ARTICLE

Methods to determine slow diffusion coefficients of biomolecules. Applications to Engrailed 2, a partially disordered protein

Rafal Augustyniak · Fabien Ferrage ·
Raphaël Paquin · Olivier Lequin ·
Geoffrey Bodenhausen

Abstract We present new NMR methods to measure slow translational diffusion coefficients of biomolecules. Like the heteronuclear stimulated echo experiment (XSTE), these new methods rely on the storage of information about spatial localization during the diffusion delay as longitudinal polarization of nuclei with long T_1 such as nitrogen-15. The new BEST-XSTE sequence combines features of Band-selective Excitation Short-Transient (BEST) and XSTE methods. By avoiding the saturation of all protons except those of amide groups, one can increase the sensitivity by 45% in small proteins. The new experiment which combines band-Selective Optimized Flip-Angle Short-Transient with XSTE (SOFAST-XSTE) offers an alternative when very short recovery delays are desired. A modification of the HSQC-edited version of the XSTE experiment offers enhanced sensitivity and access to higher resolution in the indirect dimension. These new methods have been applied to detect changes in diffusion coefficients due to dimerization or proteolysis of Engrailed 2, a partially disordered protein.

Electronic supplementary material The online version of this article (doi:10.1007/s10858-011-9510-8) contains supplementary material, which is available to authorized users.

R. Augustyniak · F. Ferrage (✉) · R. Paquin · O. Lequin · G. Bodenhausen
Département de chimie, Ecole Normale Supérieure,
24 rue Lhomond, 75231 Paris Cedex 05, France
e-mail: Fabien.Ferrage@ens.fr

R. Augustyniak · F. Ferrage · R. Paquin · O. Lequin · G. Bodenhausen
Université Pierre et Marie Curie, 4 Place Jussieu,
75252 Paris Cedex 05, France

R. Augustyniak · F. Ferrage · R. Paquin · O. Lequin · G. Bodenhausen
UMR 7203 Laboratoire des Biomolécules CNRS-UPMC-ENS,
Paris Cedex 05, France

R. Paquin · G. Bodenhausen
Institut des Sciences et Ingénierie Chimiques,
BCH, Ecole Polytechnique Fédérale de Lausanne,
1015 Lausanne, Switzerland

Present Address:
R. Paquin
CENIR—Centre de neuro-imagerie de recherche; ICM—Institut du Cerveau et de la Moelle Épinière, Hôpital Pitié-Salpêtrière, 47 bd de l'Hôpital, 75651 PARIS Cedex 13, France

Introduction

Translational diffusion coefficients can provide useful information about molecular and supramolecular systems (Cantor and Schimmel 1980). Great efforts have been made to develop and improve experimental methods to determine diffusion coefficients (Johnson 1999). Diffusion parameters can be related directly to an effective hydrodynamic radius according to the Stokes–Einstein equation, and can thus be used to predict the size and shape of molecules (Chang 2000). Although other biophysical approaches can be employed for this purpose, pulsed field gradient NMR (PFG-NMR) techniques represent a tool of choice since the measurement is performed in a homogeneous solution, at equilibrium and, if desired, under physiological conditions.

It has been known since the early days of NMR that molecular diffusion in the presence of time-independent

gradients affects the amplitudes of spin echoes (SEs) and stimulated echoes (STEs) (Hahn 1950). In their pioneering work, Stejskal and Tanner (1965) introduced time-dependent 'pulsed' magnetic field gradients. The multiple advantages of stimulated echo experiments over simple spin echoes have greatly contributed to the success of diffusion measurements. The storage of the phase information in the form of longitudinal polarization prevents evolution under scalar couplings and, most importantly, provides access to longer diffusion delays, since longitudinal relaxation is slower than transverse relaxation, particularly in biological macromolecules. The use of nuclei such as ^{15}N with long spin–lattice relaxation times allows one to extend the diffusion delay. This approach is employed in heteronuclear stimulated echo experiments (XSTE), which allows studying slowly diffusing macromolecules (Ferrage et al. 2003, 2004). Long-lived states (LLS) (Carravetta et al. 2004) associated with scalar-coupled pairs of protons have similar advantages without requiring isotopic labelling (Cavadini et al. 2005; Cavadini and Vasos 2008; Ahuja et al. 2009; Sarkar et al. 2008a, b).

Measuring translational diffusion of biopolymers like proteins provides a wealth of information. The hydrodynamic properties of globular proteins reflect their molecular mass and oligomeric state (Altieri et al. 1995). Diffusion measurements can be used to determine dissociation constants of oligomers (Danielsson et al. 2008). In disordered or unfolded proteins, diffusion coefficients provide a measure of the compactness of the structure (Choy et al. 2002; Danielsson et al. 2002) and allow one to monitor folding pathways (Buevich and Baum 2002; Li et al. 2005). Ribosomal protein synthesis could be probed by monitoring molecular diffusion (Hsu et al. 2007). Diffusion measurements can be used to probe the behaviour of proteins in crowded environments (Li et al. 2009). Undesired aggregation or proteolysis of proteins can lead to changes in diffusion coefficients. Such effects can be monitored conveniently over extensive time-scales of days without removing the sample from the spectrometer. Here, we optimize several experiments to measure diffusion coefficient of proteins in view of quality control of protein samples.

Intrinsically disordered proteins, which have been widely investigated over the last decade (Dyson and Wright 2005), are particularly unstable in solution, mainly because of their sensitivity to proteolysis. Cystein oxidation impedes proper characterization of their residual structures, dynamics and functions. NMR signals of disordered proteins are poorly dispersed and chemical shifts are expected to be fairly insensitive to structural changes, particularly if modifications occur in regions with overlapping signals. By contrast, diffusion coefficients can reveal if a protein is still in its native form.

Recently, biomolecular NMR experiments have come to benefit from longitudinal relaxation-enhanced techniques (Pervushin et al. 2002; Schanda and Brutscher 2005; Schanda et al. 2006). These experiments are based on selective manipulations of the magnetization of amide protons. If one avoids the direct perturbation of the polarization of the aliphatic protons, the polarization of the amide proton recovers more rapidly through efficient cross-relaxation with aliphatic protons. Here, we extend this approach to diffusion measurements using heteronuclear stimulated echoes. In addition, the sensitivity of HSQC-edited XSTE experiments (Orekhov et al. 1999; Choy et al. 2002) could be improved by a better use of the diffusion delay.

The separation of species with slightly different diffusion coefficients by PFG-NMR is rather poor (Jerschow and Müller 1998) since overlap in the spectra leads to low accuracy and precision in the determination of diffusion coefficients by DOSY. HSQC-edited diffusion experiments have been developed to boost the spectral resolution so that one can determine diffusion coefficients of coexisting species (Orekhov et al. 1999; Choy et al. 2002; Buevich and Baum 2002; Brand et al. 2007; McLachlan et al. 2009). This allows one to separate folded and unfolded states of a protein in very slow exchange (Choy et al. 2002). When working at low (Martinez-Viviente and Pregosin 2003) or high temperatures (Jerschow 2000), convection in the sample may have deleterious effects on the measurement of diffusion coefficients. Although pulse sequences designed to suppress the effects of convection may help, (Jerschow and Müller 1997; Jerschow 2000; Buevich and Baum 2002), the use of Shigemi tubes with restricted volumes and moderate temperatures (between 295 and 305 K) reduces misleading effects of convection. We favor the use of heteronuclear stimulated echo-type experiments (Ferrage et al. 2003, 2004; Orekhov et al. 1999; Choy et al. 2002). In order to keep the diffusion delay Δ constant, such techniques normally achieve chemical shift labelling in the indirect dimension using a constant-time evolution interval. A decrease in sensitivity is the price to pay for high resolution in the indirect dimension. In this work, the increase in the duration of the evolution interval t_1 is compensated by a (harmless) decrease in the longitudinal diffusion delay. High resolution in the indirect dimension, which is essential for partially disordered proteins, thus becomes accessible without paying an exorbitant price in sensitivity.

The new family of PFG-NMR pulse sequences is used to detect cystein oxidation and proteolysis in Engrailed 2, a homeoprotein with long disordered fragments. The construct contains a single cystein, so that oxidation results necessarily in dimerization via the formation of an intermolecular disulfide bond. The diffusion coefficient of the reduced (monomeric) form was found to be nearly twice as large as that of the oxidized (dimeric) form. HSQC-edited

experiments that provide diffusion coefficients for each residue allow one to identify the boundaries of protein fragments and identify proteolysis cleavage sites. These results were confirmed by mass spectrometry. Addition of protease inhibitors significantly increases the lifetime of our samples.

Materials and methods

Sample preparation

The recombinant chicken Engrailed 2 protein was expressed and purified as previously described (Augustyniak et al. 2011). Uniformly ^{15}N-labelled protein was used in the NMR experiments. The sample employed to study proteolysis was prepared without protease inhibitors. Likewise, the sample used to identify oxidation and dimerization was initially prepared without TCEP (tris(2-carboxyethyl)-phosphine) reducing agent. Uniformly ^{15}N-labeled human ubiquitin (1.5 mM, pH 4.5 in 50 mM acetate buffer) was used to compare the sensitivities of variants of XSTE diffusion experiments.

NMR spectroscopy

All NMR studies were carried out at 303 K on a Bruker Avance III 600 MHz spectrometer equipped with a TXI triple-resonance triple-gradient probe using 5 mm Shigemi tubes. Gradient calibration was performed by recording a 1D image of a plexiglass phantom with 1 mm diameter holes drilled at 5 mm intervals, inserted into a 5 mm NMR tube, using the gradient echo sequence described by Price (1998). The 1D image was slightly distorted due to non-uniform gradients, which led to an uncertainty on the peak amplitude of gradient pulses larger than expected from the precision of frequency difference measurements. Although the spatial distribution of the amplitude of magnetic field gradients can be mapped and taken into account in the fit of gradient echo decays (Connell et al. 2009), such an analysis was not carried out in the present study. We obtained $G_z^{max} = 0.695 \pm 0.044$ T/m (69.5 ± 4.4 G/cm, i.e., ± 6%) for the maximum amplitude of the gradient along the z-axis. This calibration is essential for the precision and accuracy of the diffusion measurements presented in this work. Error propagation of experimental noise suggests that the measurements are quite precise. Indeed, when comparing data obtained on the same apparatus, small variations are significant. However, because of the uncertainty of the gradient amplitude G_z, all the data presented here may suffer from a systematic error as large as 13%. This potential source of inaccuracy should be kept in mind when comparing results obtained with different hardware.

All encoding and decoding gradients were applied along the z-axis while most other gradients were applied in the x–y plane. This allows proper suppression of the water signal since interference of encoding and decoding gradients with WATERGATE gradient pulses is avoided. When using a probe that is only equipped with z-axis gradients, a WATERGATE sequence (Piotto et al. 1992) can be applied after the decoding gradients (Sarkar et al. 2008a, b). This may however lead to sensitivity losses and poor water suppression for some gradient amplitudes.

Longitudinal relaxation-enhanced XSTE pulse sequences are presented in Fig. 1. Both sequences are adapted from the heteronuclear stimulated echo experiment. Pulse shapes, durations and carrier frequencies are identical to those typically used in Band-selective Excitation Short-Transient (BEST) methods (Schanda et al. 2006) and band-Selective Optimized Flip-Angle Short-Transient (SOFAST) approaches (Schanda and Brutscher 2005). RE-BURP pulses where used for refocusing and inversion while E-BURP and time-reversed E-BURP pulses were employed for 90° rotations (Geen and Freeman 1991). Polychromatic PC9 pulses were used for 120° rotations (Kupce and Freeman 1994). Refocusing of the proton chemical shifts was achieved by shifting both 180° pulses from the centre of the retro-INEPT sequence by an adjustable delay δ'. Although the SOFAST-XSTE scheme must be distinguished from SOFAST-HMQC, it relies on a similar concept, with an effective 60° rotation for the proton magnetization that lies along the z-axis at the end of the longitudinal diffusion delay T_L. This delay can be as long as 1 s (Ferrage et al. 2003, 2004), a duration comparable to (or longer than) the effective longitudinal relaxation times of the amide protons. At the end of T_L, the proton polarization is close to its equilibrium value. The effective 60° rotation after the delay Δ preserves half of this proton polarization for the following scan, so that very short recovery delays can be used. Note that we must distinguish the *total diffusion delay* Δ, which is the interval between encoding and decoding gradients, from the *longitudinal diffusion delay* T_L, defined as the duration during which the phase information is stored as longitudinal polarization.

Diffusion coefficients can be measured at the residue level with the XSTE-3D sequence (see Fig. 2), in analogy to HSQC-edited XSTE experiments (Orekhov et al. 1999; Choy et al. 2002). In this experiment, a signal stemming from a given NH pair can reflect a superposition of signals of the same residue belonging to different fragments or states of the protein. Although we probe a single residue at a time, each response may thus result from a sum of fragments or states. In this case, the term 'effective diffusion coefficient' would be more appropriate. In contrast to earlier work (Choy et al. 2002), composite-pulse
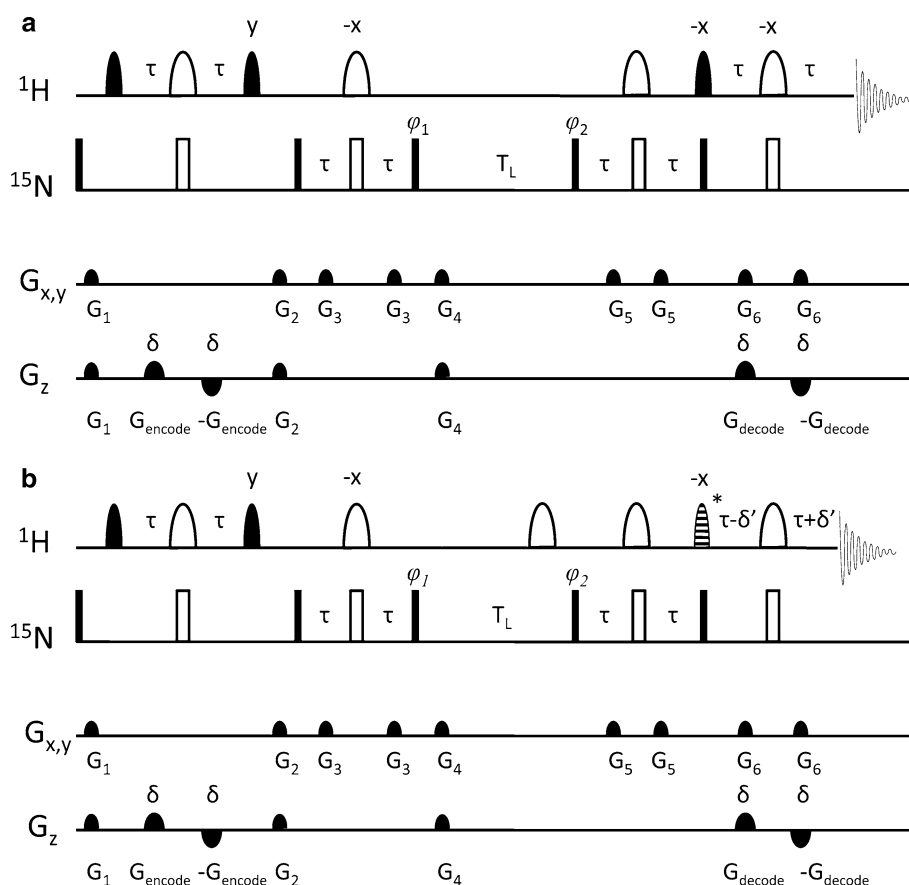
Fig. 1 Pulse sequences designed to measure diffusion coefficients of isotopically ^{15}N- or ^{13}C-enriched molecules. **a** Band-selective excitation short transient combined with heteronuclear stimulated echo (BEST-XSTE) sequence; **b** band-selective optimized flip-angle short-transient combined with heteronuclear stimulated echo (SOFAST-XSTE) sequence. *Filled* and *open symbols* indicate 90° and 180° pulses, respectively. The star identifies a polychromatic PC9 pulse (Kupce and Freeman 1994) with a 120° flip angle. Other bell-shaped pulses represent (180°) RE-BURP, (90°) E-BURP and time-reversed (90°) E-BURP pulses (Geen and Freeman 1991). The durations of shaped pulses were 2, 2.03 and 1.92 ms for PC9, RE-BURP and (time-reversed) E-BURP pulses, respectively, at 600 MHz for protons. Unless otherwise mentioned, all pulses are applied along the *x*-axis of the rotating frame. The phase cycling employed was:

$\varphi_1 = y, -y; \varphi_2 = y, y, -y, -y$. For both experiments, the receiver phase was $\varphi_{rec} = x, -x, -x, x$. Here, the total diffusion delay $\Delta = T_L + 6\tau$ was set to 200, 400 or 500 ms (see text). The INEPT delays were $\tau = 2.72$ ms $\approx |4J_{HN}|^{-1}$ ($J_{HN} \approx -92$ Hz). The adjustable δ' delay was set to 502 μs. The encoding and decoding gradients ($G_{encode} = G_{decode}$) had durations $\delta = 1$ ms and calibrated amplitudes of 4.43, 8.36, 12.3, 16.2, 20.1, 24.1, 28.0, 31.9, 35.9, 39.8 G/cm. Durations and peak amplitudes over the *x*, *y* and *z* directions of the other gradients were respectively: $G_1 = 0.6$ ms, 42.5 G/cm, 0, 8.34 G/cm; $G_2 = 0.6$ ms, 0, 0, 41.7 G/cm; $G_3 = 0.6$ ms, 27.5 G/cm, 0, 0; $G_4 = 0.6$ ms, 35 G/cm, 37.8 G/cm, 17.4 G/cm; $G_5 = 0.6$ ms, 0, 29.7 G/cm, 0; G_6, 1 ms, 30 G/cm, 32.4 G/cm, 0. In the SOFAST-XSTE experiment the duration of G_1 was 1.33 ms

decoupling (CPD) is applied to the nitrogen-15 channel during acquisition. In order to control potential heating effects, the 2D version of the X-STE experiment was run with and without CPD during detection with a recovery delay of 2 s as in the XSTE-3D experiments. The diffusion coefficients measured with and without CPD during acquisition were identical (data not shown). A critical modification is the removal of a constant-time delay for nitrogen-15 chemical shift labelling. The total diffusion delay Δ has to be kept constant whilst the chemical shift labeling delay t_1 is incremented. Indeed, if the diffusion delay Δ were to increase with t_1, this would result in line broadening of peaks in the indirect

dimension. This line broadening, which is proportional to the square of the intensity of encoding gradients, would contribute to a significant overestimation of diffusion coefficients if, for example, $t_{1max} \approx 0.25 \Delta$. In this work, we have removed a constant-time delay for the evolution of the nitrogen-15 chemical shifts. The increase in t_1 is compensated by a decrease in the longitudinal diffusion delay T_L.

One should be aware that improper suppression of N_x and N_y components at the beginning of Δ may result in significant artefacts if $T_2(^{15}N)$ is comparable to Δ. Therefore, particular attention should be given to the choice of the amplitude and direction of the purge gradient G_4.
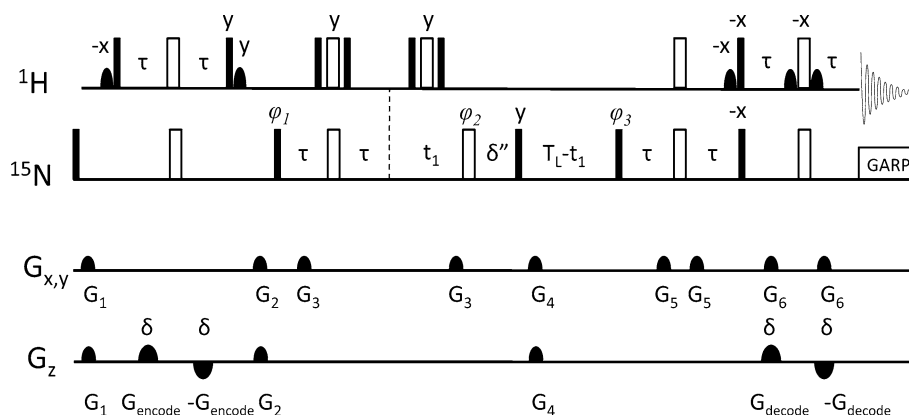
Fig. 2 HSQC-edited XSTE (XSTE-3D) sequence designed to measure diffusion coefficients of macromolecules at the residue level. *Filled* and *open* symbols indicate 90° and 180° pulses, respectively. *Bell-shaped* pulses represent water flip-back *Sinc* pulses with durations of 1.2 ms. Unless otherwise mentioned, all pulses are applied along the *x*-axis of the rotating frame. The phase cycling was:

$\varphi_1 = x,\ -x$; $\varphi_2 = x,\ x,\ x,\ x,\ y,\ y,\ y,\ y$; $\varphi_3 = y,\ y,\ -y,\ -y$ and the receiver phase was $\varphi_{rec} = x,\ -x,\ -x,\ x,\ -x,\ x,\ x,\ -x$. The total diffusion delay, the INEPT delays as well as the gradient amplitudes and durations were identical to those used in Fig. 1 for the BEST-XSTE sequence. The short delay δ'' leads to an effective value $t_1 = 0$ ms for the first time point in the indirect dimension

NMR data were processed and analyzed using NMRPipe (Delaglio et al. 1995). 1D spectra of Engrailed 2 and ubiquitin were integrated over the entire amide region, i.e., $7.5 < \delta < 9$ ppm for Engrailed 2, and $5.5 < \delta < 9.5$ ppm for ubiquitin. The backbone resonance assignments of Engrailed 2 are available at the BMRB (accession code 17325, Augustyniak et al. 2011).

Results and discussion

The sensitivities of four 2D heteronuclear stimulated echo experiments were compared. A series of 1D spectra was recorded on a sample of uniformly nitrogen-15 labelled human ubiquitin with the same total duration. Here, we define the sensitivity as the ratio of signal integrals over the amide region to the square root of the number of scans. Figure 3a and b allow one to compare the sensitivity of the BEST and SOFAST schemes with the original XSTE sequence for total diffusion delays $\Delta = 200$ and 500 ms. In both cases, the BEST-XSTE scheme is the most sensitive, with an improvement of the sensitivity up to 75 and 80 % for total diffusion delays $\Delta = 200$ and 500 ms respectively, for very short recovery delays T_{RD}, which are defined to include the acquisition delay. In most cases, the SOFAST-XSTE scheme is less sensitive than the BEST-XSTE sequence, except for $T_{RD} \leq 800$ ms. For a long total diffusion delay Δ and a very short recovery delay T_{RD} (Fig. 3b), the SOFAST-XSTE experiment is the most sensitive. SOFAST-XSTE thus appears to be adapted for large macromolecules or supramolecular assemblies that require long diffusion delays (Ferrage et al. 2003, 2004). The sensitivity of the XSTE experiment was also compared

to a variant where water flip-back elements were removed. The increase of sensitivity due to water flip-back pulses (Grzesiek and Bax 1993) is only of the order of 10%.

A potential limit of BEST-XSTE and SOFAST-XSTE experiments may be reached for very long total diffusion delays Δ. Indeed, the improvement in sensitivity is smaller for a total diffusion delay $\Delta = 500$ ms and is expected to decrease further with increasing Δ. The duration of typical double- and triple-resonance pulse sequences is on the order of 100 ms, i.e., an order of magnitude shorter than a typical recovery delay T_{RD} which can be as long as 1 s. The duration of XSTE experiments is also on the order of 1 s. Reducing the recovery delay T_{RD} from 2 s to 300 ms in XSTE experiments allows one to increase the number of scans in the same total duration from 24 to 88 (a factor 3.6) when $\Delta = 200$ ms but only from 24 to 64 (a factor 2.7) when $\Delta = 500$ ms.

We investigated the sensitivity of the modified XSTE-3D sequence by comparing intensities of the individual signals in ^1H–^{15}N planes with those obtained using a constant-time 3D XSTE sequence (CT-XSTE-3D) similar to the sequence of Choy et al. (2002). Both sequences were run on a 0.6 mM sample of a 114-residue construct of the chicken protein Engrailed 2 with identical parameters (recycle delay T_{RD}, total diffusion delay Δ, t_1^{max} and t_2^{max}) and processed in the same manner. The intensities were compared for the first 2D ^1H–^{15}N planes obtained with weak encoding and decoding gradients. Figure 4 shows the intensity ratios for many residues of Engrailed 2. Signal enhancements range from 18% to 178%. Enhancements of 100% or more are observed in the globular homeodomain region (residues 200–259) and in the hydrophobic cluster around the aromatic residues (Trp169, Trp172 and Tyr174)
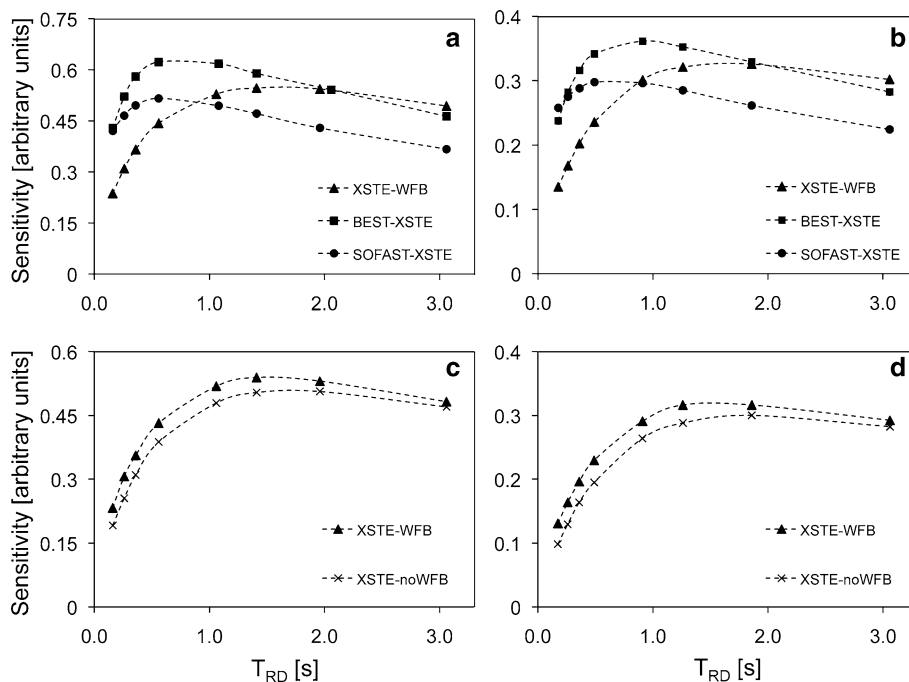
Fig. 3 Signal integral over the amide region divided by the square root of the number of scans as a function of the recovery delay T_{RD} between consecutive scans, obtained with different 2D NMR pulse schemes for a sample of 1.5 mM uniformly ^{15}N labeled human ubiquitin in 50 mM ammonium acetate, pH 4.5, at 600 MHz. The data have been extracted from 1D spectra recorded with low encoding/decoding gradient amplitudes ($G_{encode} = G_{decode} = s \, G^{max} = 4.43$ G/cm; where s is the shape factor and G^{max} the peak amplitude of the PFGs) using XSTE-WFB (with water flip-back) (*triangles*), BEST-XSTE (*squares*), SOFAST-XSTE (*circles*) and XSTE without water flip back (*crosses*) with integration over the amide region ($7 < \delta^H < 9.5$ ppm). The comparison has been performed for a total diffusion delay $\Delta = T_L + 6\tau = 200$ ms (**a** and **c**) and 500 ms (**b** and **d**). The *dashed lines* show interpolations of experimental data. The total duration of each experiment was 500 s (550 s) for $\Delta = 200$ ms (500 ms), respectively

of the non-globular N-terminal extension. Disordered regions are characterized by a smaller, yet significant, increase in sensitivity (on average by 50%). Overall, signal enhancements reflect local order that correlates with $R_2(^{15}N)$ (Klein-Seetharaman et al. 2002). Indeed, taking into account the effects of relaxation during t_1 and T_L, the ratio of signal intensities between schemes with and without constant-time evolution periods is:

$$S_{noCT}/S_{CT} = \exp\{[R_2(^{15}N) - R_1(^{15}N)](t_1^{max} - t_1)\} \quad (1)$$

where $R_2(^{15}N)$ and $R_1(^{15}N)$ are the transverse and longitudinal relaxation rates. This ratio is equal to 1 for $t_1 = t_1^{max}$, but it is larger for $t_1 < t_1^{max}$. Residues in the homeodomain that show no sign of chemical exchange have typical backbone transverse relaxation rates $R_2(^{15}N) \cong 10$ s^{-1} and longitudinal relaxation rates $R_1(^{15}N) \cong 2$ s^{-1}; and we used $t_1^{max} = 110$ ms. The intensity ratio of (1) is expected to be 2.4 for $t_1 = 0$. Compared to this estimate, some of the ratios displayed in Fig. 4 appear too large. Relaxation of antiphase coherences during part of t_1 as well as small contributions of chemical exchange to $R_2(^{15}N)$ for some residues combined with the window function employed (a sine-bell window function shifted by 90°) in the t_1 dimension may explain this higher-than-expected signal enhancement.

The resolution in the ω_1 dimension of a constant-time experiment is expected to be higher than in a non-constant-time experiment. In our case, resolution in the indirect dimension is sufficient in both experiments. All peaks that
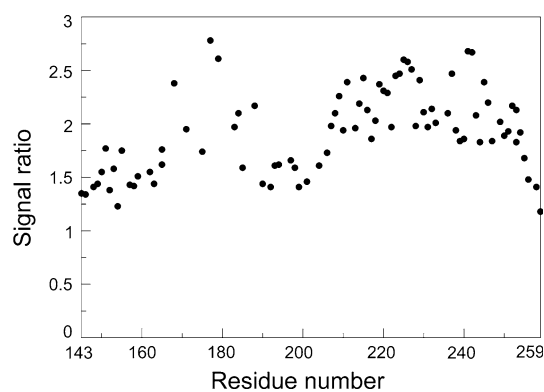

Fig. 4 Ratios of signal intensities obtained by the new XSTE-3D method to those obtained by the conventional constant-time XSTE method (Choy et al. 2002) for individual residues in uniformly ^{15}N labeled Engrailed 2 (0.6 mM). Data were recorded at 600 MHz and obtained by extracting the 2D ^1H–^{15}N planes measured with low diffusion gradient amplitudes ($G_{encode} = G_{decode} = 4.43$ G/cm) with a total diffusion delay $\Delta = 400$ ms and a recovery delay $T_{RD} = 2$ s. The total duration of each experiment was 13 h and 50 min

were sufficiently resolved for quantitative analysis in the ^1H–^{15}N planes of the constant-time experiment were also resolved when using the sequence of Fig. 2.

So far, our attempts to combine 3D and BEST approaches were unsuccessful. Sample heating by CPD during acquisition can be significant when short recovery delays are used. An in-phase/antiphase (IPAP) version of the experiment was therefore developed, in analogy to earlier work (Choy et al. 2002). Even after addition of the two resulting signals of the two proton doublet components the signal-to-noise ratios were lower than those measured with the XSTE-3D sequence. Further developments will be necessary to overcome this limit.

XSTE sequences were applied to monitor the stability of a 114-residue construct of Engrailed 2, which is partially disordered. This construct consists of a globular homeodomain (residues 200–259) and a long disordered N-terminal extension (residues 146–199). The disordered N-terminal extension is particularly sensitive to the formation of intermolecular disulfide bonds and to degradation by proteases. XSTE experiments were employed to optimize the buffer conditions to achieve sample stability over several weeks.

The BEST-XSTE sequence was used to monitor dimerization of Engrailed 2 through the oxidative formation of intermolecular disulfide bonds. Figure 5 shows the decay of the signal integrated over the amide region as a function of the amplitude of the pulsed field gradients. A freshly prepared protein sample without reducing agent was first used. The diffusion coefficient was found to be D = $(4.022 \pm 0.012) \times 10^{-11}$ m^2 s^{-1}. The buffer was then modified by the addition of 10 mM TCEP. The diffusion coefficient of Engrailed 2 increased to D = $(7.788 \pm 0.015) \times 10^{-11}$ m^2 s^{-1}. Diffusion coefficients were obtained by fitting the decay of the signal integral to the equation:

$$S/S_0 = \exp\left(-4D\gamma^2 s^2 G_{max}^2 \delta^2 \Delta\right) \qquad (2)$$

where S and S_0 are the signal integrals with and without encoding and decoding gradients; D is the diffusion coefficient; γ the gyromagnetic ratio of the proton, $s \leq 1$ the shape factor of the encoding and decoding gradient pulses; G_{max} the peak amplitude of the gradients; δ their duration; $\Delta = T_L + 6\tau$ is the total diffusion delay, and T_L is the longitudinal diffusion delay.

These results suggest that, in the absence of any reducing agent, Engrailed 2 is not monomeric, since the diffusion coefficient is too small for a protein comprising only 114 residues. The increase of the diffusion coefficient upon addition of the reducing agent TCEP indicates that the oligomerization occurs through the formation of an intermolecular disulfide bond. Our 114-residue construct of
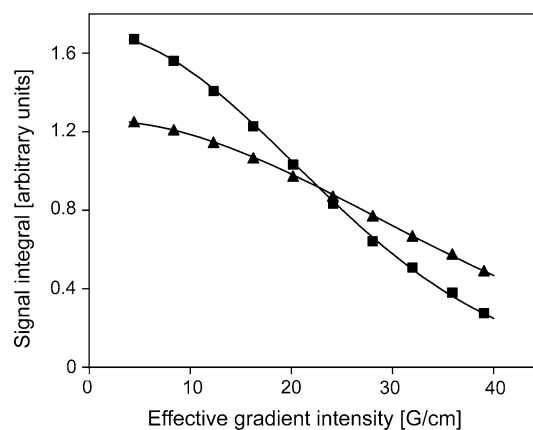

Fig. 5 Signal integrals of reduced and oxidized Engrailed 2 (∼0.6 mM, pH 6, 30°C) plotted as a function of the amplitude $G_{encode} = G_{decode}$ of the encoding and decoding gradients. (*Squares*) protein with 10 mM TCEP reducing agent to prevent dimerization. (*Triangles*) protein without TCEP so that dimerization through the oxidative formation of an intermolecular disulfide bridge could occur. The spectra were recorded using the BEST-XSTE sequence and integrated over the amide region ($7 < \delta^H < 9.5$ ppm). The total diffusion delay was $\Delta = 400$ ms. The diffusion coefficients determined by fitting to (2), are D = $(7.788 \pm 0.015) \times 10^{-11}$ m^{2}s^{-1} and D = $(4.022 \pm 0.012) \times 10^{-11}$ m^{2}s^{-1} for the monomeric and dimeric forms of the protein. The duration of each experiment was 520 s

Engrailed 2 features a single cysteine residue (Cys175). The multimeric state under oxidative conditions is most likely a dimer with an intermolecular disulfide bond. Although little is known about the structure of Engrailed since only the homeodomain was studied at atomic resolution (Fraenkel et al. 1998), there is no evidence that homeoproteins can occur in living organisms as oxidized dimers. The reduced monomer is therefore likely to be the native form of Engrailed 2.

The increase in the diffusion coefficient of Engrailed upon reduction is much larger than what would be expected for a transition between dimeric and monomeric forms of a globular protein (Augé et al. 2009). Although diffusion coefficients of disordered proteins can be reproduced in silico with a good accuracy (Bernado and Blackledge 2009), diffusion coefficients of partially disordered proteins are not yet amenable to reliable predictions. In particular, the hydrodynamic coupling between the folded and disordered regions is difficult to evaluate accurately. In the case of a dimer, with two partially disordered proteins cross-linked in the middle of their disordered regions, predictions would be particularly challenging. The large variation of diffusion coefficients could be ascribed to an oligomerization equilibrium when the protein is in the oxidized state.

The proteolysis of Engrailed 2 was followed by the measurement of diffusion coefficients at the residue level with the XSTE-3D pulse sequence. For this purpose the protein was stored at 303 K for 7 days in the absence of

any protease inhibitor. Figure 6 shows residue-by-residue estimates of diffusion coefficients for Engrailed obtained at 303 K. Attenuation of signal intensities in ^1H–^{15}N planes upon increasing the amplitudes of the encoding and decoding gradient were fitted to (2). Two different values of the diffusion coefficient can be identified. For residues 146–177, the diffusion coefficients vary between $D = (12.575 \pm 0.152) \times 10^{-11}$ m^2 s^{-1} and $D = (14.615 \pm 0.134) \times 10^{-11}$ m^2 s^{-1}, with an average value $D_{146–177} = (13.465 \pm 1.150) \times 10^{-11}$ m^2 s^{-1}. The site-to-site variations of the diffusion coefficient within this fragment are significantly larger than the uncertainty of the measurement. This feature may be due to the presence of additional proteolysis sites, although only minor peaks corresponding to fragments with lower masses have been observed on MS-MALDI spectrum. The superposition of signals from proteolysis products of various lengths would result in significant variations of apparent diffusion coefficients. Multi-exponential fitting would, in principle, allow for the determination of the diffusion coefficients of all species, but is difficult when their variations are small. However, an overestimation of the precision of our measurements cannot be excluded. For residues 179–259, the diffusion coefficients vary from $D = (8.023 \pm 0.159) \times 10^{-11}$ m^2 s^{-1} to $D = (9.581 \pm 0.263) \times 10^{-11}$ m^2 s^{-1}, with an average value $D_{179–259} = (8.765 \pm 0.816) \times 10^{-11}$ m^2 s^{-1}. These data indicate unambiguously that the protein was cut into at least two fragments, which are diffusing as independent molecules. The data corresponding to residues close to C175 are either incomplete or associated with high errors, due to low signal-to-noise ratios. The observation of significantly different diffusion coefficients for residues 177 and 179 makes it possible to determine the cleavage site precisely. The hydrolysis must have occurred between R177 and Y178, which suggests that some trypsin-like endopeptidases, which cleave peptide chains mainly on the carboxyl side of lysine or arginine (Brown and Wold 1973), may have contaminated the sample. The mass spectrum showed two major peaks at 3.77 and 9.87 kDa. The expected molecular masses of the 146–177 and 178–259 fragments of uniformly ^{15}N labeled Engrailed 2 protein are 3.77 and 9.88 kDa. These data are in remarkable agreement with results obtained by PFG-NMR.

Conclusions

Heteronuclear stimulated echoes are suitable to measure accurate and precise diffusion coefficients of large macromolecules. We have introduced several improvements to enhance the sensitivity. Shorter recovery delays T_{RD} allow one to reduce the total experimental time. Our new 3D scheme provides access to residue-by-residue estimates of diffusion coefficients with better sensitivity. The improved
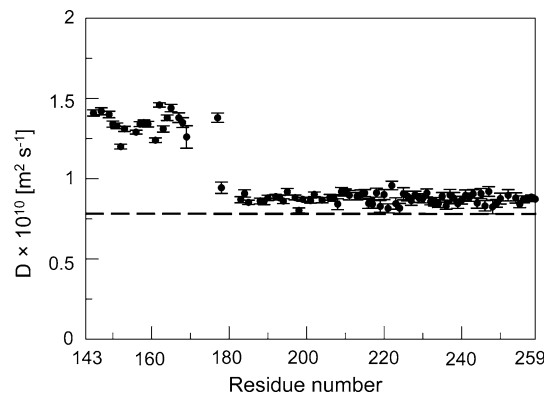

Fig. 6 Diffusion coefficients of individual amino-acid residues in Engrailed 2 incubated at 303 K for 7 days after purification without protease inhibitors. The protein was cleaved between R177 and Y178 to form two fragments. The intensities of individual H^N signals were measured with the XSTE-3D sequence. The total diffusion delay was $\Delta = 400$ ms. The *dashed line* indicates the diffusion coefficient of the uncleaved form of Engrailed 2 under the same experimental conditions determined with the BEST-XSTE sequence

heteronuclear stimulated echo methods allow one to study a wide range of molecular properties, as illustrated by the investigation of the cleavage and dimerization of the partially disordered protein Engrailed 2.

Supporting information

The pulse sequence of the constant-time XSTE-3D experiment and the MALDI-TOF spectrum of the Engrailed spectrum after proteolysis are provided in the supporting information.

References

Ahuja P, Sarkar R, Vasos PR, Bodenhausen G (2009) Diffusion coefficients of biomolecules using long-lived spin states. J Am Chem Soc 131:7498–7499. doi:10.1021/ja902030k

Altieri AS, Hinton DP, Byrd RA (1995) Association of biomolecular systems via pulsed field gradient NMR self-diffusion measurements. J Am Chem Soc 117:7566–7567. doi:10.1021/ja00133a03

Augé S, Schmit PO, Crutchfield CA, Islam MT, Harris DJ, Durand E, Clemancey M, Quoineaud AA, Lancelin JM, Prigent Y, Taulelle F, Delsuc MA (2009) NMR measure of translational diffusion and fractal dimension. Application to molecular mass measurement. J Phys Chem B 113:1914–1918. doi:10.1021/jp8094424

Augustyniak R, Balayssac S, Ferrage F, Bodenhausen G, Lequin O (2011) ^1H, ^{13}C and ^{15}N resonance assignment of a 114-residue fragment of Engrailed 2 homeoprotein, a partially disordered protein. Biomol NMR Assign. doi: 10.1007/s12104-011-9306-5

Bernado P, Blackledge M (2009) A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering. Biophys J 97:2839–2845. doi:10.1016/j.bpj.2009.08.044

Brand T, Cabrita E, Morris GA, Gunther R, Hofmann H, Berger S (2007) Residue-specific NH exchange rates studied by NMR diffusion experiments. J Magn Reson 187:97–104. doi:10.1016/j.jmr.2007.03.021

Brown WE, Wold F (1973) Alkyl isocyanates as active-site-specific reagents for serine proteases. Reaction properties. Biochemistry 12:828–834. doi:10.1021/bi00729a007

Buevich AV, Baum J (2002) Residue-specific real-time NMR diffusion experiments define the association states of proteins during folding. J Am Chem Soc 124:7156–7162. doi:10.1021/ja012699u

Cantor CR, Schimmel PR (1980) Biophysical chemistry, part II: techniques for the study of Biological Structure and Function. W.H. Freeman, San Francisco

Carravetta M, Johannessen OG, Levitt MH (2004) Beyond the T1 limit: singlet nuclear spin states in low magnetic fields. Phys Rev Lett 92:153003–153004. doi:10.1103/PhysRevLett.92.153003

Cavadini S, Vasos P (2008) Singlet states open the way to longer time-scales in the measurement of diffusion by NMR spectroscopy. Concepts Magn Reson 32A:68–78. doi:10.1002/cmr.a.20100

Cavadini S, Dittmer J, Antonijevic S, Bodenhausen G (2005) Slow diffusion by singlet state NMR spectroscopy. J Am Chem Soc 127:15744–15748. doi:10.1021/ja052897b

Chang PT (2000) Chapter 21.4 diffusion. In: Physical chemistry for the chemical and biological sciences, 2nd edn, University Science Books, Sausalito, CA, pp 876–883

Choy WY, Mulder FAA, Crowhurst KA, Muhandiram DR, Millett IS, Doniach S, Forman-Kay JD, Kay LE (2002) Distribution of molecular size within an unfolded state ensemble using small-angle X-ray scattering and pulse field gradient NMR techniques. J Mol Biol 316:101–112. doi:10.1006/jmbi.2001.5328

Connell MA, Bowyer PJ, Bone PA, Davis AL, Swanson AG, Nilsson M, Morris GA (2009) Improving the accuracy of pulsed field gradient NMR diffusion experiments: Correction for gradient non-uniformity. J Magn Reson 198:121–131. doi:10.1016/j.jmr.2009.01.025

Danielsson J, Jarvet J, Damberg P, Graslund A (2002) Translational diffusion measured by PFG-NMR on full length and fragments of the Alzheimer Aβ(1–40) peptide. Determination of hydrodynamic radii of random coil peptides of varying length. Magn Reson Chem 40:S89–S97. doi:10.1002/mrc.1132

Danielsson J, Liljedahl L, Barany-Wallje E, Sønderby P, Kristensen LH, Martinez-Yamout MA, Dyson HJ, Wright PE, Poulsen FM, Maler L, Graslund A, Kragelund BB (2008) The intrinsically disordered RNR inhibitor Sml1 is a dynamic dimer. Biochemistry 47:13428–13437. doi:10.1021/bi801040b

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6:277–293. doi:10.1007/BF00197809

Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 6:197–208. doi:10.1038/nrm1589

Ferrage F, Zoonens M, Warschawski DE, Popot J-L, Bodenhausen G (2003) Slow diffusion of macromolecular assemblies by a new pulsed field gradient NMR method. J Am Chem Soc 125:2541–2545. doi:10.1021/ja0211407

Ferrage F, Zoonens M, Warschawski DE, Popot J-L, Bodenhausen G (2004) Slow diffusion of macromolecular assemblies measured by a new pulsed field gradient NMR method. J Am Chem Soc 126:5654. doi:10.1021/ja033464g

Fraenkel E, Rould MA, Chambers KA, Pabo CO (1998) Engrailed homeodomain-DNA complex at 2.2 Å resolution: a detailed view of the interface and comparison with other engrailed structures. J Mol Biol 284:351–361. doi:10.1006/jmbi.1998.2147

Geen H, Freeman R (1991) Band-selective radiofrequency pulses. J Magn Reson 93:93–141. doi:10.1016/0022-2364(91)90034-Q

Grzesiek S, Bax A (1993) The importance of not saturating water in protein NMR. Application to sensitivity enhancement and NOE measurements. J Am Chem Soc 115:12593–12594. doi:10.1021/ja00079a052

Hahn EL (1950) Spin echoes. Phys Rev 80:580–594. doi:10.1103/PhysRev.80.580

Hsu S-TD, Fucini P, Cabrita LD, Launay H, Dobson CM, Christodoulou J (2007) Structure and dynamics of a ribosome-bound nascent chain by NMR spectroscopy. Proc Natl Acad Sci USA 104:16516–16521. doi:10.1073/pnas.0704664104

Jerschow A (2000) Thermal convection currents in NMR: flow profiles and implications for coherence pathway selection. J Magn Reson 145:125–131. doi:10.1006/jmre.2000.2083

Jerschow A, Müller N (1997) Suppression of convection artifacts in stimulated-echo diffusion experiments. Double-stimulated-echo experiments. J Magn Reson 125:372–375. doi:10.1006/jmre.1997.1123

Jerschow A, Müller N (1998) Diffusion-separated nuclear magnetic resonance spectroscopy of polymer mixtures. Macromolecules 31:6573–6578. doi:10.1021/ma9801772

Johnson CS (1999) Diffusion ordered nuclear magnetic resonance spectroscopy: principles and applications. Prog NMR Spectrosc 34:203–256. doi:10.1016/j.ab.2007.04.025

Klein-Seetharaman J, Oikawa M, Wirmer J, Duchardt E, Ueda T, Imoto T, Smith LJ, Dobson C, Schwalbe H (2002) Long-range interactions within a non-native protein. Science 295:1719–1722. doi:10.1126/science.1067680

Kupce E, Freeman R (1994) Wideband excitation with polychromatic pulses. J Magn Reson A 108:268–273. doi:10.1006/jmra.1994.1123

Li Y, Kim S, Brodsky B, Baum J (2005) Identification of partially disordered peptide intermediates through residue-specific NMR diffusion measurements. J Am Chem Soc 127:10490–10491. doi:10.1021/ja052801d

Li C, Wang Y, Pielak GJ (2009) Translational and rotational diffusion of a small globular protein under crowded conditions. J Phys Chem B 113:13390–13392. doi:10.1021/jp907744m

Martinez-Viviente E, Pregosin PS (2003) Low temperature H-1-, F-19-, and P-31-PGSE diffusion measurements. Applications to cationic alcohol complexes. Helv Chim Acta 86:2364–2378

McLachlan A, Richards J, Bilia A, Morris GA (2009) Constant time gradient HSQC-iDOSY: practical aspects. Magn Reson Chem 47:1081–1085. doi:10.1002/mrc.2518

Orekhov VY, Korzhnev DM, Pervushin KV, Hoffman E, Arseniev AS (1999) Sampling of protein dynamics in nanosecond time scale by 15 N NMR relaxation and self-diffusion measurements. J Biomol Struct Dynam 17:157–174

Pervushin K, Vögeli B, Eletsky A (2002) Longitudinal 1H relaxation optimization in TROSY NMR spectroscopy. J Am Chem Soc 124:12898–12902. doi:10.1021/ja027149q

Piotto M, Saudek V, Sklenar V (1992) Gradient-tailored excitation for single-quantum NMR spectoscopy of aqueous solutions. J Biomol NMR 2:661–665. doi:10.1007/bf02192855

Price WS (1998) Pulsed-field gradient nuclear magnetic resonance as tool for studying translational diffusion: part II. Experimental aspects. Concepts Magn Reson 10:197–237. doi:10.1002/(SICI)1099-0534(1998)10:4<197::AID-CMR1>3.0.CO;2-S

Sarkar R, Ahuja P, Vasos PR, Bodenhausen G (2008a) Measurement of slow diffusion coefficients of molecules with arbitrary scalar

couplings via long-lived spin states. ChemPhysChem 9:2414–2419. doi:10.1002/cphc.200800476

Sarkar R, Moskau D, Ferrage F, Vasos PR, Bodenhausen G (2008b) Single or triple gradients? J Magn Reson 193:110–118. doi:10.1016/j.jmr.2008.04.029

Schanda P, Brutscher B (2005) Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. J Am Chem Soc 127:8014–8015. doi:10.1021/ja051306e

Schanda P, Van Melckebeke H, Brutscher B (2006) Speeding up three-dimensional protein NMR experiments to a few minutes. J Am Chem Soc 128:9042–9043. doi:10.1021/ja062025p

Stejskal EO, Tanner JE (1965) Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. J Chem Phys 42:288–292. doi:10.1063/1.1695690