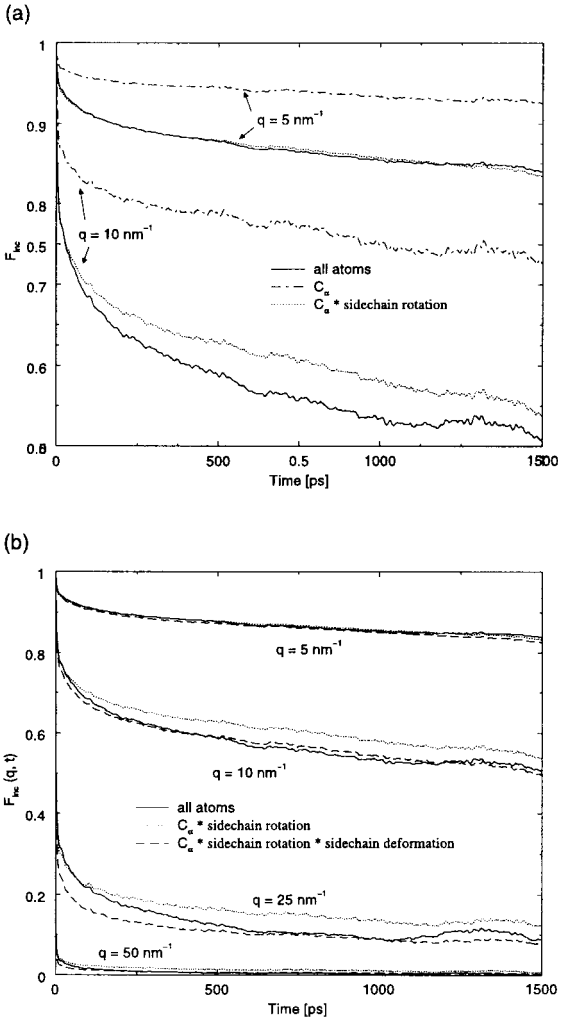

Fig. 1. The incoherent intermediate scattering function $F_{\mathrm{inc}}^{\mathrm{all}}(q,t)$ compared to the factorization approximation from Eq. (1). (a) $F_{\mathrm{inc}}^{\mathrm{C}_\alpha}$ and $F_{\mathrm{inc}}^{\mathrm{C}_\alpha} F_{\mathrm{inc}}^{\mathrm{rot}}$ compared to $F_{\mathrm{inc}}^{\mathrm{all}}$ for two values of q. The C_α contribution alone is clearly insufficient, but the product with the sidechain rotation contribution is an excellent approximation for $q = 0.5$ Å^{-1} and also for $q = 0.5$ Å^{-1} and short times. (b) $F_{\mathrm{inc}}^{\mathrm{C}_\alpha} F_{\mathrm{inc}}^{\mathrm{rot}}$ and $F_{\mathrm{inc}}^{\mathrm{C}_\alpha} F_{\mathrm{inc}}^{\mathrm{rot}} F_{\mathrm{inc}}^{\mathrm{def}}$ compared to $F_{\mathrm{inc}}^{\mathrm{all}}$ for four values of q. The addition of the sidechain deformation term provides the correct long-time values, but leads to a wrong behavior at short times. This shows that the assumption of negligible correlation between the three types of motion is not entirely justified.

We conclude from this analysis that the motions which are measured by $F_{\mathrm{inc}}(q,t)$ can to a good approximation be described by an uncorrelated combination of whole-residue motion (represented by the motion of the C_α atoms) and rigid-body rotation of the sidechains relative to the backbone.

4. Brownian modes

Our next aim is the reproduction of the long-time large-scale motions, measured by $F_{\mathrm{inc}}^{C_\alpha}(q,t)$, with a much simplified physical model, in order to verify if the assumption of harmonic motion is justified. Our protein model consists only of the C_α atoms, which however are assigned the masses of the whole residues they represent. Such a model has been used successfully for studies of thermal fluctuations [26] and domain motions [27,28]. The C_α atoms interact via an harmonic pair potential, which is derived from the MD trajectory, and are subject to friction originating from the solvent and from the other protein atoms that are not included explicitly.

We assume to be in the Brownian motion regime in which the only dynamical variables are the positions of the atoms. The usual justification for this model is that the velocities are supposed to be uncorrelated on the time scale under consideration and their distribution thus is the equilibrium distribution. In a damped harmonic system, the relaxation times of the position and velocity autocorrelation functions are the same. However, as shown in Ref. [29], on a coarse-grained time scale, the neutron scattering law for the Langevin and Brownian motion regime are the same if friction dominates. This is certainly the case for the large-scale motions that we wish to describe. Here, the local dynamics is governed by many jumps between local minima which create an effective friction on a coarse-grained time scale. The net effect is diffusion in a smooth effective potential. We will deal with this diffusion first and come back to the short-time vibrations in a local minimum in Section 4.4. All the techniques described in this section were implemented using the Molecular Modeling Toolkit [20].

4.1. Brownian motion in an harmonic potential

The equation of motion for Brownian dynamics is the Smoluchowski equation,

$$\frac{\partial}{\partial t}P = \nabla^{\mathrm{T}} \cdot \mathbf{D} \cdot \nabla P - \frac{1}{k_{\mathrm{B}}T}\nabla^{\mathrm{T}} \cdot \mathbf{D} \cdot \mathbf{f}P, \tag{2}$$

where $P(\mathbf{R},t|\mathbf{R}',t')$ is the probability for a move $\mathbf{R}' \to \mathbf{R}$ within a time interval $t' \to t$, and $\mathbf{R}$ and $\mathbf{R}'$ are the $3N$-dimensional position vectors (N being the number of particles) at time t and t', respectively. The vector $\mathbf{f}$ comprises the forces, and the diffusion matrix $\mathbf{D}$ is related to the friction matrix $\boldsymbol{\gamma}$ by $\mathbf{D} = k_{\mathrm{B}}T\boldsymbol{\gamma}^{-1}$. In the case of an harmonic system, we have $\mathbf{f} = -\mathbf{K} \cdot (\mathbf{R} - \mathbf{R}^{\mathrm{eq}})$, where $\mathbf{R}^{\mathrm{eq}}$ is a (stable) minimum of the potential energy. The force constant matrix $\mathbf{K}$ is positive (semi-)definite, and the solution of the Smoluchowski equation is a Gaussian distribution. The corresponding stochastic process described by it is known as the Ornstein–Uhlenbeck process [8]. This solution can be expressed in a form that is similar to standard normal mode analysis, and analytic expressions for the scattering functions can be derived [29]. The expressions below are taken from Ref. [29] but are rewritten in non-mass-weighted Cartesian coordinates to be suitable for direct numerical application. Here, we cannot make the assumption that the force constant matrix is positive definite, because for a system without external forces, it always has six zero-frequency eigenvalues. This requires some minor modifications to the expressions found in Ref. [29]. Essentially, the inverse of $\mathbf{K}$ has to be replaced by the pseudoinverse, $\mathbf{K}^+$, which has the properties that $\mathbf{K}^+ \cdot \mathbf{K} = \mathbf{K} \cdot \mathbf{K}^+$ is the projector on the subspace describing infinitesimal global rotations and translations.

The incoherent intermediate scattering function for a harmonic Brownian system can be written as

$$F_{\mathrm{inc,bm}}(\mathbf{q},t) = \sum_{\alpha=1}^{N} b_{\alpha,\mathrm{inc}}^2 f_{\alpha\alpha}(\mathbf{q},\infty) f'_{\alpha\alpha}(\mathbf{q},t), \tag{3}$$

with

$$f_{\alpha\alpha}(\mathbf{q},\infty) = \exp\left(-k_{\mathrm{B}}T\mathbf{q}^{\mathrm{T}} \cdot (\mathbf{K}_{\alpha\alpha})^{+} \cdot \mathbf{q}\right), \tag{4}$$

$$f'_{\alpha\alpha}(\mathbf{q},t) = \exp\left(\sum_{k=1}^{3N-6} \exp(-\lambda_k t) y^{(k)}_{\alpha\alpha}(\mathbf{q})\right), \tag{5}$$

where

$$y^{(k)}_{\alpha\alpha}(\mathbf{q}) = k_B T \frac{(\mathbf{q}^T \cdot \mathbf{u}^{(\alpha)}_k)^2}{\lambda_k m_\alpha \mathbf{u}^T_k \cdot \mathbf{M}^{-1/2} \cdot \boldsymbol{\gamma} \cdot \mathbf{M}^{-1/2} \cdot \mathbf{u}_k} \tag{6}$$

and the eigenvalues λ_k and eigenvectors $\mathbf{u}_k$ satisfy the relation

$$\mathbf{A} \cdot \mathbf{u}_k = \lambda_k \mathbf{u}_k, \tag{7}$$

with $\mathbf{A} = \mathbf{M}^{1/2}\boldsymbol{\gamma}^{-1} \cdot \mathbf{K} \cdot \mathbf{M}^{-1/2}$, where $\mathbf{M}$ is the diagonal matrix containing the particle masses. Furthermore, $\mathbf{u}^{(\alpha)}_k$ is the displacement of atom α in the eigenvector of mode k. Note, that we have eliminated the six modes with $\lambda_k = 0$ which describe global translation and rotation of the protein.

For numerical calculations, Eq. (7) has the inconvenience that the matrix $\mathbf{A}$ is not symmetric. However, an equivalent symmetric eigenvalue problem can be constructed by defining the symmetric matrix

$$\hat{\mathbf{A}} = \boldsymbol{\gamma}^{-1/2} \cdot \mathbf{K} \cdot \boldsymbol{\gamma}^{-1/2} \tag{8}$$

and the modified eigenvectors

$$\hat{\mathbf{u}}_k = \boldsymbol{\gamma}^{1/2}\mathbf{M}^{-1/2} \cdot \mathbf{u}_k. \tag{9}$$

It can easily be verified that the vectors $\hat{\mathbf{u}}_k$ are the eigenvectors of $\hat{\mathbf{A}}$ with eigenvalues λ_k.

The essential difference between Brownian mode analysis and standard normal mode analysis is thus that the force-constant matrix $\mathbf{K}$ is not mass weighted, but friction weighted. A simple physical interpretation of the Brownian modes can be given in the absence of random forces, i.e. with no coupling to a heat bath. Then each Brownian mode k will independently move from its initial amplitude along $\mathbf{M}^{-1/2} \cdot \mathbf{u}_k$ towards the stable equilibrium configuration with a time constant of λ_k; this motion is overdamped for all modes. In the presence of random forces, all modes are always excited, such that no single-mode motion can exist. The Brownian modes then only provide a convenient description of the probability density for configurational changes.

4.2. Effective potential well

In a study of domain motions in large proteins by normal mode analysis [27,28], it was found that domain motions can be reproduced using a simple harmonic potential of the form

$$U(\mathbf{R}_1, \ldots, \mathbf{R}_N) = \sum_{\text{all pairs } \alpha,\beta} U_{\alpha\beta}(\mathbf{R}_\alpha - \mathbf{R}_\beta) \tag{10}$$

with the pair potential

$$U_{\alpha\beta}(\mathbf{r}) = k\left(\mathbf{R}^{(0)}_{\alpha\beta}\right)\left(|\mathbf{r}| - \left|\mathbf{R}^{(0)}_{\alpha\beta}\right|\right)^2. \tag{11}$$

Here and in the following $\mathbf{x}_\alpha = \mathbf{R}_\alpha - \mathbf{R}^{eq}_\alpha$ are the particle displacements with respect to the equilibrium position, and $\mathbf{R}^{(0)}_{\alpha\beta} = \mathbf{R}^{eq}_\alpha - \mathbf{R}^{eq}_\beta$ is the pair distance vector in the stable equilibrium configuration. The pair force constant was chosen as

$$k(\mathbf{r}) = \exp\left(-\frac{|\mathbf{r}|^2}{r_0^2}\right) \tag{12}$$

with $r_0 = 0.7$ nm, essentially for mathematical convenience. For the determination of domain motions, the only important feature of the potential energy well is a clear distinction between slow and fast motions; reproduction of dynamical time scales or of the total atomic fluctuation is not important. This potential can therefore not be expected to yield good results in the present study.

One can derive a force constant matrix for a C_α model from the force constant matrix for an all-atom model by assuming that for any given displacement of the C_α atoms, the other atoms move along in such a way as to minimize the potential energy. For an harmonic potential energy well, this energy-minimizing displacement vector can be determined analytically. In the following, we indicate the C_α atoms by the superscript α and the other atoms by the superscript o. We thus divide the all-atom force constant matrix $\mathbf{K}$ into four submatrices,

$$\mathbf{K} = \begin{pmatrix} \mathbf{K}^{(\alpha\alpha)} & \mathbf{K}^{(\alpha o)} \\ \mathbf{K}^{(o\alpha)} & \mathbf{K}^{(oo)} \end{pmatrix} \tag{13}$$

and the vector $\mathbf{x}$ describing the atomic displacements from the energy minimum into the subvectors $\mathbf{x}^{(\alpha)}$ and $\mathbf{x}^{(o)}$. Then for given $\mathbf{x}^{(\alpha)}$, the $\mathbf{x}^{(o)}$ are given by

$$\mathbf{x}^{(o)} = -\left(\mathbf{K}^{(oo)}\right)^{1} \cdot \mathbf{K}^{(o\alpha)} \cdot \mathbf{x}^{(\alpha)} \tag{14}$$

and the effective force constant matrix $\tilde{\mathbf{K}}$ for the C_α atoms becomes

$$\tilde{\mathbf{K}} = \mathbf{K}^{(\alpha\alpha)} - \mathbf{K}^{(\alpha o)} \cdot \left(\mathbf{K}^{(oo)}\right)^{1} \cdot \mathbf{K}^{(o\alpha)}. \tag{15}$$

The inversion of $\mathbf{K}^{(oo)}$ poses no problem in principle, because this matrix does not have zero eigenvalues. However, in practice this approach can only be used for small proteins, both because of the inversion of this large matrix and because a careful energy minimization is required before the evaluation of $\mathbf{K}$.

In order to obtain a simpler model, we assume again the functional form given by Eqs. (10) and (11), but with a different distance-dependent force constant than the one from Eq. (12). We determine this force constant from the effective force constant matrix Eq. (15) which we calculate for crambin using the AMBER 94 force field [30]. Then, we assume that each pair entry $\tilde{\mathbf{K}}_{ij}$ can be derived from a pair potential that has the form of Eq. (11) and calculate the effective pair force constant

$$k_{ij} = -\mathrm{Tr}\left(\tilde{\mathbf{K}}_{ij}\right). \tag{16}$$

By plotting this quantity against the distance between atoms α and β (Fig. 2), we find that the points can be approximated well by

$$k(r) = \begin{cases} 8.6 \times 10^5\ \mathrm{kJ\,mol^{-1}\,nm^{-3}} \cdot\ r - 2.39 \times 10^5\ \mathrm{kJ\,mol^{-1}\,nm^{-2}} & \text{for } r < 0.4\ \mathrm{nm}, \\ 128\ \mathrm{kJ\,nm^4\,mol^{-1}} \cdot r^{-6} & \text{for } r \geqslant 0.4\ \mathrm{nm}. \end{cases} \tag{17}$$

The two distance categories are required to describe the substantial difference between nearest-neighbor pairs along the backbone (distances below 0.4 nm) and all other pairs.

It can be seen from Fig. 2 that this fit yields a very good approximation of the original effective force constant matrix $\tilde{\mathbf{K}}$. However, this matrix, like the all-atom force constant matrix $\mathbf{K}$ from which it was derived, describes a single local minimum in the AMBER 94 potential energy surface. What we need for our Brownian modes study is an effective potential well that describes large-scale motions. There is no a priori reason why these two potential wells should have anything in common; the effective potential well is expected to contain many local minima that even might differ from each other significantly. However, there is evidence that the local minima are very similar to each other [31], and also, up to a global scaling factor, to the effective potential well describing large-scale motions. The evidence for this latter similarity comes from numerous normal mode studies on proteins which, in spite of being limited to a single local minimum of the potential energy surface, can reproduce experimental information that is clearly related to large-scale motions, such as domain motions or the residue dependence of atomic fluctuations, if only a scaling factor is applied to the amplitude of the motions. There seems to be a fundamental self-similarity principle in protein energy surfaces which has not yet been studied in detail.

We thus apply a global scaling factor s to our potential energy surface, which we obtain by fitting to the MD trajectory for the C_α atoms. From Eq. (4), it is evident that the long-time limit $F_{\mathrm{inc}}(\mathbf{q}, t \to \infty)$, also known as the elastic incoherent structure factor (EISF), depends only on the shape of the potential well, described by $\mathbf{K}$, but not on the friction constants $\boldsymbol{\gamma}$. We thus determine the scaling factor s by fitting to the EISF obtained from the MD trajectory; the resulting value is $s = 0.115$. The fit is shown in Fig. 3. In view of the completely different potential energy surfaces and sampling techniques in the two approaches, the agreement is remarkably good, even considering the fact that the EISF is not highly sensitive to force field details. It also provides a first indication that the assumption of an effective harmonic

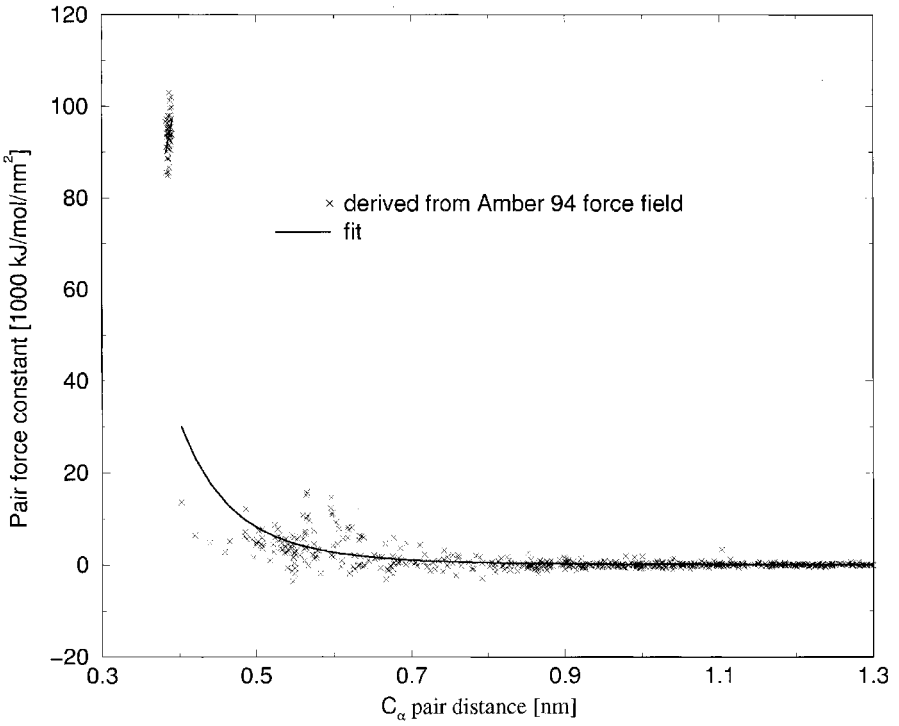

Fig. 2. The pair force constants (Eq. (16)) obtained from the effective C_α force constant matrix (Eq. (15)) derived from the AMBER 94 force field, and the fit given by Eq. (17). The force constants derived from the AMBER 94 force field can become negative because the assumption that the effective force constant matrix is a sum of pair terms is not entirely justified. Distances less than 0.4 nm correspond to neighboring C_α atoms along the backbone; the force constant in this region increases linearly with the distance. The fit for larger distances has the form a/r^6.

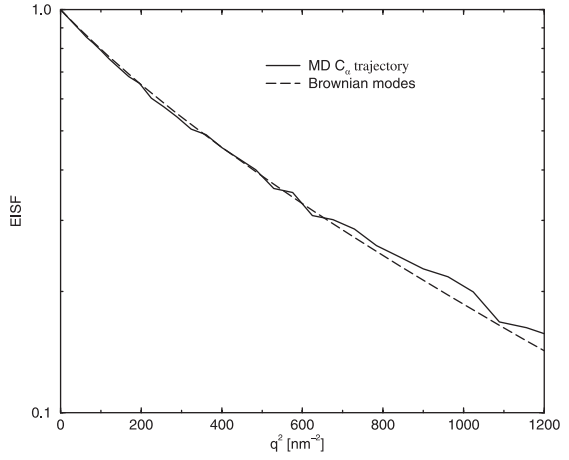

Fig. 3. The EISF from the MD trajectory for the C_α atoms compared to the results from Brownian mode analysis (long-time limit of Eq. (3)).

potential for the long-time dynamics of C-phycocyanin is reasonable.

4.3. Friction constants

In the most general case, the friction matrix $\boldsymbol{\gamma}$ is an arbitrary symmetric and positive-definite $3N \times 3N$ matrix that depends on the configuration $\mathbf{x}$. In the absence of a well-defined microscopic model for the origin of friction in solvated proteins, we prefer to limit ourselves to the simplest case in which there is one friction constant γ_i or each atom; the off-diagonal elements of $\boldsymbol{\gamma}$ are assumed to be zero. We thus have only N friction parameters which can be obtained from the MD trajectory.

The most convenient dynamical quantity for fitting the friction constants is the mean-square

displacement of atom α, which can be expressed in terms of Brownian modes as

$$\left\langle [\mathbf{x}_\alpha(t) - \mathbf{x}_\alpha(0)]^2 \right\rangle = 2k_\mathrm{B}T \sum_{k=1}^{3N} \frac{\left|\hat{\mathbf{u}}_k^{(\alpha)}\right|^2}{\gamma_i \lambda_k} \left(1 - \mathrm{e}^{-\lambda_k t}\right). \tag{18}$$

For short times, we obtain

$$\left\langle [\mathbf{x}_\alpha(t) - \mathbf{x}_\alpha(0)]^2 \right\rangle = \frac{6k_\mathrm{B}T}{\gamma_\alpha} t + \mathcal{O}(t^2), \tag{19}$$

the well-known Einstein relation for Brownian motion. In principle, the friction constants can therefore be calculated from the initial linearly increasing part of the mean-square displacements for the C_α atoms, which we obtain from the MD trajectory by an efficient FFT-based method [25]. However, our model also contains a short-time vibrational component which contributes to the mean-square displacement; the presence of such a term is even visible in the simulation data, which shows oscillatory behavior for short times. It is therefore preferable to include data at longer times into the fit as well. We have used a very simple two-parameter model that describes each atom by three Brownian oscillators with the same relaxation time, i.e.

$$\left\langle [\mathbf{x}_\alpha(t) - \mathbf{x}_\alpha(0)]^2 \right\rangle = \frac{6k_\mathrm{B}T}{\gamma_\alpha} \frac{1 - \mathrm{e}^{-\lambda t}}{\lambda}. \tag{20}$$

This model is certainly not perfect, and the simulation data suffers from insufficient statistics at long times, which results in a noticeable dependence of the fitted parameters of the time interval in which the fit was performed; the friction constants can differ by up to a factor two. Fortunately, the Brownian modes are not particularly sensitive to such variations, as will be shown in Section 4.5. In the following, we use the friction constants obtained from a fit in the interval $0 < t < 40$ ps.

One could use the fitted friction constants directly in a Brownian mode calculation, but a few-parameter model is clearly preferable because it allows a transfer to other proteins. Moreover, a model for the friction constants can help to gain a better understanding of the origin of friction in macromolecular systems. A suitable model can be constructed based on the observation that the friction constants inside the protein are larger than those at the surface. This is in contradiction to a frequently used model in which the friction constants are assumed to be proportional to the solvent-exposed surface of each atom [9]. This model was developed for liquid alkanes in which practically all atoms have exposed surfaces, and whose physical properties differ fundamentally from densely packed viscoelastic systems such as proteins. There is no reason why such a model should work well for proteins, and to our knowledge, it has never been validated by experiment or simulation for this application. Since a friction constant can be interpreted as the number of collisions per time unit with other particles, it seems reasonable that there are more collisions inside the very dense protein than between the protein surface and the comparatively dilute solvent.

A possible objection to our findings is the small amount of water that was used in the simulation; a standard water box with well-defined thermodynamic parameters, which would be appropriate to model a protein in solution, might result in higher collision frequencies for the solvent-exposed atoms. However, preliminary results from a simulation of lysozyme in a periodic water box at constant pressure indicate that in such a more realistic system, the friction constants are also highest inside the protein. Clearly, this subject requires more detailed study.

We found empirically that the friction constants show a strong correlation with the quantity

$$d_\alpha = \frac{3}{4\pi R^3} \sum_{\beta=1} N m_\beta \Theta(R - |\mathbf{R}_\beta - \mathbf{R}_\alpha|), \tag{21}$$

i.e. the average density of the protein atoms inside a sphere of radius R around atom α. The solvent atoms are ignored in the calculation of d_α. We found that $R = 1.5$ nm yields a sufficiently good correlation to permit substituting the original friction constants by the simple linear relation

$$\gamma_\alpha = a + b d_\alpha, \tag{22}$$

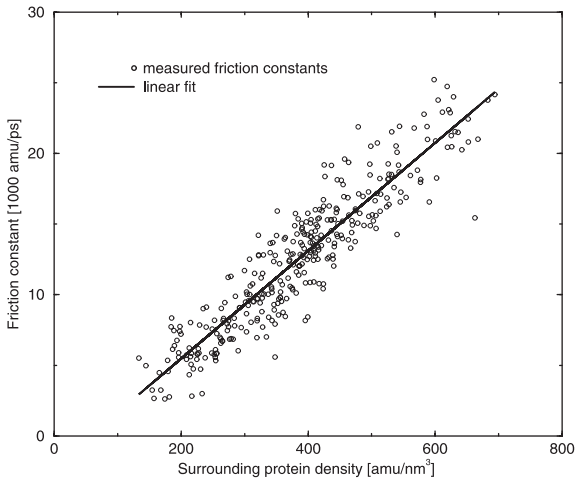

Fig. 4. The friction constants γ_α fitted from the MD trajectory vs. the average protein density around atom α (see Eq. (21)), and the linear fit to this relation given by Eq. (22).

with $a = -2160$ amu ps^{-1} and $b = 38.2$ nm^3 ps^{-1}. This linear relation and the original fitted friction constants are shown in Fig. 4.

As mentioned above, fitting the friction constants in a different time interval yields somewhat different values, but there is always a strong linear correlation with the densities d_α. For comparison, we will use the values fitted in the interval $0 < t < 100$ ps, which lead to a linear relation with the parameter $a = -500$ amu ps^{-1} and $b = 20.4$ nm^3 ps^{-1}. The friction constants are thus systematically lower and grow less with increasing density. It will be shown in Section 4.5 that the influence on the incoherent intermediate scattering function is nevertheless small.

4.4. Short-time vibrational contributions

We have mentioned before that our model consists of two contributions, diffusive motion on long-time scales, and vibrational motion on short-times scales defined by the frequency of jumps between different local minima. We assume these two contributions to be independent, such that the total incoherent intermediate scattering function is given by the product

$$F_{\mathrm{inc}}(\mathbf{q},t) = F_{\mathrm{inc,bm}}(\mathbf{q},t)F_{\mathrm{inc,vib}}(\mathbf{q},t), \tag{23}$$

with $F_{\mathrm{inc,bm}}(\mathbf{q},t)$ given by Eq. (3). We describe the short-time vibrational motion by a standard normal mode calculation for the C_α atoms, using the potential energy function described in Section 4.2 but without the scaling factor s, which was introduced specifically for describing the amplitudes of slow large-scale motions. The result of the normal mode analysis is a set of eigenfrequencies ω_k, $k = 1,\ldots,3N$, with corresponding eigenvectors $\mathbf{v}_k$. As for the Brownian modes, we eliminate the six modes with $\omega_k = 0$ that describe global translation and rotation of the protein.

The incoherent intermediate scattering factor for an harmonic oscillator is given by

$$F_{\mathrm{inc,vib}}(\mathbf{q},t) = \sum_{\alpha=1}^{N} b_{\alpha,\mathrm{inc}}^2 f_{\alpha\alpha,\mathrm{vib}}(\mathbf{q},t) \tag{24}$$

with

$$f_{\alpha\alpha,\mathrm{vib}}(\mathbf{q},t) = \exp\left(\sum_{k=1}^{3N-6}(\mathbf{q}^{\mathrm{T}}\cdot\mathbf{v}_k^{(\alpha)})^2[1-w(t)\cos\omega_k t]\right), \tag{25}$$

where we have added a windowing function $w(t)$. Without this function (i.e. $w(t) = 1$), the scattering function shows an oscillatory behavior after an initial decay. This oscillatory behavior reflects the existence of nonzero correlations over long-time scales, as is to be expected for a purely oscillatory system. However, in our model, the position correlations should vanish on a time scale corresponding to jumps between local minima. This can be achieved by introducing a Gaussian windowing function

$$w(t) = \exp\left(-\frac{t^2}{\tau^2}\right). \tag{26}$$

We do not know the relaxation time τ, of course, but its value is not critical; the only reason for introducing the windowing function is to obtain a smooth function at long times that converges to the EISF. We use a value of $\tau = 20$ ps in the results shown and discussed below.

A straightforward calculation of $F_{\mathrm{inc}}(\mathbf{q},t)$ from Eq. (23) would result in a wrong long-time limit (the EISF) because the potential energy for the

Brownian modes term alone was already scaled to reproduce the EISF. We thus have to replace the scaling coefficient s (Section 4.2) by a factor s' which takes into account the contribution of the vibrational term to the total atomic fluctuations. Since the fluctuations are proportional to $1/s$, and the vibrational potential energy has $s = 1$, we find $s' = 1/(1/s - 1) = 0.13$.

4.5. Results

We have calculated the incoherent intermediate scattering function using the potential energy function and friction constant model described above, both with pure Brownian modes (Eq. (3)) and with a combined Brownian modes/vibration approach (Eq. (23)). As for the MD calculations, we calculate an isotropic average over the directions of $\mathbf{q}$ by repeating the calculation for several randomly chosen vectors of length $q = |\mathbf{q}|$; 10 vectors per q value proved to be sufficient to obtain a converged average.

Fig. 5 shows the incoherent intermediate scattering function for $q = 5\ \mathrm{nm}^{-1}$. It is clear from this picture that Brownian modes alone cannot reproduce the MD data; the convergence to the asymptotic value is much too slow, even for low values of the friction constants. It is worth pointing out again that all the functions in Fig. 5 have the same asymptotic value, defined by the EISF. The MD results seem to be systematically lower at long times, but this is due to insufficient statistics in calculating the averages; the incoherent intermediate scattering function in fact reaches values that are far below its asymptotic limit, the EISF. This apparent contradiction is caused by different calculation procedures for F_{inc} and the EISF: the EISF is calculated as a static average over all configurations in the trajectory, whereas F_{inc} is calculated as a correlation function, with much poorer statistics for long times than for short ones.

The different Brownian mode results show that it is clearly not sufficient to get the asymptotic value and the initial slope (i.e. the friction constant) correct in order to reproduce the complete time correlation function. The very good agreement between MD and our Brownian modes approach with a vibrational contribution is therefore far from evident, and the vibrational contribution is essential for obtaining good results. Fig. 6 demonstrates that this very good agreement extends to much higher values of q, i.e. more localized motions. The highest q value shown in Fig. 6 corresponds to a wavelength of 0.25 nm, i.e. less than the nearest-neighbour C_α distance of about

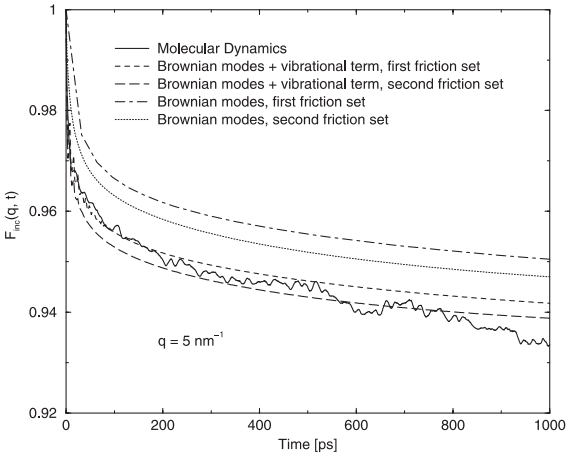

Fig. 5. The incoherent intermediate scattering function $F_{\mathrm{inc}}(\mathbf{q}, t)$ for $q = 5\ \mathrm{nm}^{-1}$. The Brownian modes alone cannot reproduce the MD curve, even for the lower friction constants. The larger friction constants together with the vibrational contribution yield a very good agreement with the MD result.

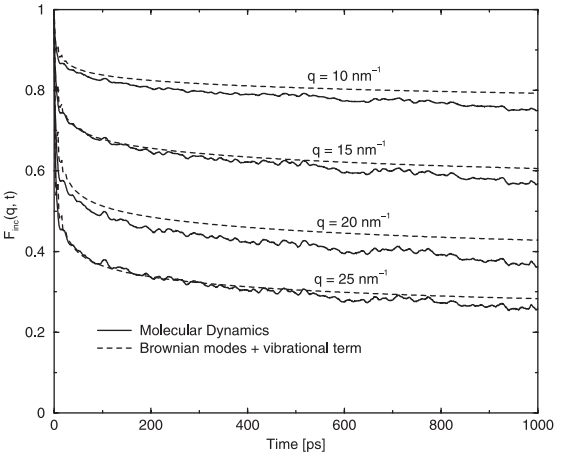

Fig. 6. The incoherent intermediate scattering function $F_{\mathrm{inc}}(\mathbf{q}, t)$ for three values of q. The Brownian mode results were obtained with the friction constants labelled "first set" in Fig. 5.

0.4 nm; an agreement beyond this value is thus not to be expected.

We thus find that the C_α dynamics of a globular protein can be reproduced with a simple harmonic model that contains only six parameters of which three were fitted to the MD trajectory and three to the AMBER 94 force field.

5. Conclusions

We have shown that the dynamics of a globular protein, as measured by the incoherent intermediate scattering function $F_{\mathrm{inc}}(q,t)$, can be described by two approximately independent contributions: whole-residue motion, described by the C_α trajectories, and rotational sidechain diffusion. Furthermore, we have shown that the first contribution can be reproduced extremely well from a model that consists of vibration in local energy minima plus jumps between minima that lead to an effective diffusive motion in a smooth harmonic potential well.

An obvious question is how general our model is, i.e. if it can be transferred to other proteins. The three parameters fitted to the AMBER 94 force field were determined for crambin and thus have already been transferred to another protein. Moreover, previous experience with normal mode calculations on simplified protein models [27] suggests that the pair interaction term should not differ among proteins. Obtaining an effective potential well for long-time diffusion from the description of a local minimum by scaling is expected to work for all proteins with a single stable conformation, but the scaling factor s might depend on the size of the protein or even on its structure. As for the friction constants, our current understanding of the origin of friction in proteins is too incomplete to permit any statement about the generality of our findings. More studies on different proteins will be required before a definite answer can be given.

The question of generality also arises for the rotational sidechain diffusion term in the incoherent intermediate scattering function. Since the sidechains and their packing density is the same in all proteins, we would expect this term to have a universal shape and a magnitude proportional to the size of the protein. Again, only further studies on different proteins can clarify this point.

Our main conclusion is that, in spite of the apparent complexity of proteins, many of their properties can be described by very simple models, which can be treated by computationally cheap operations. This has been shown earlier for atomic fluctuations [26] and domain motions [27,28], but none of these are dynamic, i.e. time-dependent, quantities. The results presented in this paper extend the applicability of simple models to dynamic quantities that are observed by neutron scattering, and thereby opens new possibilities for the simulation-aided interpretation of experimental results on large proteins.

Acknowledgements

This project was supported by a grant from the CNRS program "Physique et Chemie du Vivant". One of the authors (AJP) was partially financed by a PECO grant obtained by the Laboratoire Léon Brillouin. He also wishes to acknowledge the European Community mobility schemes and the Romanian Academy which made possible this collaboration.

References

[1] J.C. Smith, Protein dynamics: comparison of simulations with inelastic neutron scattering experiments, Q. Rev. Biophys. 24 (1991) 227.
[2] P. Martel, Biophysical aspects of neutron scattering from vibrational modes of proteins, Prog. Biophys. Mol. Biol. 57 (1992) 129.
[3] J.C. Smith, S. Cusack, B. Tidor, M. Karplus, Inelastic neutron scattering analysis of low-frequency motions in proteins: harmonic and damped harmonic models of bovine pancreatic trypsin inhibitor, J. Chem. Phys. 93 (1990) 2974.
[4] S. Hayward, A. Kitao, F. Hirata, N. Go, Effect of solvent on collective motions in globular protein, J. Mol. Biol. 234 (1993) 1207.
[5] J. Kottalam, D.A. Case, Langevin modes of macromolecules: applications to crambin and DNA hexamers, Biopolymers 29 (1990) 1409.
[6] A. Ansari, Langevin modes analysis of myoglobin, J. Chem. Phys. 110 (1999) 1774.

[7] G. Lamm, A. Szabo, Langevin modes of macromolecules, J. Chem. Phys. 85 (1986) 7334.
[8] M.C. Wang, G.E. Uhlenbeck, On the theory of Brownian motion II, Rev. Mod. Phys. 17 (1945) 323.
[9] R.W. Pastor, M. Karplus, Parametrization of the friction constant for stochastic simulations of polymers, J. Phys. Chem. 92 (1988) 2636.
[10] W. Nadler, A.T. Brnger, K. Schulten, M. Karplus, Molecular and stochastic dynamics of proteins, Proc. Natl. Acad. Sci. USA 84 (1987) 7933.
[11] R. Elber, M. Karplus, Multiple conformational states of proteins: a molecular dynamics analysis of myoglobin, Science 235 (1987) 318.
[12] H. Frauenfelder, F. Parak, R.D. Young, Conformational substates in proteins, Ann. Rev. Biophys. Biophys. Chem. 17 (1988) 451.
[13] A. Kitao, S. Hayward, N. Go, Energy landscape of a native protein: jumping-among-minima model, Proteins 33 (1998) 496.
[14] S. Dellerue, Ph.D. thesis, Université Paris Sud, submitted for publication.
[15] S. Dellerue, A.J. Petrescu, J.C. Smith, S. Longeville, M.C. Bellissent-Funel, Collective dynamics of a photosynthetic protein probed by neutron spin-echo spectroscopy and molecular dynamics simulation, Physica B, in press.
[16] S. Dellerue, A.J. Petrescu, J.C. Smith, M.C. Bellissent-Funel, in preparation.
[17] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, CHARMM: a program for macromolecular energy, minimization, and dynamics calculations, J. Comp. Chem. 4 (1983) 187.
[18] A.D. MacKerell, D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D.J. McCarthy, L. Kuchnic, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguien, B. Prudhom, W.E. Reicher, B. Roux, M. Schlenkirch, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiokiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of protein, J. Chem. Phys. B 102 (1998) 3586.
[19] M. Duerring, G.B. Schmidt, R. Huber, Isolation, crystallization, crystal structure analysis and refinement of constitutive c-phycocyanin from the chromatically adapting cyanobacterium fremyella diplosiphon at 1.66 Angstroms resolution, J. Mol. Biol. 217 (1991) 577.
[20] K. Hinsen, The molecular modeling toolkit: a new approach to molecular simulations, J. Comp. Chem., in press.
[21] S. Furois-Corbin, J.C. Smith, G.R. Kneller, Picosecond timescale rigid-helix and side-chain motions in deoxymyoglobin, Proteins 16 (1993) 141.
[22] G.R. Kneller, J.C. Smith, Liquid-like sidechain dynamics in myoglobin, J. Mol. Biol. 242 (1994) 181.
[23] G.R. Kneller, Superposition of molecular structures using quaternions, Mol. Sim. 7 (1990) 113.
[24] T. Róg, K. Hinsen, G.R. Kneller, submitted for publication.
[25] G.R. Kneller, V. Keiner, M. Kneller, M. Schiller, NMOLDYN: A program package for neutron scattering oriented analysis of molecular dynamics simulations, Comp. Phys. Comm. 91 (1995) 191, Report ILL95-KN02T, Institut Laue-Langevin, 156 X, F-38042 Grenoble Cedex, France.
[26] I. Bahar, A.R. Atilgan, B. Erman, Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential, Folding Design 2 (1997) 173.
[27] K. Hinsen, Analysis of domain motions by approximate normal mode calculations, Proteins 33 (1998) 417.
[28] K. Hinsen, A. Thomas, M.J. Field, Analysis of domain motions in large proteins, Proteins 34 (1999) 369.
[29] G.R. Kneller, Chem. Phys. 261 (2000) 1.
[30] W.D. Cornell , P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz Jr., D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, A second generation force field for the simulation of proteins and nucleic acids, J. Am. Chem. Soc. 117 (1995) 5179.
[31] A.V. Lamy, M. Souaille, J.C. Smith, Simulation evidence for experimentally detectable low temperature vibrational inhomogeneity in a globular protein, Biopolymers 39 (1996) 471.