ELSEVIER

Biophysical Chemistry 53 (1994) 131–144

Biophysical Chemistry

# Protein interactions and dynamics probed by quantum chemistry, computer simulations and neutron experiments

J.C. Smith

*Section de Biophysique des Protéines et des Membranes, Département de Biologie Cellulaire et Moléculaire, Commissariat à L'Energie Atomique, Centre d'Etudes Saclay, 91191 Gif-sur-Yvette Cedex, France*

Received 9 October 1993; accepted 10 October 1993

**Abstract**

We review some recent experiments and calculations on aspects of the structure and dynamics of proteins and related systems. The use of quantum chemical techniques to determine geometries and energies of supramolecular complexes of biological interest is illustrated, and the concomitant development of empirical energy functions for use in protein simulations outlined. We describe how simulations of crystalline peptides and amino-acids using an empirical force field can be combined with appropriate coherent and incoherent inelastic neutron scattering experiments to elucidate the characteristics of lattice vibrations and diffusive atomic motions in the crystals. The application of molecular dynamics simulations to the interpretation of incoherent neutron scattering experiments on proteins is examined and the resulting ideas on the general characteristics of protein motions discussed in terms of their functional implications.

*Keywords:* Interactions in protein; Dynamics; Inelastic neutron scattering

## 1. Introduction

Rapidly increasing available computer power has encouraged the development of numerical calculations in investigations of the chemical properties underlying protein function. In this approach quantum chemical calculations are performed on small molecules and complexes of interest and the results mapped onto empirical potential energy functions. The empirical energy functions are used to simulate the structure and dynamics of larger-scale systems including macromolecules. This approach provides a means of continuous transfer of information derived from ab initio molecular orbital theory to investigations into protein structure and dynamics.

Although the computational methods used are becoming more and more reliable there is room for substantial improvement. The empirical potential functions and simulation methodologies are imperfect. Moreover, computer power limitations restrict the volume of the configurational space of the system under investigation that can be sampled. Therefore, most researchers rely on input of experimental data to check and correct the computational results. Conversely, the simulations can often be used in the interpretation of experimental data on complex systems, and can


*SSDI* 0301-4622(1994)00084-W

provide a stepping stone from which experimental data can be analysed in terms of simplified descriptions and analytical theory.

In this article we discuss examples of the above approach to protein investigation. We illustrate the use of quantum chemical calculations in characterising the geometries and energies of supramolecular complexes of biological interest. The development of force fields for the simulation of biological macromolecules is discussed. Simulations of vibrational and diffusive motions of biomolecules in the crystalline state are described and compared to corresponding neutron scattering work. Finally, neutron scattering and theoretical work on vibrations and diffusive motion in proteins is examined and the relevance of collective and random motions to protein function discussed.

## 2. Born–Oppenheimer potential surfaces from ab initio quantum chemistry

Much information on the equilibrium geometries, conformational energetics and interactions of simple organic molecules can be gained from experiments such as gas-phase electron diffraction, X-ray crystallography and microwave spectroscopy. This information can be used to construct potential energy functions that are then transferred to larger molecules.

However, for many small molecules and supramolecular complexes of biological relevance geometries and energies are not accessible from experiments. To obtain this information quantum chemical calculations can be used. In these calculations the electrons are considered explicitly and the Schrödinger equation is solved numerically in the Born–Oppenheimer approximation for a given nuclear configuration. In the Born–Oppenheimer approximation the nuclear kinetic energy is neglected; this corresponds to the case where the electrons readjust infinitesimally quickly to a change in nuclear configurations, a valid approximation in most cases of biological interest. The results of the quantum chemical calculations are the electronic wavefunction and the energy of the system. Repeating the optimization for different nuclear configurations allows the construction of a Born–Oppenheimer potential energy surface for the system, i.e. the energy as a function of the nuclear coordinates. Using ab initio molecular orbital methods it is possible, with a sufficiently large basis set and appropriate electron correlation corrections, to reliably determine Born–Oppenheimer energy surfaces for systems with a few heavy atoms [1].

An example of the ab initio approach is recent work [2] on a complex of direct biological interest, that between water and the Schiff base moiety in the light-driven bacterial proton pump pro-

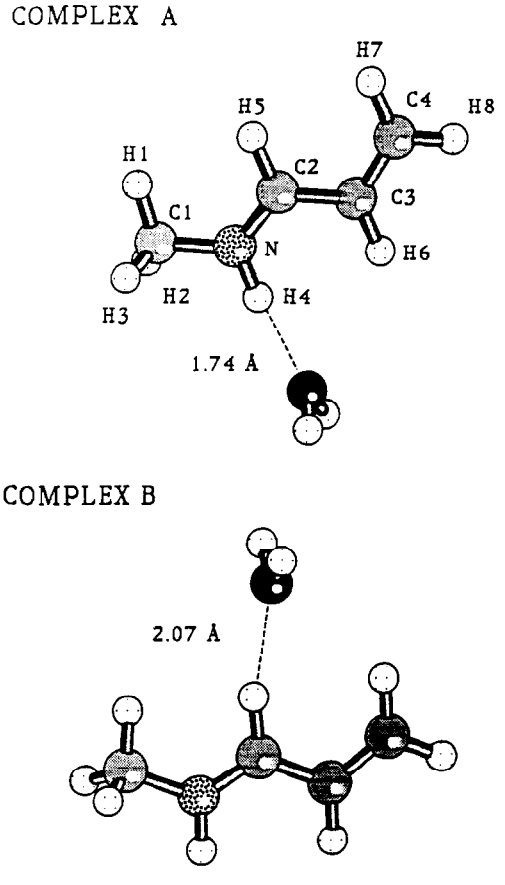


Fig. 1. MP2/6-31G** fully-optimised geometries of model Schiff-base complexes with water. Calculations were performed using the GAUSSIAN program [41]. From Ref. [2].

tein, bacteriorhodopsin. A characterization of Water–Schiff base interactions is prerequisite to a complete understanding of the mechanism of the bacteriorhodopsin photocycle. Two stable, geometry-optimized hydrogen-bonded Schiff-base: water complexes were found in the calculations and are illustrated in Fig. 1. In one of the complexes the water molecule hydrogen bonds to the NH group of the protonated Schiff base (complex **A**). On the opposite side of the molecule the water molecule hydrogen bonds to a CH group (complex **B**). The latter is an example of a C–H...O hydrogen bond, less common than other types of hydrogen bond. The interaction energies for the two complexes were found to be $\approx -18$ kcal/mol for complex **A** and $\approx -13$ kcal/mol for complex **B**. These interaction energies are large compared to typical values derived from ab initio calculations on neutral, hydrogen-bonded complexes, although the N–H...O interaction is in the same range as other cation–water hydrogen bonded species. The calculations indicate that the net positive charge on the protonated Schiff base is delocalized over the NH and the adjacent CH group. This may explain why the C–H...O interaction energy is large. When the optimized complexes were transferred into the three-dimensional structure of bacteriorhodopsin [3] it was found that they could be accommodated by the protein and that additional hydrogen bonds form between the water molecules and the side-chains of the protein.

### 2.1. Empirical potential energy functions

The information gained from quantum chemical calculations can be combined with experimental data on small molecules to construct an empirical potential energy function for simulations of larger, condensed phase systems [4–6]. An example of such an energy function is that in the CHARMM program used by our laboratory [7]. This has the following form:

$$V = \sum_{\text{bonds}} k_b(b - b_0)^2 + \sum_{\text{angles}} k_\theta(\theta - \theta_0)^2 + \sum_{\text{dihedrals}} k_\phi[1 + \cos(n\phi - \delta)] + \sum_{\text{impropers}} k_\omega(\omega - \omega_0)^2 + \sum_{i,j} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{i,j} \frac{q_i q_j}{r_{ij}}. \quad (1)$$

The force field includes *bonded interactions*, comprising bond stretches, bond angle bends, and dihedral angle contributions, and *nonbonded interactions* between pairs $(i, j)$ of atoms. In Eq. (1) $b$, $\theta$, $\phi$ and $\omega$ are the bonded lengths, angles, dihedrals and improper torsions at any given configuration and $b_0$, $\theta_0$, $\phi_0$ are reference values for these properties; the associated force constants are $k_b$, $k_\theta$, $k_\phi$ and $k_\omega$. For the intrinsic torsions $n$ is the symmetry number of the rotor (e.g., 3 for a methyl group) and $\delta$ is the phase angle. The improper dihedral contributions are used to represent out-of-plane deformations of the $sp_2$ groups.

The nonbonded interactions are included for atoms on different molecules and for atoms on the same molecule separated by three or more bonds. The nonbonded interactions between pairs of atoms $i$, $j$ consist of a van der Waals term, with well-depth and van der Waals radius, and an electrostatic interaction between partial charges $q_i$, $q_j$. The dielectric constant, $\epsilon = \epsilon_0\epsilon_r$. There is no explicit term for hydrogen bonding in Eq. (1). This is because in most cases the quantum chemical potential energy surfaces for hydrogen bonding interactions can be well modelled with a combination of the van der Waals and electrostatic terms in Eq. (1).

### 2.2. Conformational search techniques

Once a potential energy function is derived it can be put to many uses. One of these is in the determination of stable structures for biopolymers. A major problem in structure determination is that the potential energy function possesses multiple minima, and that it is presently impossible to exhaustively search the different minima of a peptide longer than five or six amino-acid residues. Therefore, computational methods that search intelligently part of the accessible conformational space have been devised.

A new method for the conformational search of biomolecules has recently been developed [8]. The method grows the peptide chains atom-by-atom or group-by-group using a Monte Carlo algorithm, replicating and deleting them according to their partial energies. The result of the calculation is a collection of configurations with energies obeying a Boltzmann distribution. An potential stumbling block for such a method is that the partial chains might not be able to find long-range interactions, such as hydrogen bonds, because these low-energy interactions are relatively far away in the sequence. As a test of the method it was applied to heptaalanine [8]. The results showed that the low-energy conformers indeed contained a full complement of hydrogen bonds. This is encouraging for future research using this type of chain-growth method.

## 3. Dynamics of small biomolecules in the crystalline state

Empirical potential energy functions can be used for simulations of systems containing a few thousands of atoms, such as a solvated protein. An important use of simulations is to provide a sampling of the phase space accessible to the system under study. This enables experimentally accessible statistical mechanical quantities to be computed as averages over the simulation. Both molecular dynamics and Monte Carlo simulations have been used in this respect. Molecular dynamics can also be used to compute time-dependent, dynamical quantities as an average over the simulation timeframes.

Examples of protein simulations will be given later. However, much useful information can be obtained by investigating condensed-phase systems consisting of peptides and amino-acids. The structurally well-characterized crystalline state is particularly useful in this respect. In the following discussion we briefly examine two molecular crystals exhibiting markedly different dynamical behaviour – zwitterionic L-alanine and the alanine dipeptide. L-alanine forms crystals with strong hydrogen-bond mediated electrostatic interactions that result in the existence of well-defined, vibrating lattice modes. Crystals of the alanine dipeptide are more anharmonic and exhibit diffusive rotational motions of the methyl groups.

*Combination of neutron scattering and molecular dynamics simulation.* In all dynamical studies neutron scattering and computer simulations were combined. The experimental quantity of interest is the dynamic structure factor [9]. Two types of scattering, coherent and incoherent, can be investigated. $S_{\mathrm{coh}}(\boldsymbol{q},\omega)$, the coherent dynamic structure factor, is written as

$$S_{\mathrm{coh}}(\boldsymbol{q},\omega)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\mathrm{d}t\ \mathrm{e}^{-\mathrm{i}\omega t}F_{\mathrm{coh}}(\boldsymbol{q},t), \tag{2}$$

$$F_{\mathrm{coh}}(\boldsymbol{q},t)=\frac{1}{N^2}\sum_{\alpha}\sum_{\beta}b_{\mathrm{coh},\alpha}b_{\mathrm{coh},\beta}\times\langle \mathrm{e}^{-\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{R}_{\alpha}(0)}\ \mathrm{e}^{\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{R}_{\beta}(t)}\rangle. \tag{3}$$

We see from Eq. (2) that $S_{\mathrm{coh}}(\boldsymbol{q},\omega)$ is the time Fourier transform of the coherent intermediate scattering function, $F_{\mathrm{coh}}(\boldsymbol{q},t)$. The sum in Eq. (2) is over the $N$ atoms, $\alpha$ in the sample. The atomic positions are specified by their time-dependent position vector operators $\boldsymbol{R}_{\alpha}(t)$. $b_{\mathrm{coh},\alpha}$ is the coherent scattering length of atom $\alpha$. Clearly, the coherent scattering contains a component due to pair correlations between different particles in the sample.

In a similar fashion, the incoherent dynamic structure factor, $S_{\mathrm{inc}}(\boldsymbol{q},\omega)$, can be expressed in terms of single particle correlations:

$$S_{\mathrm{inc}}(\boldsymbol{q},\omega)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\mathrm{d}t\ \mathrm{e}^{-\mathrm{i}\omega t}F_{\mathrm{inc}}(\boldsymbol{q},t), \tag{4}$$

$$F_{\mathrm{inc}}(\boldsymbol{q},t)=\frac{1}{N}\sum_{\alpha}b_{\mathrm{inc},\alpha}^{2}\langle \mathrm{e}^{-i\boldsymbol{q}\cdot\boldsymbol{R}_{\alpha}(0)}\ \mathrm{e}^{i\boldsymbol{q}\cdot\boldsymbol{R}_{\alpha}(t)}\rangle. \tag{5}$$

Whether a sample scatters predominantly coherently or incoherently depends on the isotopic distribution. $b_{\mathrm{inc}}$ for hydrogen is sufficiently large that hydrogen incoherent scattering dominates the measured profiles from nondeuterated proteins. Incoherent scattering can be classified as three types: elastic, quasielastic and inelastic. The inelastic scattering arises from vibrational motion in the sample. Quasielastic scattering is manifest as a broadening of the elastic peak and indicates the presence of nonvibrational, diffusive motions.

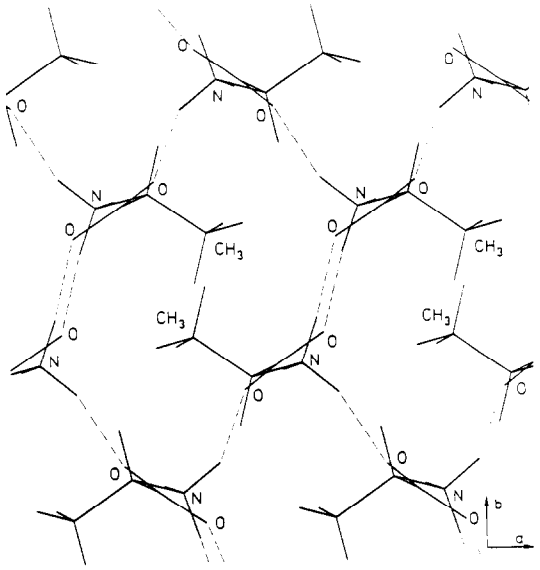


Fig. 2. Crystal geometry of zwitterionic L-alanine projected on the ($a$, $b$) plane. Hydrogen bonds are represented as dashed lines. From Ref. [10].

The elastic scattering gives information on the geometry of the motions involved and the associated atomic mean square displacements in the long time limit.

### 3.1. Lattice vibrations in L-alanine

Zwitterionic L-alanine, $CO_2^-$–$CH(CH_3)$–$NH_3^+$ is an excellent system for investigating nonbonded interactions in biological molecules. A diagram of the crystal structure in the $a$, $b$ plane is shown in Fig. 2. All the available protons of the ammonium groups are used to form hydrogen bonds with the carboxyl oxygens. One of these hydrogens links the alanine molecules together to form chains parallel to the $c$ axis of the crystal. The other two hydrogen bonds link columns of molecules in a three-dimensional network. The hydrogen bonds are anisotropic, running mostly along the $b$ direction in the $a$,$b$ projection.

The hydrogen-bonded crystal possesses well-defined lattice vibrations. The study of lattice vibrations (phonons) is standard in solid state physics but has not yet been fully exploited in biology. These intermolecular vibrations are low frequency, on the picosecond timescale, and are strongly influenced by the nonbonded interactions. A characterization of lattice vibrations in zwitterionic amino-acid crystals promises to provide particularly useful information on short and long range electrostatic interactions in proteins.

A detailed description of the dynamical behaviour of a harmonic crystal requires the determination of the dependence of the frequency, $\omega$ of the lattice modes on their wavevector, $\boldsymbol{q}$, i.e. their phonon dispersion relations. To determine these relations measurements of the coherent one-phonon inelastic scattering of neutrons by the crystal modes can be made. In recent work [10] dispersion relations in a deuterated L-alanine crystal were measured using a triple-axis neutron spectrometer at the Saclay reactor and the results compared with corresponding theoretical curves

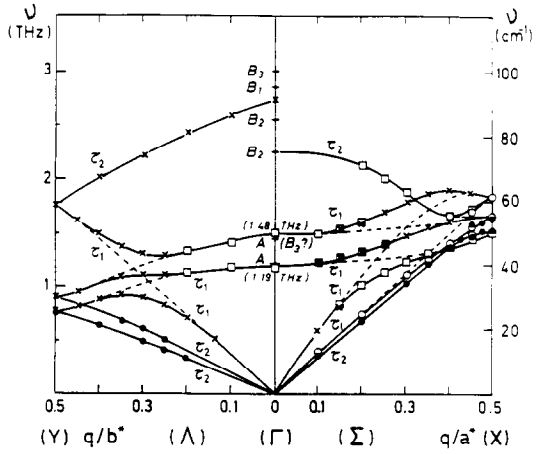


Fig. 3. Phonon dispersion curves for perdeuterated crystalline zwitterionic L-alanine determined by coherent inelastic neutron spectroscopy. The curves are along the $a^*$ and $b^*$ directions. The full circles, × symbols and open square symbols correspond to neutron data points obtained with the $a^*$, $b^*$ plane as the scattering plane and represent modes observed in predominantly transverse, purely longitudinal and mixed configurations, respectively, corresponding to measurements made around the (330) reciprocal lattice point. The open circles are data points obtained around the (220) reciprocal lattice point in a mostly transverse configuration. The minuses at $q = 0$ are Raman scattering results [11]. Solid lines indicate the most probable connectivity of the dispersion curves and dashed lines indicate branch anticrossing. From Ref. [10].

calculated using a harmonic approximation to the empirical energy function in Eq. (1).

In Fig. (3) are shown measurements of positions on the $q$, $\nu$ surface ($\nu = \omega/2\pi$) of the acoustic and lowest-frequency optic modes propagating along the $a^*$ and $b^*$ directions. Also shown are lines representing the most probable paths for the dispersion curves $\nu(q)$. From the linear portion close to $q = 0$ of the acoustic branches one can deduce the velocity of sound of the corresponding modes. The longitudinal speeds of sound along the $a^*$ and $b^*$ direction are 3.6 km/s and 4.6 km/s respectively. The $b^*$ velocity is much faster than that commonly obtained for organic compounds ($\approx 3.5$ km/s). This may be related to the geometry of the hydrogen bonds as depicted in Fig. 2. The interchain hydrogen bonds are anisotropic and run mostly along the $b$ direction. Consequently, the interchain interactions in L-alanine are expected to be strong along the $b$ axis and to result in a fast longitudinal sound velocity in this direction.

The phonon dispersion curves calculated using the empirical potential energy function in the harmonic approximation are shown in Fig. 4; this can be directly compared with its experimental counterpart, Fig. 3. In most cases there is good agreement between the forms of the measured and calculated curves. In some cases the calculated frequencies are somewhat too high. This is most apparent with the two lowest optical curves, the calculated zero-$q$ frequencies of which are around 65 $cm^{-1}$ at zero wavevector whereas experimentally they are around 45 $cm^{-1}$. Along the $b^*$ direction all the lines have approximately the same form in the experimental and calculated curves. In the $a^*$ direction the dispersion curves are somewhat more complicated; several maxima and minima and crossings are seen. Some anticrossing effects, indicated by the dashed lines in Fig. 3, were observed in the experiments and in the calculations. Anticrossing effects occur when phonon branches belonging to the same irreproducible representation exchange their eigenvectors at different values of $q$.

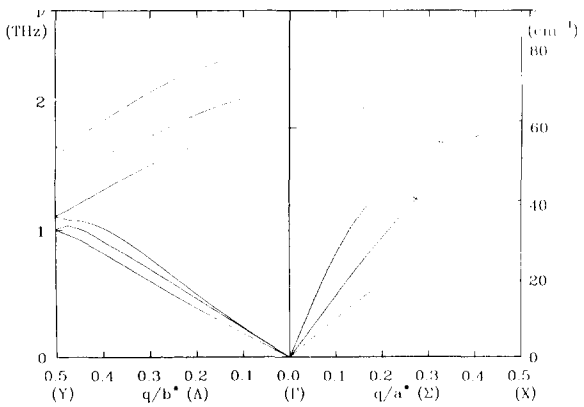


Fig. 4. Theoretical phonon dispersion curves obtained from normal model analyses of the L-alanine crystal. The diagram is directly comparable with Fig. 3. From Ref. [10].

### 3.2. Methyl group rotational dynamics in crystals of the alanine dipeptide

The rotational dynamics of methyl groups in molecular crystals serve as examples where nonbonded interactions influence librational oscillations and transitions between three fold degenerate conformers. In a recent study [12] we examined in detail the dynamics of the three methyl groups in the alanine dipeptide, $CH_3$–CONH–$CH(CH_3)$–CONH–$CH_3$.

The potential of mean force for a methyl group rotation can be considered as having two contributions; an intrinsic contribution represented by the sinusoidal dihedral term in Eq. (1) and one originating from intra- and inter-molecular nonbonded interactions. The side-chain ('hindered') methyl group has an intrinsic barrier of $\approx 3$ kcal/mol whereas the N-terminal and C-terminal methyl groups, both bonded to peptide groups, have intrinsic barriers $\approx 0$ kcal/mol. Thus, for the terminal ('free') methyl groups the rotational barriers originate from only the nonbonded interactions in the crystal.

In the recent work elastic and inelastic incoherent neutron scattering experiments and molecular dynamics simulations were combined to characterise the methyl group dynamics in the crystal over a wide range of temperatures. Neutron scattering spectra were calculated from the simulations and directly compared with the experimental profiles.

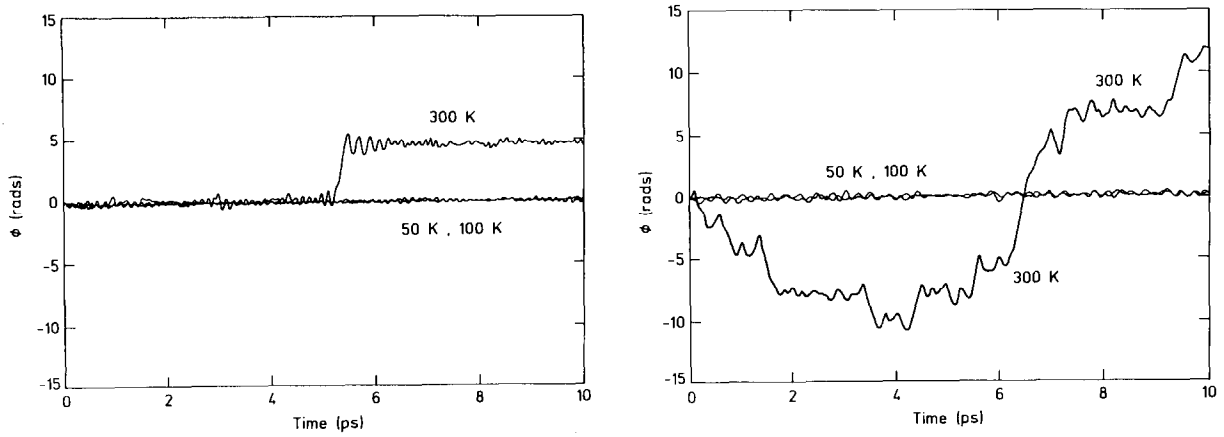


Fig. 5. Time series for methyl group dihedral angles ($C_3$ axes) in crystalline alanine dipeptide from molecular dynamics simulations of the molecule including the full crystal environment. (a) A hindered methyl; (b) a free methyl. From Ref. [12].

Time series of a free and a hindered methyl group rotational angle, calculated from molecular dynamics simulations, are shown in Fig. 5. The hindered methyl group undergoes a forced oscillation followed by jump transitions between potential wells. The free methyl group undergoes more continuous diffusion.

The different dynamical behaviours of the methyl groups are manifest in their contributions to the measured elastic incoherent structure factor (EISF). The $q$-dependence of the EISF gives information on the geometries of the motions concerned. $S_{\mathrm{inc}}(\boldsymbol{q}, \omega)$ can be divided into a part arising from purely elastic scattering with $\omega = 0$ and a part with $\omega \neq 0$.

Writing

$$F_{\mathrm{inc}}(\boldsymbol{q}, t) = \mathrm{EISF}(\boldsymbol{q}) + F'_{\mathrm{inc}}(\boldsymbol{q}, t), \tag{6}$$

$$\mathrm{EISF}(\boldsymbol{q}) = \lim_{t \to \infty} F_{\mathrm{inc}}(\boldsymbol{q}, t) \tag{7}$$

we have

$$S_{\mathrm{inc}}(\boldsymbol{q}, \omega) = \mathrm{EISF}(\boldsymbol{q})\delta(\omega) + S'_{\mathrm{inc}}(\boldsymbol{q}, \omega). \tag{8}$$

The EISF is determined by the geometry of the volume explored by the atom and is written as

$$\mathrm{EISF}(\boldsymbol{q}) = \frac{1}{N} \sum_{\alpha} |\langle e^{i\boldsymbol{q}\cdot\boldsymbol{R}_\alpha} \rangle|^2. \tag{9}$$

The EISF depends on the Fourier transform of the average probability distribution of the protons.

The molecular dynamics simulation-derived scattering curves can be decomposed into contributions from translational and rotational motion of the methyl groups. The rotational contributions are shown in Fig. 6. A characteristic spherical Bessel function behaviour is visible in the 300 K free methyl EISF($q$). That of the hindered methyl is not yet fully developed in the simulation.

The direct comparison with experiment of the simulation-derived dipeptide dynamic structure

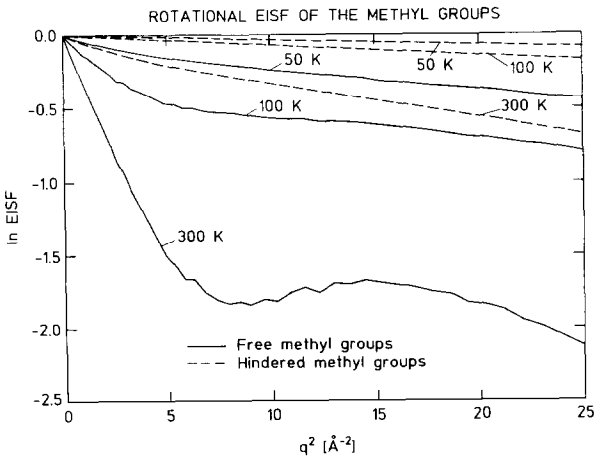


Fig. 6. Rotational components to the simulation-derived EISF($q$) for the alanine dipeptide. From Ref. [12].

factor demonstrates the usefulness of the simulation approach for decomposing and interpreting contributions to observed quantities. However, in the case of the dipeptide the simulations were not long enough to sample all the configurational space accessible in the experiment. This sampling problem was taken into account in comparisons with experiment [12].

The sampling problem was not present in another application, to polyacetylene, of the combined molecular dynamics/neutron scattering analysis [13]. This investigation was of the low-frequency vibrations in the system; these are sufficiently fast that they were well sampled in the simulation. The low-frequency vibrations in crystalline polyacetylene are anisotropic and depend on the conformation of the molecule, *cis* or *trans*. It was found that molecular dynamics simulations could be used effectively to interpret the measured vibrational density of states, and that both the anisotropy and the conformation-dependence of the measured scattering were well reproduced by the simulations. The results may be of use in understanding the role of electron–phonon coupling in highly conducting Na-doped polyacetylene.

## 4. Dynamics of folded proteins

X-ray crystallography and NMR spectroscopy furnish detailed information on the average structures of folded proteins. However, it is becoming increasingly evident that the structural picture of proteins must be supplemented with dynamical information on the fluctuations of the atoms about their average positions. Vibrational and diffusive internal dynamics are present in proteins and must be characterized. Although dynamical events in folded proteins occur on a wide range of timescales, picoseconds (ps) motions make a particularly important contribution to the internal fluctuations of the atoms from their mean positions. These 'low-frequency' protein motions have been extensively investigated using computer simulation techniques [14,15]. Incoherent neutron scattering has been able to provide complementary experimental information on the timescales, forms and amplitudes of the picosecond-timescale dynamics. Recent progress was made possible by comparing experimentally-derived spectra with those calculated from the results of simulations. A review of comparisons between incoherent neutron scattering experiments and simulations has been published [16].

*Low-frequency vibrations in proteins.* The simplicity of interpretation and the existence of analytical expressions for dynamic and thermodynamic quantities makes the normal mode description of protein dynamics a good starting point for comparison of theoretical and experimental inelastic neutron scattering spectra. In addition, the relatively large amplitude of the low-frequency internal vibrations in proteins makes it of considerable interest to probe them experimentally. Inelastic neutron scattering is well placed for such an investigation because the intensity of the scattering from a vibration depends directly on its *amplitude*; the low-frequency vibrations that dominate the mean-square fluctuations in the harmonic models will also scatter neutrons intensely.

Detailed analyses have been made of the neutron scattering properties expected from harmonic and damped harmonic models of the bovine pancreatic trypsin inhibitor, BPTI [17–19]. Experimental spectra were obtained and it was shown that the amplitudes and frequencies of the low-frequency (10–200 $cm^{-1}$) modes calculated from harmonic model are in good agreement with experiment. However, the frequencies of the very low-frequency modes ($< 50$ $cm^{-1}$) were found to be highly sensitive to the method of representation of the long-range electrostatic interactions.

At the available instrumental resolution low-frequency ($< 80$ $cm^{-1}$) inelastic spectra from the various proteins measured so far (lysozyme, BPTI, myoglobin, cytochrome C, phosphoglycerate kinase, etc.) have similar forms. The lowest frequency vibrations ($< 15$ $cm^{-1}$) do not appear as inelastic scattering in the spectrum. If present, they must be overdamped. This has particular implications for double-lobed proteins such as lysozyme. The characteristic lysozyme hinge-bending motion, if vibrational, would show up as a strong peak in the inelastic neutron scattering spectrum. No such peak is seen [16]; even in the

dried protein the hinge-bending mode, if present, is overdamped.

The addition of water molecules to a dry protein can be expected to produce a change in the internal protein dynamics. Neutron experiments on protein powders indicate that hydration increases the picosecond-timescale mean square displacements and the quasielastic scattering from small proteins [18–20]. This may be the result of a water-induced increase in the rate of picosecond-timescale jumps between free energy minima (see below). An alternative model describes the effect of hydration as the addition of low-frequency overdamped modes to the system [18].

Optical spectroscopies also show promise for investigations in the low-frequency regime. Recently it was demonstrated that with the use of synchrotron radiation far-infrared spectra from lysozyme can be derived with good signal-to-noise ratios [21]. Moreover, a hydration-dependent change in the absorption spectrum was seen. Corresponding calculations to interpret in detail the measured spectra are underway [22].

*Diffusive motions at room temperature.* Vibrational models of isolated proteins neglect several factors likely to be present in real proteins in physiological conditions. One of these is the anharmonicity of the intramolecular potential energy surface. Intriguing data pertaining to the anharmonicity of protein motions have been obtained by determinations of the temperature-dependence of mean-square displacements and fluctuations of protein atoms. Neutron scattering [23,24] and crystallographic studies on myoglobin have indicated that the average mean-square displacement, $\langle R^2 \rangle$ increases approximately linearly with temperature below $\approx 200$ K, in accord with harmonic models for the internal dynamics. At higher temperatures $\langle R^2 \rangle$ increases more rapidly with temperature, indicating the presence of additional motions. The nonlinear increase in $\langle R^2 \rangle$ is accompanied by the presence of quasielastic neutron scattering indicating the presence of picosecond-timescale nonvibrational motions in the protein. There is increasing evidence that these additional motions are required for the functioning of some proteins. For example, crystallographic analyses of ribonuclease A have demonstrated that inhibitor molecules will bind only at temperatures above the dynamical transition [25]. The bacteriorhodopsin photo-cycle works only at temperatures where the additional motions are present [26]. Molecular dynamics simulations of a series of cyclic peptide analogs indicate that the marked flexibility differences between the different peptides arise principally from differences in the nonvibrational component of the dynamics [27].

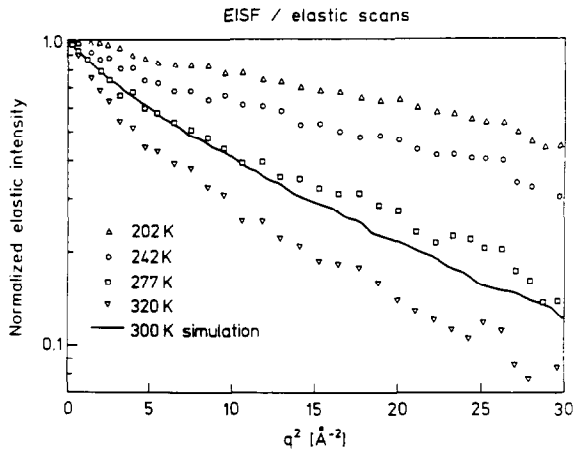


Fig. 7. Log of the elastic incoherent structure factor versus $q^2$ for myoglobin. The experimental data points are shown for four temperatures (from Ref. [23]) and the corresponding quantity derived from a molecular dynamics simulation of myoglobin at 300 K is given by a solid line. From Ref. [28].

In Fig. 7 is shown the elastic scattering as derived experimentally [23] and from a 100 ps molecular dynamics simulation of myoglobin [28]. In both experiment and simulation a nonGaussian form is seen from the elastic scattering indicating that the dynamics cannot be described in terms of all the hydrogen atoms undergoing identical harmonic motion [29]. The calculated EISF is in excellent agreement with experiment. The gradient of log(EISF) versus $q^2$ as $q \to 0$ gives the average mean square displacement of the hydrogen atoms. The low $q$ behaviour in Fig. 7 indicates that the mean square displacements are slightly underestimated in the 100 ps simulation; the low $q$ scattering matches that experimentally obtained at 277 K. This can be attributed to the fact that the hydrogen motions contributing to

the experimental profile (which is sensitive to motions faster then ~ 500 ps) are not quite completely sampled in the simulation. However, the overall agreement indicates that the amplitudes of the vibrational and diffusive motions of myoglobin are well represented in the simulations over the accessible timescale, 1–100 ps.

Two qualitatively different models for the anharmonic dynamics can be considered. In one the atoms jump between potential wells in a confined geometry. Experimental evidence for the presence of multiple minima has lead to the conclusion that at low temperatures proteins are structurally inhomogeneous in a fashion similar to the glassy state [30]. In a given sample different protein molecules explore configurations associated with different potential energy wells. The wells have been termed 'conformational substates'. At 300 K transitions between the substates occur. The 'jumps between substates' picture has been invoked in molecular dynamics simulation analyses [31,32]. It was suggested that for most atoms the harmonic approximation is valid whereas some, mostly side-chain atoms spend some time vibrating in one harmonic well before making a conformational transition into another one [31]. In myoglobin picosecond-timescale jumps between substates were described as rigid-body helix motions accompanied by loop rearrangements [32]. That effective force constants between atoms

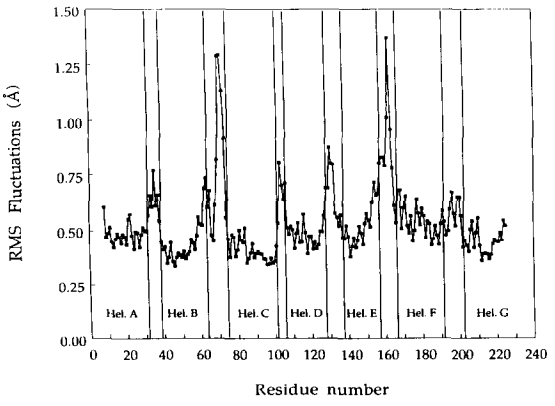


Fig. 8. Residue-dependence of the root mean square atomic fluctuations calculated from a 100 ps molecular dynamics simulation of the light-driven proton pump, bacteriorhodopsin. From Ref. [34].

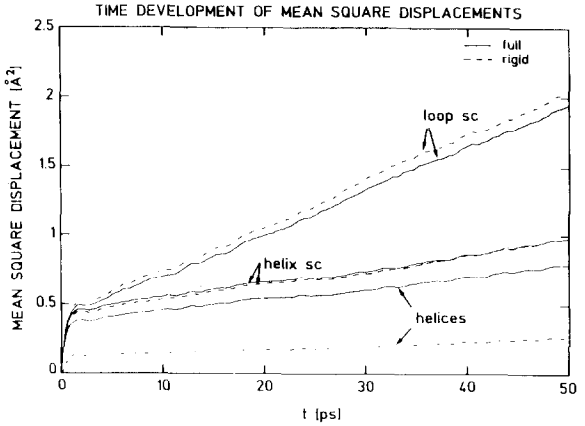


Fig. 9. Time development of the average mean square displacement (normalised per atom) from a 100 ps molecular dynamics simulation of myoglobin. From Ref. [36].

in helices are larger than for interhelix motions is substantiated by normal mode analyses of small proteins [33]. Often in protein simulations it has been observed that secondary structural elements exhibit smaller mean-square fluctuations than connecting loops. An example of this is seen in recent simulations of bacteriorhodopsin [34], as shown in Fig. 8.

As the EISF of the simulated dynamics is in good agreement with experiment (Fig. 7) it is of interest to try to identify from the simulation the nature of the nonvibrational motions concerned. The question therefore arises as to whether the full atomic trajectories can be simply described. A model alternative to substate jumps describes the 300 K dynamics as continuous diffusion of collections of atoms in confined volumes specified by the EISF. To investigate the contribution from this type of dynamics we have examined a dynamical quantity closely related to the neutron scattering – the time-dependent mean square displacement $\langle R(t)^2 \rangle$.

In Fig. 9 are presented curves of $\langle R(t)^2 \rangle$ per atom for the helices, loops and side chains. After an initial fast increase, the mean square displacements increase linearly with time. This is clear indication of the presence of *diffusive* motion in the molecule with an effective diffusion constant given by the gradient of the curves. The diffusion constant varies according to the structural ele-

ment considered. Extending the plots to timescales beyond 50 ps (not shown) leads to a reduction in the gradients although the mean square displacements do not converge in 200 ps, consistent with the presence of characteristic relaxation times of the order of hundreds of ps experimentally. Also shown as dashed lines in Fig. 9 are rigid-body contributions. These were obtained by fitting rigid-body reference structures to the time frames of the simulation. Two types of rigid-body reference structures were fitted; one consisting of the helices (backbone plus side-chains) and one consisting of the side-chains (with the centroid of the rigid body structure pinioned on the $C_\alpha$ atom).

It is clear from Fig. 9 that the rigid-helix contribution to the helix atom mean square displacements is small (about 30%) whereas the side-chain displacements are well represented by rigid-body side-chain motions. These results indicate that the major part the picosecond-timescale nonvibrational mean-square displacements arises from collision-determined diffusive motions of the side-chains acting as rigid bodies. Torsional jumps of the side-chains do occur, but too rarely to influence the $\langle R(t)^2 \rangle$ on the timescales considered. Direct calculations of the neutron scattering properties of the rigid-body trajectories have been performed and indicate that rigid side-chain diffusion dominates the nonvibrational contribution to the measured profiles [35]. A more detailed analysis of the fitted rigid-body trajectories in terms of fluctuations, mean-square displacements, velocity correlation functions and vibrational frequency distributions has been published [36].

## 5. Conclusions

Until recently the atomic-detail molecular biophysics of proteins was mostly confined to crystal structure determinations and phenomenological descriptions of these based on relative average atomic positions. Whereas the protein 'anatomical' approach has been very fruitful in unveiling structural relationships between proteins, a description of protein structure, dynamics and function directly in terms of the governing quantum and statistical mechanical principles has been, unsurprisingly, harder to achieve, The recent rapid increase in computer power and simultaneous methodological improvements have rendered approachable the self-consistent application of theoretical chemistry continuously from the quantum chemistry of small ensembles of atoms through simulations of large, biological systems.

Much information on the physical properties of proteins is to be gained from designing and carrying out neutron scattering experiments on systems amenable to computer simulation using empirical potential energy functions. The calculation of neutron scattering intensities from the simulation results and direct comparison with experiments provides a means of unequivocally interpreting the measured profiles in the cases where sufficient agreement with experiment is seen. When there are differences between experiment and theory care is often necessary in pinpointing the sources of error. These can arise from errors in the potential energy function, in the simulation methodology or in the experimental data collection and reduction.

The simplicity of neutron–nucleus interaction makes the combination of molecular dynamics (or harmonic analysis) and incoherent neutron scattering an excellent strategy for the determination of the picosecond dynamics of biomolecular systems. It is now clear from comparisons of incoherent inelastic neutron spectra calculated from normal mode analyses with experimental results that global, collective, underdamped, low-frequency vibrations do exist in proteins at all temperatures. The vibrational motion dominates the mean-square displacements at low temperatures. At higher temperatures and in solution additional, nonvibrational motions occur. Molecular dynamics simulations reproduce well the measured neutron scattering profiles for myoglobin. Further analysis of the simulations suggests that the mean-square displacements of atoms in helices and associated neutron scattering quantities are not dominated by rigid-helix motion. Rather, the nonvibrational component to protein atom motions can be considered as diffu-

sive dynamics of the side-chains acting as rigid bodies on the picosecond timescale.

Coherent neutron scattering probes pair correlations in biological macromolecules. Low resolution time-averaged radial density distribution functions can be derived from solution small angle scattering [37]. A combination of this technique with statistical mechanical theory to probe the configurational distribution of a denatured protein is reported elsewhere in this Issue [38]. The use of energy-resolved coherent inelastic scattering to determine phonon dispersion relationships in suitable biological materials, as described above, enhances our understanding of the nonbonded interactions in macromolecules and provides a challenge to theorists, who are required to reproduce detailed curves rather than single numbers as one encounters in some other branches of theoretical chemistry. The dispersion curves can be considered as energy-analysed thermal diffuse scattering as is seen associated with Bragg peaks on diffraction patterns; a close association exists between the dispersion curves and thermal diffuse scattering. Computer simulation provides a means of uniting these observations with a single model. Recently, we have been investigating very diffuse X-ray scattering in lysozyme crystals. A good agreement is seen between experimental diffuse scattering characteristics and those derived from molecular dynamics simulations and normal mode analyses of an isolated lysozyme molecule [39]. Diffuse scattering probes correlated displacements in proteins; this is an aspect of protein dynamics that is complementary to the side-chain diffusive motion accessed by the quasielastic incoherent neutron scattering discussed above.

The functional role of the various protein motions described above has still yet to be elucidated. Proteins exhibit both solid-like and liquid-like motions at physiological temperatures. Collective solid-like correlated motions clearly exist and may provide indications on accessible dynamical pathways for information transfer across proteins. For example, they have recently been clearly implicated as being coupled to functional states in photosynthetic electron transfer [40]. The picosecond-timescale diffusive motions may turn out to be largely noise with little functional relevance. However, the existence of the diffusive motions is correlated with the presence of function in some proteins. The diffusive dynamics results in the exploration by the protein of extended regions of its potential surface. This may be necessary to allow the protein to sample configurations appropriate for functional steps such as ligand binding.

## Acknowledgement

The author would like to thank past and present members of the Molecular Simulations Laboratory at Saclay for their role in developing this field at the Commissariat à L'Energie Atomique; Jerome Baudry, Anass Ech-Cherif, Michel Ferrand, Nicolas Foloppe, Sylvie Furois-Corbin, John Garen, Marc Gingold, Gerald Kneller, Alex Micu, Mafalda Nina, Edgar Rittger, Benoit Roux, Marc Souaille, Catherine Tenette, Aline Thomas and Boris Velikson. The author also thanks the numerous external collaborators who have contributed to the work outlined here.

## References

[1] W. Hehre, L. Radom, P. von R. Schleyer and J. Pople, Ab initio molecular orbital theory (Wiley–Interscience, New York, 1986).
[2] M. Nina, J.C. Smith and B. Roux, J. Mol. Struct. THEOCHEM (Special Biomolecular Edition) 286 (1993) 231–245.
[3] R. Henderson, J.M. Baldwin, T.A. Ceska, F. Zemlin, E. Beckman and K.H. Downing, J. Mol. Biol. 213 (1990) 899.
[4] J.C. Smith and M. Karplus, J. Am. Chem. Soc. 114 (1992) 801–812.
[5] N. Foloppe, J. Breton and J.C. Smith, in: NATO ASI series A: Life Sciences, Vol. 237. The photosynthetic bacterial reaction center II. Structure, spectroscopy and dynamics, eds. J. Breton and A. Vermeglio (Plenum Press, New York) pp. 43–38.
[6] M. Nina, B. Roux and J.C. Smith, in: Structures and functions of retinal proteins, Vol. 221, ed. J.L. Rigaud, (John Libbey, Eurotext, 1992) pp. 17–20.
[7] B. Brooks, R. Bruccoleri, B. Olafson, D. States, S. Swaminathan and M. Karplus, J. Comput. Chem. 4 (1993) 187.

[8] B. Velikson, T. Garel, J.C. Niel, H. Orland and J.C. Smith, J. Comput. Chem. 13 (1992) 1216.
[9] S. Lovesey (1984), International Series of Monographs on Physics, No. 72, Theory of neutron scattering from condensed matter (Oxford Science Publications, Oxford, 1984).
[10] D. Durand, M. Field, M. Quilichini and J.C. Smith, Biopolymers 33 (1993) 725.
[11] C.H. Wang and R.D. Storms, J. Chem. Phys. 55 (1971) 5110.
[12] G.R. Kneller, W. Doster, M. Settles, S. Cusack and J.C. Smith, J. Chem. Phys. 97 (1992) 8864.
[13] A.J. Dianoux, G.R. Kneller, J.L. Sauvajol and J.C. Smith, J. Chem. Phys. 99 (1993) 15586–15595.
[14] R. Lavery, J.L. Rivail and J.C. Smith, eds., American Institute of Physics Conference Proceedings Vol. 239, Advances in biomolecular simulations, (1991).
[15] C.L. Brooks III, M. Karplus and M. Pettitt, Advances in Chemical Physics, Vol. 71, eds. I. Prigogine, S.A. Rice (Wiley, New York, 1988).
[16] J.C. Smith, Q. Rev. Biophys. 24(3) (1991) 227–291.
[17] J.C. Smith, S. Cusack, U. Pezzeca, B. Brooks and M. Karplus, J. Chem. Phys. 85 (1986) 3636.
[18] J.C. Smith, S. Cusack, B. Tidor and M. Karplus, J. Chem. Phys. 93 (1990) 2974.
[19] S. Cusack, J.C. Smith, J.L. Finney, B. Tidor and M. Karplus, J. Mol. Biol. 202 (1988) 903.
[20] J.C. Smith, S. Cusack, P. Poole and J.L. Finney, J. Biomol. Struct. Dyn. 4 (1987) 583.
[21] K.D. Moeller, G.P. Williams, S. Steinhauser, C. Hirshmugl and J.C. Smith, Biophys. J. 61 (1992) 276–280.
[22] M. Souaille, E. Rittger and J.C. Smith, in preparation.
[23] W. Doster, S. Cusack and W. Petry, Nature 337 (1989) 754.
[24] W. Doster, S. Cusack and W. Petry, Phys. Rev. Letters 65 (1989) 1080.
[25] B.F. Rasmussen, A.M. Stock, D. Ringe and G.A. Petsko, Nature 357 (1992) 423.
[26] M. Ferrand, A.J. Dianoux, W. Petry and G. Zaccai, Proc. Natl. Acad. Sci. USA, in press.
[27] A. Thomas, B. Roux and J.C. Smith, Biopolymers 33 (1993) 1249.
[28] J.C. Smith and G.R. Kneller, Mol. Simul. 10 (1993) 363–375.
[29] J.C. Smith, K. Kuczera and M. Karplus, Proc. Natl. Acad. Sci. USA, 87 (1990) 1601.
[30] H. Frauenfelder, F. Parak and R.D. Young, Ann. Rev. Biophys. Chem. 17 (1988) 451.
[31] T. Ichiye and M. Karplus, Proteins. Struct. Funct. Gen. 2 (1988) 236.
[32] R. Elber and M. Karplus, Science 235 (1987) 318.
[33] T. Nishikawa and N. Go, Proteins Struct. Funct. Gen. 2 (1987) 308.
[34] M. Ferrand, G. Zaccai, M. Nina, J.C. Smith, C. Etchebest and B. Roux, FEBS Letters 327 (1993) 256–260.
[35] G.R. Kneller and J.C. Smith, J. Mol. Biol., submitted for publication.
[36] S. Furois-Corbin, J.C. Smith and G.R. Kneller, Proteins Struct. Funct. Gen. 16 (1993) 141.
[37] G. Zaccai and B. Jacrot, Ann. Rev. Biophys. Bioeng. 12 (1983) 139.
[38] P. Calmettes, B. Roux D. Durand, M. Desmadril and J.C. Smith, J. Mol. Biol. 231 (1993) 840.
[39] P. Favre, A. Micu, D. Perahia, J. Doucet, J.C. Smith and J.P. Benoît, Nature Structural Biol. 1 (1994) 124–128.
[40] M.H. Vos, F. Rappaport, J.C. Lambry, J. Breton and J-L. Martin. Nature 363 (1993) 320.
[41] M.J. et al., GAUSSIAN 90 (Carnegie-Mellon University Pittsburgh PA 1990).