Exploring internal protein dynamics by neutron spin echo spectroscopy

Ralf Biehl,* Michael Monkenbusch and Dieter Richter

The activity of proteins is often related to configuration changes that concern single atoms or amino acids or entire subdomains within the protein. The corresponding length and timescale reach from sub-Angstrom and picoseconds to nanometers and several tens of nanoseconds and beyond. We focus here on the slow motions on several ten nanosecond timescales of complete domains and show that and how these can be accessed by means of small angle neutron scattering and neutron spin-echo spectroscopy. In particular neutron spin echo spectroscopy is able to access timescales up to several hundred nanoseconds. Further insight into domain dynamics can be achieved by modelling the dynamics in comparison with the experimental data.

Introduction

Proteins are highly dynamic objects that are the subject of numerous experimental and theoretical studies.[1] On the scale of atoms we find rather unspecific stochastic fluctuations on the picosecond timescale. On an amino acid length scale motions are connected to chemical activity during binding of substrates or to diffusive motions of amino acid loops at the exterior of the protein. Complete subdomain movements on the nanosecond timescale allow conformational changes to enable function and on microsecond timescales proteins fold to their equilibrium structure. In particular the domain movements are important for the specific function of a protein as they facilitate the transfer of atomic groups and increase the specificity of reactions.[2]

We focus here on these large scale domain movements on the timescale of several tenths of nanoseconds with several Angstrom in amplitude as they are expected for proteins with separated domains connected by hinge regions. Examples for such proteins are alcohol dehydrogenase with a cleft opening motion,[3] hexokinase with a locking domain motion[4] or phosphoglycerate kinase with a large hinge bending motion.[5,6]

There are only a few experimental techniques available that access these length and timescales like NMR[7] or fluorescence resonance energy transfer.[8] A strong point of neutron spin echo spectroscopy (NSE) compared to other non-scattering techniques is the ability to study nanosecond dynamics in the coherent scattering on different length scales by changing the scattering vector Q in a non-destructive experiment without mutation of the protein.

The ability to measure internal motions in proteins by NSE has initially been demonstrated by Alpert et al.[9] The main focus of recent publications was the interprotein diffusion,[10,11] the collective protein dynamics in membranes[12] or cluster formation.[13] Internal dynamics was revisited by Bu et al.[14] followed by successful attempts to separate interprotein diffusion and internal dynamics.[3,6] In the following we will discuss the neutron

scattering of proteins with respect to the protein structure and dynamics and how to access these by small angle neutron scattering (SANS) and NSE. Therefore we first describe, with three different sized proteins as examples (see Fig. 1), the static scattering and introduce the expected dynamics of a rigid protein to observe the change due to internal dynamics. Then we give an introduction to NSE with the necessary background correction and how to access the inelastic scattering due to protein dynamics. We close with a first model to interpret the internal dynamics by means of a normal mode analysis and some general considerations how to select a protein to observe large scale domain fluctuations.
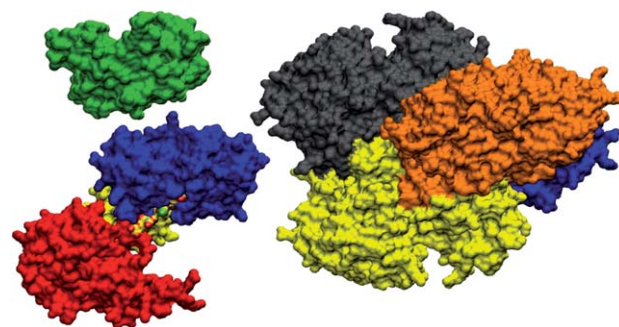

Fig. 1 Surface plots of three exemplary proteins with different size. The smallest is ribonuclease A (Rnase: PDB code 3rn3; green), an endonuclease that cleaves single-stranded RNA with a molecular weight of 13 700 Da. Below phosphoglycerate kinase (PGK: PDB code 3pgk) is a protein involved in glycolytic pathway with a molecular weight of 44 500 Da builds up by 2 main domains (blue, red) connected by a hinge (yellow) allowing fluctuations between the subdomains.[15,16] The largest protein is yeast alcohol dehydrogenase with a molecular weight of 147 kDa (ADH: PDB code 2hcy), which converts ethanol to acetaldehyde, a reaction that can be catalyzed in both directions. ADH is a tetramer with each monomer (black, orange, yellow, blue) having a dumbbell like structure with two globular domains separated by a cleft, which is visible in the yellow and black domain. The maximum length of PGK is about 8 nm.

Institute for Solid state Research, Forschungszentrum Jülich, 52425 Jülich, Germany. E-mail: ra.biehl@fz-juelich.de

Neutron scattering of proteins

Structure

The crystal structure of individual proteins with atomic resolution is determined by X-ray crystallography[17] or NMR.[18] This structure is for a large class of proteins found in the Protein Data Bank (PDB) and can be accessed by individual codes. These crystal structures build the basis for any calculation to describe the structure or the dynamics of a protein. Geometrical restrictions from the crystal structure are not present in solution and the structure may differ due to rearrangements of exterior loops or of subdomains, but the general structure is conserved in most cases.

Changes of the configuration which influence the shape of individual proteins are visible in the coherent form factor of the protein. Interference between different proteins is dependent on the spatial distribution and modifies the coherent scattering intensity $I(Q) \approx NS(Q,c)F(Q)$ for a set of N scatterers according to the structure factor $S(Q,c)$, which is dependent on the concentration c and the interactions between different proteins. $F(Q) = \langle \Sigma_{i,j} b_i b_j \exp(iQ(r_i - r_j)) \rangle$ the orientation averaged, concentration independent form factor reveals the atomic configuration and the shape of the protein. $S(Q,c)$ and $F(Q)$ depend on the scattering vector Q and the internal configuration of atoms at position r_i with scattering lengths b_i. The scattering length has to account for the exchange of labile hydrogen and for the contrast relative to the solvent, which for proteins typically is a D_2O buffer solution to increase the visibility of the protein compared to the solvent. The SANS intensity has additional contributions from the protein incoherent scattering and the coherent and incoherent scattering of the solvent buffer. Both incoherent contributions are independent of Q. The coherent contribution of D_2O is dependent on the water structure factor, which is constant for small Q. The background scattering of the buffer is measured independently and dominates the scattering at large Q limiting the accessible Q range for coherent scattering. From the protein solution scattering the background and the incoherent protein scattering have to be subtracted to result in the coherent scattering of the protein. From a measurement with different concentrations the form factor can be extracted by scaling the background corrected SANS intensity with the concentration and extrapolating the result to $c = 0$ for each individual Q. As an approximation also a low concentration below 0.3% w/v can be chosen if $S(Q) = 1$ can be assumed. With the form factor $F(Q)$ the structure factor can be calculated by $S(Q,c) = I(Q)/c/F(Q)$. For a high salinity as found in typical buffers the charges on the protein surface are screened and a hard sphere like structure factor is found.

Fig. 2a shows the form factor of the exemplary proteins in comparison to the expected buffer background for a protein solution of 1% w/v concentration. $F(Q = 0)$ shows the expected dependence on molecular weight leading to an increase with molecular weight. Above 0.08 $Å^{-1}$ the scattering signal is below the buffer background. With decreasing size the form factor shifts to larger Q reflecting the decrease of the radius of gyration R_g, which determines the slope at small Q according to the Guinier[19] approximation $F(Q) \approx \exp(-Q^2 R_g^2/3)$. For the tetramer ADH we observe a pronounced shoulder, which is
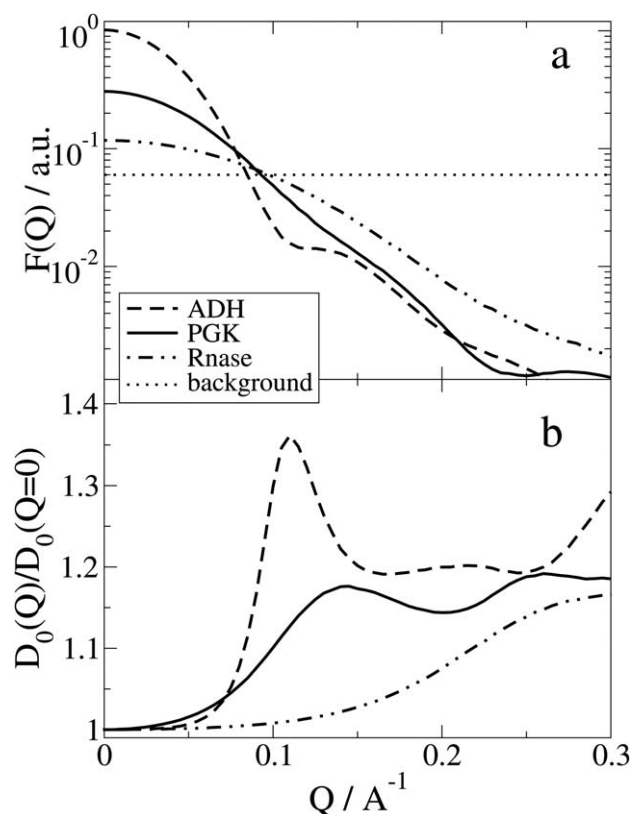

Fig. 2 (a) Form factor of ADH, PGK and Rnase in comparison to the expected background due to buffer scattering in a 1% w/v protein solution. For a concentration of 5% the expected buffer background is around 0.01. (b) $D_0(Q)/D_0(Q = 0)$ for the three proteins. The diffusion coefficient $D_0(Q = 0)$ at a temperature of 10 °C according to the HYDROPRO result is 2.7 $Å^2$ ns^{-1}, 3.94 $Å^2$ ns^{-1} and 6.3 $Å^2$ ns^{-1} for ADH, PGK and Rnase, respectively.

characteristic for the tetrameric structure of the protein. PGK has a much smaller shoulder around 0.1 $Å^{-1}$, while Rnase has no remarkable structure in the shown Q range.

Rigid protein diffusion

The dynamics of proteins in solution is dominated by the translational and rotational diffusion of the protein. The single protein diffusion may be approximated by the diffusion of a non-spherical rigid particle at a finite dilution. We consider the Brownian dynamics of a rigid body in a viscous fluid. The observable single protein diffusion $D_0(Q)$ at an infinite dilution for a neutron scattering experiment is given by (see Appendix):

$$D_0(\boldsymbol{Q}) = \frac{1}{\boldsymbol{Q}^2 F(Q)} \sum_{j,k} \left\langle b_j \exp(i\boldsymbol{Q}r_j) \begin{pmatrix} \boldsymbol{Q} \\ \boldsymbol{Q} \times \boldsymbol{r_j} \end{pmatrix} \right.$$
$$\left. \boldsymbol{D} \begin{pmatrix} \boldsymbol{Q} \\ \boldsymbol{Q} \times \boldsymbol{r_k} \end{pmatrix} b_k \exp(i\boldsymbol{Q}r_k) \right\rangle \tag{1}$$

with the 6×6 diffusion tensor $\boldsymbol{D}$ comprising translational and rotational contributions of the rigid protein. The 6×6 diffusion tensor can be split into 4 3×3 diffusion tensors with $\boldsymbol{D}_{TT}$ as the translational diffusion tensor, $\boldsymbol{D}_{RR}$ as rotational diffusion tensor and $\boldsymbol{D}_{TR} = \boldsymbol{D}^T_{RT}$ as the translation–rotation coupling. The main

contributions arise from the diagonal elements of $\boldsymbol{D}_{TT}$ and $\boldsymbol{D}_{RR}$ and correspond to the scalar diffusion coefficients as the average over the corresponding trace.

By evaluating the hydrodynamic friction of an equivalently shaped object $\boldsymbol{D}$ can be calculated by the computer code HYDROPRO[20] for a given protein structure. At large length scales (small Q) the rigid body is seen as shapeless, quasi point-like object, showing only the constant Q independent translational diffusion D_{T0} as observed by dynamic light scattering (DLS) at infinite dilution. The additional contribution from rotational diffusion leads to an increase in the effective rigid body diffusion $D_0(Q)$ above D_{T0} proportional to the rotational diffusion. It is observed, if the investigated length scale ($\sim 2\pi/Q$) is smaller than a length typical for the deviation of the protein shape from spherical symmetry. The resulting $D_0(Q)/D_0(Q = 0)$ is shown in Fig. 2b. The diffusion coefficient at a temperature of 10 °C according to the HYDROPRO result is 2.7 A^2 ns^{-1}, 3.94 A^2 ns^{-1} and 6.3 A^2 ns^{-1} for ADH, PGK and Rnase, respectively. The effective single particle diffusion $D_{eff}(Q)$ of ADH shows a pronounced maximum at the onset of the shoulder in the form factor. The tetrameric structure of the protein increases at this specific length scale the contrast between the protein subdomains enhancing the contribution of rotational diffusion on top of the translational diffusion. The shape of the tetrameric ADH is close to a spherical symmetry leading to equal diagonal elements in the rotational diffusion matrix. As a result the slope before the maximum is steep. Because of the strong asymmetry of PGK the initial increase in $D_0(Q)$ is broader reflecting the differences in the diagonal elements of the rotational diffusion ([2.6, 2.5, 3.5] × 10^6 s^{-1}) with a smaller maximum at larger Q. Rnase diffuses much faster than the larger proteins and shows the initial increase at even larger Q.

At finite concentrations interparticle interactions cannot be neglected, and the collective translational diffusion coefficient is $D_t(Q) = D_{t0}H_t(Q)/S(Q)$, where $H_t(Q)$ is the hydrodynamic function representing hydrodynamic interparticle interactions mediated by the solvent and the structure factor $S(Q)$ representing the direct interactions[21–24] as measured by SANS. D_{t0} is the Q independent single particle scalar translational diffusion. The hydrodynamics influences also the rotational diffusion D_r but independent from Q by $D_r = D_{r0}H_r$ with the single particle scalar rotational diffusion constant D_{r0}([25]).

Neutron spin-echo spectroscopy

NSE is a high-resolution inelastic scattering technique that measures the temporal and spatial correlation of atomic or molecular items that have scattering contrast for neutrons. Inelastic neutron scattering experiments on proteins are usually performed at high Q where the incoherent scattering dominates and atomic fluctuations on a picosecond timescale are measured. At low Q the coherent scattering dominates and in general slower large scale collective dynamics in the protein can be observed. Inelastic neutron scattering techniques usually measure the Fourier transform of the space-time correlation function $S(Q,\omega)$. The neutron spin-echo technique implicitly performs a cos-Fourier transformation back to (Fourier) time and therefore yields the intermediate scattering function $I(Q,t)$ in the time domain together with a high resolution in time.[26] This is achieved

by coding the incoming velocity of a neutron from a polarized beam in the spin precession angle that results after a primary coding section. After scattering from the sample the outgoing velocity is decoded again by a symmetric (de)coding section. In high resolution instruments the coding sections consist of solenoids that generate a magnetic field along the neutron path. The spin precession happens with the Larmor frequency γ, the accumulated angle is proportional to the path integral J, over the magnetic field in the coding section and the time the neutron needs to travel through it, i.e. inversely proportional to the neutron velocity. The two coding precession sections are symmetric and if a neutron does not change its velocity when scattered at the sample, the total precession angle in both arms is equal. Since the setup is such that the second arm precessions count negative the effective final angle is zero—independent of the neutron velocity. The polarization is restored (spin echo) at the point of exact symmetry between both arms. On the other hand a small velocity change during scattering creates a difference in precession angles that yields a polarization change at the end of the coding–decoding sections. Thus the instrument is sensitive to very small neutron velocity changes even when the velocity distribution of incoming neutrons is much larger (typically 10 percent FWHM). The final polarization signal is converted to intensity with the help of an analyzer in front of the detector.

The Fourier time—the correlation time in NSE experiments— $t = \gamma J m^2 \lambda^3/h^2$ is dependent on the neutron mass m_n, the third power of the neutron wavelength λ and the path integral of the magnetic field J. Depending on the expected relaxation time the wavelength can be chosen to meet the accessible times before the signal is lost in the background.

Protein scattering and background correction

The basic observable is the loss of polarization A—as amplitude of the spin echo—in a polarized neutron beam for a given Fourier time t due to an inelastic scattering of the neutrons. The initial polarization $I(Q,t = 0)$ is measured without Larmor precessions in the main magnetic field under utilization of spin flippers in the instrument setup. I_{up} is measured in a transmission setup "parallel polarizers", while I_{down} is measured with "crossed polarizers". The polarization averaged signal for $t = 0$ is $I(Q,t = 0) = (I_{up} - I_{down})/2$ and equivalent to $I(Q)$ in the previous section.

The intermediate scattering function of the protein results from the background correction as:

$$\frac{I(Q,t)}{I(Q,0)} = \frac{\dfrac{2\left(A_{sample} - TA_{bgr}\right)}{\left(I_{up,sample} - I_{down,sample}\right) - T\left(I_{up,bgr} - I_{down,bgr}\right)}}{\dfrac{2A_{ref}}{I_{up,ref} - I_{down,ref}}} \qquad (2)$$

The loss of polarization in a real measurement is partly due to imperfections of the experimental setup, different flight path of neutrons in the magnetic field or inhomogeneities of the magnetic field and imperfect flippers and polarizers. This loss of polarization can be measured with a purely elastic scatterer like carbon black powder and serve as a reference measurement to determine the instrumental resolution of the instrument. The sample measurement is normalized by this measurement (subscript ref in

eqn (2)). Due to the different attenuation between sample and background measurement we have to account for the ratio of the transmissions T of the sample and the background, which is measured in a transmission configuration ($Q = 0$).

The measured signal as the echo amplitude A is shown in Fig. 3 for PGK and corresponding background buffer measurement corrected only for the reference (see ref. 6 for experimental details on all PGK measurements). The measured signal is composed of the scattering of all components in the sample (or background) and due to the sample cell. The sample cell as a weak elastic scatterer adds a small Fourier time independent (*i.e.* elastic) contribution to the echo amplitude A and to I_{up} and I_{down}. The D_2O buffer contributes mainly to I_{up} and I_{down} but less to the echo amplitude, because the water dynamics is much faster than the time window of the experiment at low Q. As a result the background scattering (triangles) shows a more or less constant echo amplitude over the full time range, but does not increase to the value of $(I_{up} - I_{down})/2$ at $t = 0$. The value of the plateau is due to the weak elastic sample cell scattering. Nevertheless for larger Q the incoherent D_2O dynamics becomes visible at short Fourier times leading to a small increase for the largest Q in the background measurement. The larger scattering of the sample is due to the protein scattering but includes the background scattering. In general the scattering is mainly due to the coherent scattering of the protein in the buffer solution and the background is a correction of a few percent but important to determine the exact initial slope of the sample scattering. The scattered intensity at $t = 0$ reflects the combined effects of form factor and structure factor, which means usually that the intensity decreases with increasing Q for small concentrations. Typically the
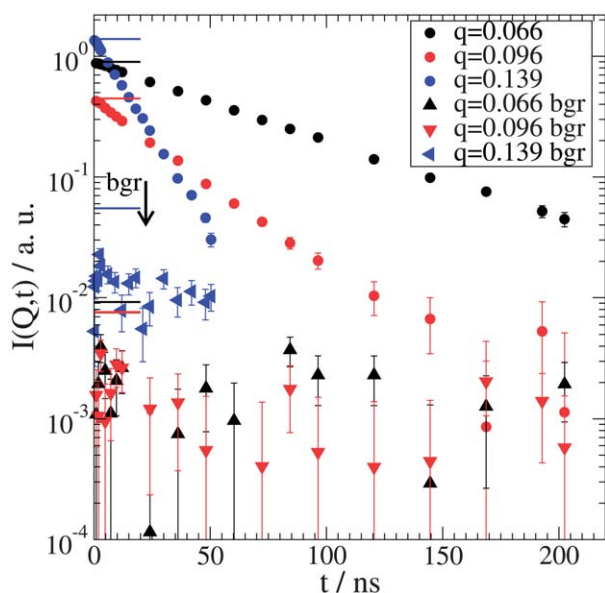
maximum accessible Fourier time t is limited; if not by the largest field integral J of the solenoids, then by the level of the background, which is reached depending on the relaxation rate for the measured Q. As a rule of thumb, the background correction can achieve a reasonable signal up to the crossover, where the coherent protein scattering equals the D_2O background in a SANS measurement, because the dominating incoherent signal of the D_2O background only contributes with 1/3 to the echo amplitude due to the spin flip in incoherent scattering.

The $Q = 0.139$ Å^{-1} data have a shorter relaxation time and are measured with a shorter wavelength (shorter maximum Fourier time with the same maximum field integral J) to increase the intensity of the incoming neutron beam and reduce the statistical error. It is clear that for a lower concentration *e.g.* 1% the signal is 5 times lower and reaches the crossover earlier with the result that only short Fourier times can be accessed within acceptable statistical limits. The crossover is also dependent on the relaxation time, $(Q^2 D_{eff})^{-1}$. By changing the temperature the viscosity and so the relaxation time can be changed to meet the accessible timescales.

The uncorrected echo amplitude of the sample does not reach the initial polarization because of the background contribution with fast dynamics (solvent) to the polarization, which, however, is missing in the echo already for the shortest available/measured Fourier time. Therefore we can use the extrapolated amplitude for $t = 0$ after correction, which should reach the initial polarization of 1, as a measure of quality for the background correction. Fig. 4 shows a semi-logarithmic plot of results for the intermediate scattering function $I(Q,t)/I(Q,t = 0)$ for 5% PGK including a bound substrate as obtained after these standard correction procedures. For the rigid protein the accessible Q and Fourier time range can be estimated assuming an exponential decay with $\exp(-Q^2 D_0(Q)t)$ according to the rigid protein diffusion with the shown accuracy in Fig. 4. It has to be taken into account that additional internal dynamics will result in a faster decay.

Analysis

The initial slope of the intermediate scattering function for $t \rightarrow 0$ comprises the dynamics due to the translational and rotational
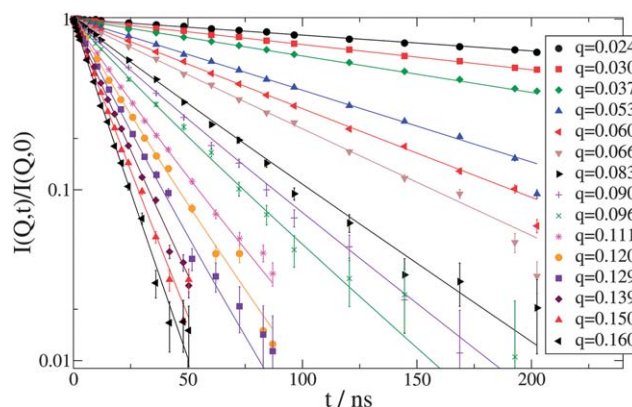

Fig. 3 Scattering intensity of a 5% PGK solution (dots) in comparison with the D_2O buffer scattering (triangle) for three exemplary Q. The data are only corrected by the reference elastic scatterer. The used wavelength is 10 Å for $Q = 0.139$ Å^{-1} and 15.9 Å^{-1} for the others. Horizontal lines indicate the polarization $(I_{up} - I_{down})/2$ corresponding to the sample or background in the same colour aside or below. The shorter wavelength has a higher intensity overcompensating the lower form factor at larger Q leading to a higher scattering for $t = 0$.

Fig. 4 Semi-logarithmic plot of the intermediate scattering function $I(Q,t)/I(Q, t = 0)$ of 5% PGK with a bound substrate. Q values are given in Å^{-1}. The lines correspond to a single least square fit including the Q dependence according to eqn (3).

diffusion of the protein together with any additional internal dynamics of the protein. The initial slope can be analysed by a cumulant analysis with $I(Q,t)/I(Q, t = 0) = \exp(K_1 t + 1/2 K_2 t^2)$ with the initial decay rate $K_1 = -Q^2 D_{eff}$. For a non-exponential decay a single exponential fit over larger times will always result in a reduced value compared to the initial decay rate. In the limit for small times $t \rightarrow 0$, $-K_1/Q^2 = D_{eff}$ represents the effective initial diffusion constant D_{eff}. For small second cumulant values K_2 (less than the statistical error), K_2 can be fixed to zero to perform a single exponential fit.

Fig. 5 shows the results for PGK and ADH in comparison to the results of dynamic light scattering experiments (as the limit for small Q) under the same conditions (see ref. 3 for experimental details on ADH). For both proteins we find a decrease at low Q values followed by an increase at higher Q. The lines represent the expectation for the rigid protein according to the results of HYDROPRO used in eqn (1) including the effect of the structure factor on the translational diffusion, which was measured independently by SANS, and the effect of the hydrodynamic function H. Already the modification due to the structure factor results in a good description of the low Q diffusion data leading to the observed decrease in $D_{eff}(Q)$ under the assumption of a constant $H_T(Q)$. H_T values of 0.81 for ADH and 0.74 for PGK were found, both smaller than the expected values for hard sphere systems around 0.91 for the same volume fraction. H_R was found to be close to 1, which is consistent to a weaker coupling of hydrodynamic interactions to rotational diffusion.[21,25] For ADH we extracted the long time limit for Fourier times 60 ns < t < 160 ns from a set of data, where the accessible timescale was long enough to yield reliable results. We find (blue data points in Fig. 5) that the long time limit is reasonably described by the assumptions made. The change in the effective diffusion is due to additional internal dynamics.

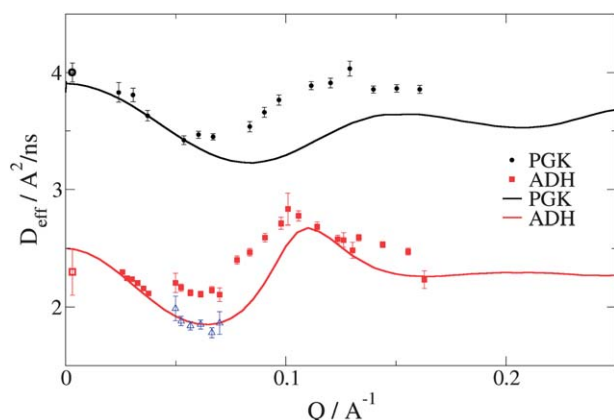From the comparison of the expected behaviour for a rigid protein we can conclude that large scale fluctuations of entire domains are mostly visible when also the rotational diffusion starts to contribute to the measured initial slope. This can be understood as large domain fluctuations mainly change the shape of the protein and therefore have the same length scale as shape fluctuations seen by the scattering vector Q due to rotations.

A direct observation of the crossover due to additional internal dynamics is possible if the internal relaxation time is well separated from the diffusion relaxation times and the amplitude is large enough to be discriminated within statistical errors. Fig. 6 shows selected Q values of PGK with bound substrate and ADH. At small Q we find for PGK a single exponential decay. For $Q = 0.11$ A^{-1} we have a faster decay due to the $Q^2 D_T$ dependence of the translational diffusion decay rate. The red broken line shows the initial slope extrapolated to longer times, which is faster than the expected decay due to the rigid protein behaviour according to eqn (1) including the corrections (blue broken line as long time behaviour extrapolated to $t = 0$). The larger protein ADH shows at $Q = 0.085$ A^{-1} a change from the initial slope (red broken line) to the long time behaviour (blue broken line), but in general slower compared to the faster diffusing PGK.

To extract these internal dynamics from the measured intermediate scattering function we can use the decomposition of rotational diffusion into spherical harmonics by Lindsay et al.[27] with scalar translational and rotational diffusion constants D_t and D_r of the rigid protein to describe the diffusion contribution including the time evolution. D_t and D_r are adjusted such that the initial slope of eqn (2) yields values that match $D_0(Q)$ of eqn (1) in order to effectively average over the anisotropy of PGK. Both
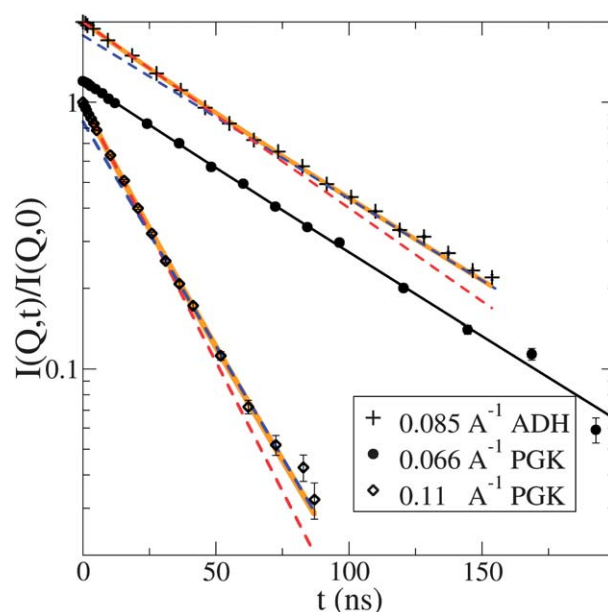

Fig. 5 Effective diffusion as a result of a second order cumulant fit for PGK at a temperature of 10 °C (black) and ADH at 5 °C (red). The blue points show the effective diffusion as a result of a single exponential fit for times t > 60 ns as the long time limit for ADH. For both proteins the first point at $Q = 0.003$ A^{-1} represents DLS measurement, which resulted in concentration independent diffusion constants below 5% w/v. The lines are expectations for the effective diffusion of a rigid protein including effects of the structure factor and the hydrodynamic function.

Fig. 6 Intermediate scattering function of ADH and PGK with substrate for selected Q. The red broken lines are the initial slope extrapolated to long times. The blue broken lines are the long time limit as expected for $D_0(Q)$ including corrections by the structure factor and the hydrodynamic function according to eqn (3) extrapolated to $t = 0$. Orange lines correspond to a fit according to eqn (3). For PGK $Q = 0.066$ A^{-1} is shifted by a factor of 1.2, for ADH $Q = 0.067$ A^{-1} is shifted by a factor of 2 for clarity.

are also corrected for interactions as described by $S(Q)$, H_T and H_R. The internal dynamics is described by an additional exponential decay with rate constant Γ. The internal dynamics contribution $A(Q)$ is taken as Q dependent resulting in:

$$\frac{I(Q,t)}{I(Q,t=0)} = [(1 - A(Q)) + A(Q)\exp(-\Gamma t)]\left(\sum_{l=0}^{\infty} S_1(Q)\right)^{-1}$$

$$\times \exp\left(-Q^2 D_t \frac{H_T}{S(Q)} t\right)$$

$$\left(\sum_{l=0}^{\infty} S_1(Q)\exp(-l(l+1)D_r H_R t)\right)$$

$$\text{with } S_1(Q) = \sum_m \left| \sum_i b_i j_l(Q r_i) Y_{l,m}(\Omega_i)\right|^2$$

$$(3)$$

$S_l(Q)$ are the terms of a multipole expansion with scattering length b_i of the atom i at position $\mathbf{r}_i$ and orientation Ω_i, $j_l(Qr)$ are the spherical Bessel functions and $Y_{l,m}$ the spherical harmonics. The fits are shown in Fig. 6 as orange lines for the selected Q. Compared to eqn (1) this treatment has the advantage that instead of the initial slope the full non-exponential decay due to rotational and translational diffusion is described and can be used in a simultaneous least square fit over all Q. If there exists one dominant motion for the internal dynamics, one characteristic time constant should be observed regardless of Q, hence we can use one Γ value for all Q values and focus on the Q dependence of the internal dynamics contribution $A(Q)$.

From this direct comparison we can extract the relaxation time of about 30 ns for PGK with substrate with an additional amplitude of about 0.12 for the high Q data in Fig. 6. Analysis of all available Q data allows the determination of the Q dependent amplitude $A(Q)$ of the specific internal protein dynamics as shown for PGK with and without substrate in Fig. 7.

Modelling of protein dynamics

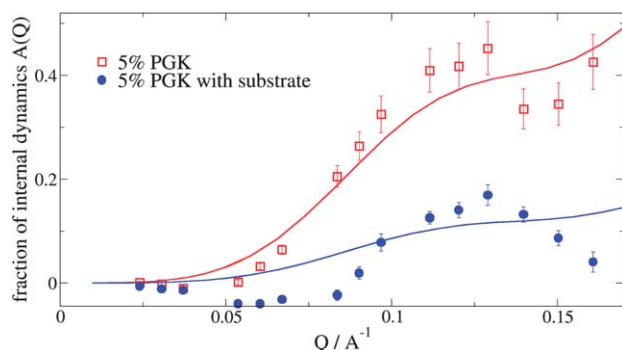To describe the effect of specific protein movements onto the intermediate scattering function we need models for the

Fig. 7 Fraction $A(Q)$ of the internal dynamics to the intermediate scattering function for PGK with substrate (blue, relaxation time 30 ns) and without substrate (red, relaxation time 60 ns). The first three normal modes display the simplest relative motions of a two body system with a flexible hinge, which are torsion and 2 perpendicular hinge motions. The corresponding description according to eqn (5) is shown as solid lines.

occurrence of large scale domain motions. Rather geometrical models with artificial lines or points in the protein as hinge or a torsion line do not respect the internal structures and forces in the protein. On the other side full molecular dynamics simulations allowing a full calculation of the intermediate scattering function along a trajectory are rare because of the huge amount of computing time necessary to get hundreds of nanoseconds to resolve possible 30 ns timescales as in the case of PGK.

An intermediate method is the usage of normal modes or even simpler elastic normal modes as overdamped movement patterns respecting the protein structure. The protein is represented by a reduced set of atoms linked with equal springs neglecting the real forcefields, if their distance is smaller than a specific cutoff distance. The used potential is harmonic around the equilibrium distance between two neighboring atoms as it is suggested by a spring model. By solving the eigen equation of the generalized force matrix the eigenvectors are obtained. The cutoff and the simplified model mainly affect the higher order modes.[28] The low normal modes are mainly determined by the shape of the protein independent of the atomic details of the interaction. The absolute value of the eigenmode frequency depends on the force constant of the individual springs. The low eigenmodes describe distortions with small relative displacements of neighboring atoms which need a minimum energy for the global deformation. These are the soft modes that allow an easy configurational change of the whole protein and can be used as displacement patterns to change the protein configuration.

In a first order approximation for small displacements along a normal mode eigenvector the scattering intensity including the effect of overdamped normal modes may be approximated by:[29]

$$I(Q,t) \propto I(Q) + \sum_\alpha a_\alpha e^{-\lambda_\alpha t} P_\alpha(Q)$$

$$P_\alpha(Q) = \left\langle \sum_{k,l}^N b_k b_l \exp(i\mathbf{Q}(\mathbf{r}_k - \mathbf{r}_l)) \cdot (\mathbf{Q}\cdot\mathbf{e}_k^a)(\mathbf{Q}\cdot\mathbf{e}_l^a)\right\rangle \quad (4)$$

where $I(Q)$ is the elastic form factor, a_α is the corresponding mode amplitude squared, e^α is the eigenvector of mode α and λ_α is the relaxation rate of the overdamped mode. For a combination of modes the amplitude of an eigenmode is weighted by the ratio of their squared eigenfrequency and the lowest eigenfrequency ω_7^2 to result in equal energies to satisfy equipartition of energies. The contribution of a set of normal modes to the intermediate scattering function $I(Q,t)/I(Q,t=0)$ is:

$$\hat{P}_\alpha(Q)e^{-\lambda_\alpha t} = \sum_\alpha a_\alpha P_\alpha(Q)e^{-\lambda_\alpha t} \left/ \left(I(Q) + \sum_\alpha a_\alpha P_\alpha(Q)\right)\right. \quad (5)$$

which is equivalent to the additional relaxation assumed in eqn (3). Calculating the initial slope results in the additional effective diffusion compared to the rigid body expectation.[3] The resulting contribution to the intermediate scattering function for PGK is shown in Fig. 7 in comparison to the experimental results as described in the Analysis section. As a result it was found that for PGK the bound substrate reduces the relaxation time from 60 ns to 30 ns in combination with a reduction of the amplitude by a factor of two. For PGK the thermal domain dynamics are necessary to reach the catalytic active configuration from the equilibrium configuration (see ref. 6 for a detailed discussion).

Any model of small displacements—respecting the real force-fields or simplifying them—extrapolated to larger elongations disregards the restrictions due to atomic excluded volumes and bond length. Also the type of relaxation is here limited to the assumed single exponential relaxation. The small displacement assumption used to result in eqn (5) disregards the large amplitudes of the motions. A full theory needs to respect the atomic restrictions and to account for the details of the scattering. Nevertheless this simple model gives a first insight into the possible principal movements and how they affect the intermediate scattering function. Better models for the possible configurational changes within thermal fluctuations and exact calculations of the resulting scattering are necessary to improve the interpretation of the underlying mechanisms.

Selecting a protein

As can be concluded from the previous sections the protein has to meet several criteria to be a good candidate to observe internal dynamics. Large scale thermal fluctuations are expected if the protein has well separated domains as in the case of PGK where in the cleft between the main domains the catalytic centre is located. In the case of ADH the two movable domains are located in each monomer, which are separated by a cleft which has to be opened and closed to bind and release the cofactor and the substrate. The amplitude of such motions can reach several Angstrom. For smaller proteins the expected amplitudes of internal motions are on a smaller length scale and move out of the accessible scale.

As a starting point biochemical characterisations have to be made to guarantee a high purity better than 90% of the protein solution and to exclude contaminations by other proteins or fragments. A main point for the data evaluation is that the protein solution has a defined state during the measurements. The aggregation to dimers and trimers or larger aggregates during the measurement time of about 3 day's beamtime (IN15, Institut Laue-Langevin, Grenoble at a wavelength of 15 Å) has to be prevented to measure the dynamics of a single species. This is possible for small concentrations resulting in a low scattering, but at higher concentrations the tendency of most proteins to have a strong aggregation will result in a mixture of protein clusters and single proteins. As a result even at low Q, where for a non-aggregating protein a single exponential decay due to translational diffusion is expected (see Fig. 6), a slowing down at large Fourier times will be observed due to the different diffusive timescales of single proteins and slower aggregates. An excellent method to observe aggregation processes under different conditions over several days is DLS.[30] If DLS is used with a CONTIN[31] analysis even smaller protein fragments and differently sized aggregates can be separated from the protein diffusion. Additionally a low Q limit at long times is measured to be compared to the NSE results. Removal of aggregates by centrifugation, fragment removal in dialysis or aggregation under different buffer conditions can easily be checked.

Increasing the protein concentration will also lead to a stronger scattering signal and improve the statistical errors. The influence of the structure factor and the hydrodynamic interactions will increase with increased concentration. For large anisotropic proteins the free rotational diffusion will be hindered

if the mean distance between proteins is smaller than the maximum length of the protein.

The expected amplitude due to fluctuations can be approximated by a Debye–Waller estimate as the reduction of the elastic signal due to small displacements u according to $Q^2 \langle u^2 \rangle / 3 = -\ln (1 - A(Q))$. If the movable domains are too small compared to the protein size the detectable amplitude $A(Q)$ will be too small to be discriminated as an independent fraction of the intermediate scattering function, because of the small mean square displacements $\langle u^2 \rangle$. Small amplitude displacements or fluctuations in small proteins can only be detected at larger Q, but due to the faster diffusion with a signal near to the background.

Very slow relaxation times prevent the independent measurement of the relaxation times, if the relaxation time $1/\Gamma$ is not separated from the diffusion relaxation time. If we subsume the diffusion contribution in eqn (3) to a single exponential decay with $\exp(-Q^2 D_{\text{eff}} t)$ the measured effective diffusion as the initial slope is $A(Q)\Gamma + Q^2 D_{\text{eff}}$. With decreasing relaxation rate Γ the additional amplitude vanishes even for large amplitudes.

If a discrimination between initial slope and long time behaviour is not expected, it is preferable to concentrate on the initial slope at short times, which can be measured at a shorter wavelength with higher intensity (e.g. a gain of 15 by changing the wavelength from 16 Å to 8 Å while decreasing the Fourier times from 200 ns to 25 ns for IN15) and reduced error bars.

Conclusions

We demonstrated with some examples the possibilities to measure the internal dynamics of large scale motions in proteins. It is not only possible to measure the change of the effectively measured diffusion in the initial slope; it is possible to measure the crossover from the combined diffusive and internal motions to the purely diffusive long time behaviour. The extraction of the internal dynamics contribution to the intermediate scattering function including the timescale of the protein motions allows new insights into the softness of proteins and into the mechanical properties of single proteins in their nearly natural environment. The contribution of the internal dynamics to promote or enable configuration changes to reach a specific catalytic active configuration or to enable binding and release of substrate and products of chemical reactions can be examined. The sensitivity of the protein to the environmental conditions can be investigated before configuration changes like partly denaturation appear and change the complete structure of the protein and prevent function. Biochemical characterisations like activity measurements can be related to direct dynamic measurements of the internal dynamics.

Appendix

Translational and rotational diffusion as seen by NSE

In general, the configuration of a protein can be described as an ensemble of N subunits (atoms or amino acids on a lower level of resolution). The general configuration is described by a $6N$ dimensional vector X with the first $3N$ being the components of the position vectors r_i and the last $3N$ being the components of orientation vectors n_i. The probability distribution $P(X,t)$ obeys

the generalized Smoluchowski equation $\partial P/\partial t = DP$ with the Smoluchowski operator:[32]

$$D = \frac{\partial}{\partial X} D \left[\frac{\partial}{\partial X} + \frac{\partial \Phi(X)}{kT \partial X} \right] \quad (A1)$$

Here, $D(X)$ is the $6N \times 6N$ diffusion matrix which incorporates hydrodynamic interactions between individual subunits and the potential $\Phi(X)$ depends on direct interactions and additional external fields.

The orientational part of the gradient operator $\partial/\partial X$ may then be expressed with the moment of momentum operator $L_i = r_i \times \partial/\partial r_i$.[30,33]

$$\partial/\partial X = \begin{pmatrix} \partial/\partial r \\ r \times \partial/\partial r \end{pmatrix} \quad (A2)$$

With the scattering length of a subunit b_j and the scattering vector Q the scattering density as seen by a neutron scattering experiment is[34] $\rho(Q,t) = \sum b_j e^{iQr_j(t)}$. The time dependent intermediate scattering function is:

$$F(Q,t) = \langle \rho(Q,t), \rho(Q,0)^* \rangle = \langle \rho(Q,t), \rho(-Q,0) \rangle \quad (A3)$$

where the time dependence $\rho(Q,t)$ is given by the adjoint Smoluchowski operator L and

$$\rho(Q,t) = e^{Lt} \rho(Q,t) \quad (A4)$$

$$L = \left[\frac{\partial}{\partial X} - \frac{\partial \Phi(X)}{kT \partial X} \right] D \frac{\partial}{\partial X} \quad (A5)$$

Inserting eqn (A4) with eqn (A5) in eqn (A3) yields for the scattering function

$$F(Q,t) = \left\langle \exp\left(\left[\frac{\partial}{\partial X} - \frac{\partial \Phi(X)}{kT \partial X} \right] D \frac{\partial}{\partial X} t \right) \rho(Q,0), \rho(-Q,0) \right\rangle \quad (A6)$$

The brackets $\langle \rangle$ represent the ensemble average.

The equilibrium distribution is $P_{eq}(X) = \exp(-\Phi(X))/Z(X)$ with $Z(X) = \int \exp(-\Phi(X))/\mathrm{d}X$. For a rigid protein configuration the potential can be described by an expansion around the equilibrium positions X^0 of the atoms in the centre of mass system of the protein with

$$\Phi(X) = \Phi(X^0) + \sum_i \frac{\partial^2 \Phi(X^0)}{\partial x_i \partial x_j} \mathrm{d}x_i \mathrm{d}x_j \quad (A7)$$

where the first derivatives are equal to zero in the equilibrium minimum of the potential. For a stepper potential minimizing deviations from the equilibrium positions, this represents a series $\delta(d) \propto \lim_{a \to 0} e^{-d^2/a^2}$ with the delta function $\delta(d)$ with the rigid protein as the limit for fixed positions. Accordingly after integration over the rigid body potential the equilibrium particle density is $\rho(r) = \sum \delta(r_i^0)$ and the scattering density in Fourier space[34] is $\rho(Q) = \sum e^{iQr_i^0}$ justifying the scattering density seen by neutrons. The potential only limits $6N - 6$ degrees of freedom as the rigid body potential only restricts relative bond length between atoms or positions relative to the centre of mass R with $r_i = R + r_i^0$. The orientational vectors $n_i = n$ are fixed in the rigid protein. The remaining degrees of freedom represent translation and rotation of the rigid configuration. The $6N \times 6N$ diffusion matrix $D(X)$ reduces to a common 6×6 diffusion matrix which

represents the diffusion of the rigid protein. Applying the gradient operator eqn (A2) on the scattering density with $\partial/\partial r_i \rho(Q,0) = Q\rho(Q,0)$ and $r_i \times \partial/\partial r_i \rho(Q,0) = r_i \times Q\rho(Q,0)$ the intermediate scattering function is:

$$F(Q,t) \propto \left\langle e^{-\begin{pmatrix} Q \\ r_i \times Q \end{pmatrix} D \begin{pmatrix} Q \\ r_k \times Q \end{pmatrix} t} \rho(Q,0), \rho(-Q,0) \right\rangle \quad (A8)$$

The brackets $\langle \rangle$ represent the ensemble average over the remaining variables and the exponent is the remaining part of the adjoint Smoluchowski operator for the rigid protein. While the integration over the position space for the single particle is 1 the orientational average can be replaced by an averaging over Q space.

First cumulant of diffusion

Using the first cumulant[34,35]

$$\Omega(Q) = \lim_{t \to 0} \frac{\partial \ln F(Q,t)}{\partial t} \quad (A9)$$

together with eqn (A3) for the scattering function yields for the first cumulant $\Omega(Q) = \langle \rho L \rho^* \rangle / \langle \rho \rho^* \rangle$.[35] With $\Omega(Q) = Q^2 D_0(Q)$ we get for the Q dependent diffusion coefficient of a single rigid object

$$D_0(q) = \frac{1}{Q^2 F(Q,0)} \sum_{j,k} \left\langle b_j e^{-iQr_j} \begin{pmatrix} Q \\ r_j \times Q \end{pmatrix} D \begin{pmatrix} Q \\ r_k \times Q \end{pmatrix} b_k e^{iQr_k} \right\rangle \quad (A10)$$

The denominator is equal to the form factor of the protein. The sum runs over all subunits respectively atoms.

References

1 J. A. McCammon and S. C. Harvey, *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, Cambridge, 1987; D. A. Case, *Curr. Opin. Struct. Biol.*, 1994, **4**, 285; L. E. Kay, *Nat. Struct. Biol.*, 1998, **5**, 513; *Neutron Scattering in Biology*, ed. J. Fitter, T. Gutberlet and J. Katsaras, Springer Verlag, Berlin, 2006.
2 M. Gerstein, A. M. Lesk and C. Chotia, *Biochemistry*, 1994, **33**, 6739.
3 R. Biehl, B. Hoffmann, M. Monkenbusch, P. Falus and S. Prevost, et al., *Phys. Rev. Lett.*, 2008, **101**, 138102.
4 W. S. Bennett and T. A. Steitz, *Proc. Natl. Acad. Sci. U. S. A.*, 1978, **75**, 4848.
5 R. D. Banks, C. C. F. Blake, P. R. Evans, R. Haser and D. W. Rice, et al., *Nature*, 1979, **279**, 773.
6 R. Inoue, R. Biehl, T. Rosenkranz, J. Fitter, M. Monkenbusch, A. Radulescu, B. Farago and D. Richter, *Biophys. J.*, 2010, **99**, 2309.
7 A. Mittermaier and L. E. Kay, *Science*, 2006, **312**, 224.
8 G. Haran, E. Haas, B. K. Szpikowska and M. T. Mas, *Proc. Natl. Acad. Sci. U. S. A.*, 1992, **89**, 11764.
9 Y. Alpert, L. Cser, B. Faragó, F. Franĕk, F. Mezei and Y. M. Ostanevich, *Biopolymers*, 1985, **24**, 1769; Z. Bu, R. Biehl, M. Monkenbusch, D. Richter and D. J. E. Callaway, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 17646.
10 S. Longeville, W. Doster and G. Kali, *Chem. Phys.*, 2003, **292**, 413.
11 C. Le Coeur and S. Longeville, *Chem. Phys.*, 2008, **345**, 298.
12 M. C. Rheinstädter, *Biointerphases*, 2008, **3**, FB83.
13 L. Porcar, et al., *J. Phys. Chem. Lett.*, 2010, **1**, 126.
14 Z. Bu, R. Biehl, M. Monkenbusch, D. Richter and D. J. E. Callaway, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 17646.
15 R. K. Scope, *The Enzymes*, ed. P. D. Boyer, Academic Press, New York, 3rd edn, 1973.
16 B. E. Bernstein, P. A. M. Michels and W. G. J. Hol, *Nature*, 1997, **385**, 275.

17 J. Drenth, *Principles of Protein X-Ray Crystallography*, Springer, New York, 3rd edn, 2007.

18 G. S. Rule and T. K. Hitchens, *Fundamentals of Protein NMR Spectroscopy*, Springer, Netherlands, 2006.

19 A. Guinier and G. Fournet, *Small Angle Scattering of X-Rays*, Wiley Interscience, New York, 1955.

20 J. G. de la Torre, M. L. Huertas and B. Carrasco, *Biophys. J.*, 2000, **78**, 719.

21 P. N. Pusey, *J. Phys. A: Math. Gen.*, 1975, **8**, 1433.

22 B. J. Ackerson, *J. Chem. Phys.*, 1976, **64**, 242.

23 B. J. Ackerson, P. N. Pusey and R. J. A. Tough, *J. Chem. Phys.*, 1981, **76**, 1279.

24 A. J. Banchio and G. Nägele, *J. Chem. Phys.*, 2008, **128**, 104903.

25 V. Degiorgio, R. Piazza and R. B. Jones, *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.*, 1995, **52**, 2707.

26 M. Monkenbusch, R. Schätzler and D. Richter, *Nucl. Instrum. Methods Phys. Res., Sect. A*, 1997, **399**, 301.

27 H. M. Lindsay, R. Klein, D. A. Weitz, M. Y. Lin and P. Meakin, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1988, **38**, 2614.

28 M. Tirion, *Phys. Rev. Lett.*, 1996, **77**, 1905.

29 A. C. Zemach and R. J. Glauber, *Phys. Rev.*, 1956, **101**, 118.

30 B. J. Berne and R. Pecora, *Dynamic Ligth Scattering*, Wiley, New York, 1976.

31 S. W. Provencher, *Comput. Phys. Commun.*, 1982, **27**, 213.

32 B. U. Felderhof and R. B. Jones, *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.*, 1993, **48**, 1048.

33 R. B. Jones, *Phys. A (Amsterdam, Neth.)*, 1988, **150**, 339.

34 A. Z. Akcasu, *Macromolecules*, 1982, **15**, 1321.

35 A. Z. Akcasu, M. Benmouna and C. C. Han, *Polymer*, 1980, **21**, 866.