Rapid Characterization of Structural and Behavioral Changes of Therapeutic Proteins by Relaxation and Diffusion ^{1}H-SOFAST NMR Experiments

Xingjian Xu[1,*], Guilherme Dal Poggetto[1,*], Mark McCoy[2], Mikhail Reibarkh[1], and Pablo Trigo-Mourino[3,*]

1. Analytical Research & Development, MRL, Merck & Co., Inc., Rahway, NJ, USA 2. Quantitative Biosciences, MRL, Merck & Co., Inc., Rahway, NJ, USA 3. Analytical Research & Development, MRL, Merck & Co., Inc., Boston, MA, USA

ABSTRACT: Biologic drugs have emerged as a rapidly expanding and important modality, offering promising therapeutic solutions by interacting with previously 'undruggable' targets, thus significantly expanding the range of modern pharmaceutical applications. However, the inherent complexity of these drugs also introduces liabilities and poses challenges in their development, necessitates efficient screening methods to evaluate the structural stability and behavior. Although NMR spectroscopy is well-suited for detecting weak interactions, changes in dynamics, high-order structure, and association states of macromolecules in fully formulated samples, the inherent low sensitivity limits its utility as a fast screening and characterization tool. In this study, we present two fast pulsing NMR experiments, namely the SOFAIR (band-Selective Optimized Flip-Angle Internally-encoded Relaxation) and the SOFIT (band-Selective Optimized Flip-angle Internally-encoded Translational diffusion), which enable rapid and reliable measurements of transverse relaxation rates and diffusion coefficients with more than 10-fold higher sensitivity compared to commonly used methods, like CPMG (Carr-Purcell-Meiboom-Gill) and DOSY (Diffusion-Ordered Spectroscopy), allowing the rapid assessment of biologics even at low concentrations. We demonstrated the effectiveness and versatility of these experiments by evaluating several examples, including thermally stressed proteins, proteins at different concentrations, and a therapeutic protein in various formulations. We anticipate that these novel approaches will greatly facilitate the analysis and characterization of biologics during drug discovery.

INTRODUCTION

The discovery and development of biologics have undergone remarkable progress in recent decades, making substantial contributions to the treatment of diverse diseases and conditions[1-3]. However, it is important to recognize the costly and challenging nature of progressing a biologic from discovery to the market, which is often accompanied by high attrition rates[4, 5]. Consequently, significant efforts have been made in the industry to enhance success rates and reduce overall discovery costs by developing methods that can detect liabilities early in the discovery process[6].

One crucial aspect that has garnered increased attention in biologics discovery and development is the characterization of their biophysical properties[6, 7]. This entails assessing key attributes of the target molecules such as stability, solubility, aggregation propensity, formulation suitability, and resistance to various stressors like UV light, temperature, humidity, agitation, and freeze/thaw cycles[8, 9]. Such evaluations are essential in helping researchers identify and eliminate candidates with poor developability at an early stage[10].

An integral component of the biophysical characterization of biologics involves understanding the structural and dynamic behavior of biologics in solution, which are often closely tied to their functions[11]. Some of the widely used biophysical tools include size-exclusion chromatography (SEC), static light scattering/dynamic light scattering (SLS/DLS), UV, fluorescence, circular dichroism (CD), and Fourier transform infrared spectroscopy (FTIR), which can be straightforwardly implemented, but lack in resolution[12]. In this regard, Nuclear Magnetic Resonance (NMR) spectroscopy offers a unique advantage, as it possesses sensitivity to even subtle changes in protein behavior, sparsely populated species, and weak interactions[13-16]. Moreover, NMR also has the unique capability to detect proteins, excipients, and their interactions *in situ*[17-20], eliminating the need for buffer dilution, exchange, or the addition of organic modifiers.

Typically, obtaining atomic resolution details of proteins in NMR experiments entails the use of multidimensional methods with isotopically labeled samples[21, 22]. Two-dimensional (2D) experiments may be acquired at natural abundance, but doing so can be time-consuming and demand a substantial sample. As a practical alternative, one-dimensional (1D) NMR-based approaches can offer

valuable insight and provide a wealth of useful information through the analysis of signal dispersion, line shape, or chemical shift changes[23]. More recently, 1D NMR spectra have proven valuable in molecular fingerprinting and profiling of biologics[24]. Furthermore, some advanced NMR methods can also be employed in their 1D forms to investigate the dynamic and translational properties of proteins[17, 25]. This can be achieved by monitoring changes in the signal intensity of the proton spectra as specific parameters are incrementally varied. The Carr-Purcell-Meiboom-Gill (CPMG) and Diffusion-Ordered Spectroscopy (DOSY) methods, for instance, have enabled the measurements of proton transverse relaxation rate (R_2) and translational diffusion coefficient (D), which are two parameters widely applied for characterizing proteins' behavioral changes under diverse conditions[25-27]. The correlation between these two parameters has been proposed as a means to profile therapeutic protein motions and interactions[26]. While useful, the relatively long data acquisition required to obtain high signal-to-noise ratios in these NMR experiments can limit their practical utility.

The SOFAST (band-Selective Optimized Flip-Angle Short-Transient) technique is an NMR method originally developed for fast 2D data acquisition[28, 29], but its potential applications as a 1D experiment have also been explored in conjunction with other NMR building blocks commonly utilized in atomic-resolution methods. One notable example is the HET-SOFAST[30], which could be used to rapidly obtain information on protein folding and structural compactness. In this work, we present a pair of modified ^{1}H SOFAST NMR experiments, which we termed SOFAIR (band-Selective Optimized Flip-Angle Internally-encoded Relaxation) and SOFIT (band-Selective-Optimized-Flip-angle Internally-encoded Translational diffusion), to enable more time-efficient measurements of proton R_2 and D compared to CPMG and DOSY. These experiments were tested on various examples, including thermally stressed proteins, proteins at different concentrations, and a proprietary therapeutic protein in different formulations. Through our investigations, we demonstrated that these modified SOFAST experiments provide estimations of R_2 and D values that are sensitive to structural stability and molecular interactions. Notably, the improved sensitivity and short acquisition time make these methods highly suitable for screening purposes, particularly in the drug discovery stage, where sample quantity may be limited and multiple sequence variants are compared.

EXPERIMENTAL SECTION

Materials. Lyophilized Bovine Serum Albumin (BSA) powder (≥98%) was acquired from Sigma-Aldrich. BSA was dissolved in the desired buffer (10 or 20 mM sodium phosphate, pH 7.5). Biologic A was acquired internally.

The catalytic domain of diphtheria toxin (DTA) was expressed and purified in-house. DTA tagged with an uncleavable His_6 tag on the C-terminal end was overexpressed in BL21(DE3) *E. coli*. Cells were grown in Luria-Bertani (LB) broth and induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at 37 °C after reaching an OD600 of 0.6. Cells were incubated overnight at 20 °C (16 hrs) and harvested the next day. The purification of the protein followed closely to what was described by Sauvé *et al*[31]. The protein was exchanged into a final 10 mM sodium phosphate buffer, pH 7.0 at a concentration of 14 mg/mL.

BSA Thermal Stress. A BSA sample at 20 mg/mL concentration was prepared. To increase the susceptibility of BSA to thermal stress, the BSA sample was mixed with a denaturant, 1M guanidinium hydrochloride (GdnHCl). The resulting mixture was then divided equally into two halves. One-half of the GdnHCl-treated BSA sample was subjected to thermal stress by incubating it at 57 °C for 60 minutes. Following the denaturation process, this sample was stored at 4°C. The other half of the sample remained at 4°C throughout this process without undergoing thermal denaturation. In order to create BSA samples with varying levels of stressed protein for subsequent NMR experiments, the two samples were combined in different proportions, including 0%, 25%, 50%, 75%, and 100% of the thermally stressed sample. These samples were mixed right before data acquisition.

NMR Experiments. 8 – 10% of D_2O was added to each sample. Generally, relaxation experiments were conducted with 5 mm NMR tubes and 600 µl total sample volume, diffusion experiments were conducted in 3 mm NMR tubes and 180 µl sample volume to minimize convection. All experiments on biologic A were performed in 3 mm NMR tubes.

NMR data for BSA and Biologic A were collected on a Bruker Avance Neo 500 MHz spectrometer equipped with a 5mm Prodigy TCI probe. NMR data for DTA were acquired on a Bruker Avance III HD 600 MHz spectrometer equipped with a 5 mm Prodigy BBO probe. All experiments were collected at 298K unless specified otherwise. Relaxation experiments were processed and analyzed using MestReNova (Version 14.2.1). Diffusion experiments were processed and analyzed in Topspin (Version 4.0.6), MestReNova (Version 14.2.1), and Python (Version 3.8.5).

Diffusion-Filtered CPMG. The diffusion-filtered CPMG pulse sequence (in-house coding) used to measure protein proton R_2 values contained a standard CPMG block followed by a pulse field gradient stimulated echo block to suppress any signals from small molecules, buffer, and water[26]. Specifically, a bipolar gradient of 1500 μs at 100% amplitude and a diffusion of 100 ms was used. For the CPMG echo train, the relaxation time was varied up to 0.2 s with a fixed ν_{cpmg} of 500 Hz. The recycle delay time was set to 5 s. The total data acquisition time for each experiment was approximately 40 minutes.

SOFAIR Experiments. To observe amide protons or methyl protons, the selective excitation pulse was typically applied to somewhere between 9.0 – 9.5 ppm or 0 – 1.5 ppm, respectively, covering a bandwidth of 2.5 – 3.0 ppm. The specific selection depends on the presence of any overlapping small molecule or buffer signals. The relaxation encoding period (2τ), was varied from 0 s to up to 0.2 s. Unless otherwise specified, the recycle delay, d1 was set to 0.25 s. All experiments were collected with a total data acquisition time of approximately 10 minutes for each experiment.

DOSY Experiments. The Bruker ^{1}H 2D DOSY pulse sequence stebpgp1s19 was used as the conventional DOSY experiment to measure *D*. A total of 300 ms diffusion encoding period (Δ) was used, along with 1.5 ms bipolar SMSQ10.100 gradient (δ/2) shape pulses varied quadratically from 10 to 95%. Each experiment was acquired with a total of 16 gradient increments for a total data acquisition time of approximately 20 minutes. For the catalytic domain of diphtheria toxin, the experiment was collected using 16 increments with a total experiment duration of ~5 hours to highlight the sensitivity differences between DOSY and the novel SOFIT experiment.

SOFIT Experiments. The excitation pulse was applied to somewhere between 9.0 – 9.5 ppm or 0 – 1.5 ppm to selectively observe amide or methyl protons, respectively, each covering a bandwidth of 2.5 – 4.0 ppm. A diffusion encoding period (Δ) of 20 ms and 6 ms SMSQ10.100 gradient shape pulses (δ) were used. The gradient was quadratically varied from 10% to 95% of the maximum gradient amplitude. A total of 16 gradient increments were acquired with a total data acquisition time of approximately 10 minutes.

Relaxation (R_2) and Diffusion (D) Measurements. Unless otherwise specified, R_2 and D measurements were performed by area integration over two regions: the methyl proton region (1.2 to -0.8 ppm, denoted 'Methyl') and the amide proton region (9.5 to 7.5 ppm, denoted 'H(N)'). For R_2, the resulting integrals were fit to Eq. 1:

$$I = I_\circ e^{-tR_2} + B \qquad (1)$$

where t is the delay time and R_2 is the transverse relaxation. For *D*, measurements were performed by area integration over the same methyl and amide proton regions. The resulting integrals were fit to Eq. 2:

$$I = I_\circ e^{-(\gamma\delta G)^2 D(\Delta - \frac{\delta}{3})} \qquad (2)$$

where $I_\circ$ is the signal integral at a gradient strength of zero, γ is the observed nuclei gyromagnetic ratio, G is the gradient strength, δ is the gradient pulse duration, Δ is the diffusion encoding period, and *D* is the diffusion coefficient (for a squared shape field gradient pulse). The new pulse sequences and corresponding analysis are detailed in the Supporting Information.

For biologic A characterizations, hydrodynamic radius (R_h) was reported instead of *D*. R_h was derived from the Einstein-Stoke equation:

$$D = \frac{k_B T}{6\pi\eta R_h} \qquad (3)$$

https://doi.org/10.26434/chemrxiv-2024-nj5cc **ORCID:** https://orcid.org/0000-0003-0822-882X

where k_B is Boltzmann constant, T is temperature, and η is viscosity, which was determined using a calibration curve from a previous study accounting for viscosity changes of additives[32].

RESULTS AND DISCUSSION

^{1}H SOFAIR and SOFIT Experiments. *D* and R_2 are typically obtained through the use of the DOSY and the CPMG experiments, respectively[26]. The DOSY experiment involves acquiring a series of 1D spectra applying a series of increasing gradient pulse strengths, leading to the distance-dependent dephasing of spins based on the translational diffusion of the species in solution[33]. As such, the diffusion coefficient of a molecule is tied to its hydrodynamic properties and interactions with other molecules and can be extracted by fitting the attenuation of the NMR signal as a function of the gradient strength. For measuring *D* of biologics, the standard Bruker library pulse sequence stebpgp1s19 is commonly used. This pulse program utilizes stimulated echo with bipolar gradient pulses for diffusion encoding and also 3-9-19 water suppression.

On the other hand, The CPMG experiment involves the repetition of refocusing pulses separated by equidistantly spaced dephasing delays. By increasing the number of repetitions, one can observe signal attenuation caused by transverse relaxation[34]. The R_2 relaxation rate can then be estimated from the decay curve. The relaxation behavior of macromolecules, such as proteins, is linked to their molecular motions, which in turn reflect size, conformational changes, unfolding, interactions, or aggregation[27]. In our experience, the inclusion of an additional diffusion filter in the pulse sequence is highly beneficial[26]. This filter effectively suppresses signals originating from excipients, buffer, and water, and is commonly employed for measuring R_2 of biologics using CPMG.

Ideally, R_2 and *D* should be measured by examining isolated peaks' relaxation and diffusion-induced decays. However, for biologics where peaks are broad and contain many overlapping signals, analyses often rely on integration over selected regions of the 1D ^{1}H spectrum. In these cases, particular emphasis is placed on the amide and methyl protons.

While the CPMG and DOSY experiments have proven valuable for characterizing protein behavior in solution, their relatively low sensitivity has limited their utility, primarily to study proteins with solution concentrations > 1-5 mg/ml. The incorporation of water and/or excipient suppression blocks in the pulse sequences further compounds this limitation, making them less suitable for screening purposes. To address these potential drawbacks, we have decided to take advantage of the highly sensitive SOFAST experiment[28, 29].

The 1D SOFAST experiment, illustrated in Fig. 1A, employs a band-selective PC9 excitation pulse with a 120° flip angle, followed by an RSNOB refocusing pulse. The distinctive characteristic of the SOFAST experiment is its ability to achieve high sensitivity through a very short recycle delay, thereby enabling an increased repetition rate. The band-selective nature of the SOFAST experiments also allows for the selective excitation of regions devoid of excipient or other undesired signals, eliminating the need for additional pulse sequence blocks for signal suppression. Both relaxation and diffusion elements were incorporated into the SOFAST pulse sequence through internal encoding (Fig. 1B/C), allowing R_2 and *D* to be measured without sensitivity loss and need for additional phase cycling[35-37].

For R_2 measurements, the SOFAST pulse sequence is modified to include delays (τ) flanking the refocusing pulse[34, 38]. This delay is incremented in pseudo-2D experiments to encode the transverse relaxation, which we named SOFAIR, as shown in Fig. 1B.

Unlike the CPMG experiment, this modified SOFAST approach, in certain cases, does not suppress *J*-modulation. When selectively exciting and refocusing amide protons, the *J*-evolution is refocused because the Hα atoms, to which the amide protons are coupled with, are outside the refocusing bandwidth. However, when analyzing methyl protons, the measured R_2 could be affected by *J*-modulation. While these effects exist, the influences, as examples will demonstrate, is limited only to the absolute rate values. These are less impactful for developability than the relative values and trends observed when comparing different sequences or the effects of different excipients.

For *D* measurements, instead of changing the delays, the gradient strength G_2 is varied to encode diffusion, which we named SOFIT, depicted in Fig. 1C. By incrementing the gradient strength G_2, the diffusion coefficient is encoded through a single echo scheme, differ from the stimulated echo employed for typical DOSY experiments. As a result, during the diffusion period, the sample is subject to effects from transverse relaxation. Therefore, the diffusion encoding period (Δ) was kept at a maximum of 20 ms to mitigate the impact of relaxation. In contrast, the gradient pulse (δ) duration was set to a very high 6 ms to maximize the efficiency of the diffusion-encoding process.

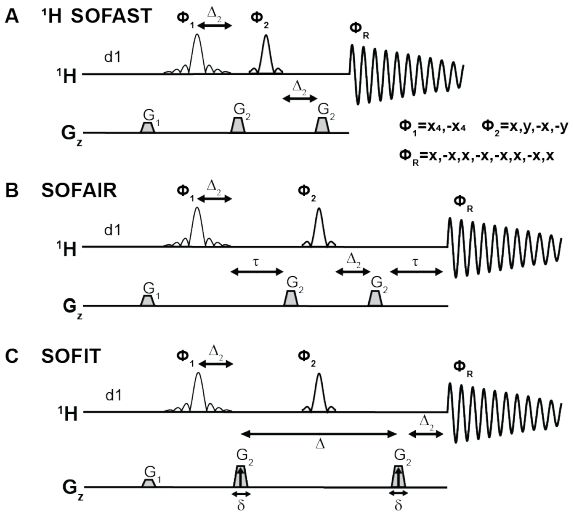

Figure 1. Pulse sequence schemes to record A) ^{1}H SOFAST, B) ^{1}H SOFAIR, and C) ^{1}H SOFIT spectra. The band-selective excitation is achieved by a 120° PC9 shaped pulse with a typical duration of 4.67 ms; the selective refocusing uses a RSNOB shaped pulse with a typical duration of 1.56 ms; both selective pulses, unless specified otherwise, are centered at 1 ppm for Methyl and 9 ppm for HN excitation, with a bandwidth of 3 ppm. The recycle delay d1 is set to 250 ms. Delay Δ_2 is to compensate for chemical shift evolution during the excitation pulse, which corresponds to 52.9% of PC9 pulse width. For SOFAIR (B), the transverse relaxation is encoded through the incrementation of the delay τ in pseudo-2D experiments, which typically ranges from 0 to 100 ms for therapeutic proteins. For SOFIT C), the diffusion encoding period Δ is set to 20 ms, and the diffusion coefficient is encoded through a single echo scheme by incrementing the gradient pulse strength G_2, which ranges between 4.82 – 45.74 G/cm with a duration (δ) of 6 ms. The phase cycles used for SOFAIR and SOFIT are identical to that from ^{1}H SOFAST; ϕ_1 = x, x, x, x, -x, -x, -x, -x; ϕ_2 = x, y, -x, -y; ϕ_R = x, -x, x, -x, -x, x, -x, x.

***R_2* Measurements from SOFAIR Are Sensitive to Protein Structural Stability and Self-Interactions.** To demonstrate the reliability of the new SOFAIR experiment, we evaluated the results of this approach and the traditional CPMG experiments.

using a well-behaved protein, bovine serum albumin (BSA, 66.5 kDa) under different conditions.

First, we tested the sensitivity of these NMR methods to detect protein destabilization and unfolding. To this end, a sample of BSA mixed with 1M guanidinium hydrochloride (GdnHCl) was prepared. Although GdnHCl alone was insufficient to fully denature BSA, it increased the protein's susceptibility to thermal stress. Initially, half of the GdnHCl-containing sample was incubated at an elevated temperature, resulting in a higher degree of protein denaturation, which was evident from the narrower dispersion of the amide and methyl proton signals observed (Fig. S1). To generate a series of samples with varied quantities of heat stressed protein, the thermally stressed BSA sample was mixed with the unstressed sample (also containing 1M GdnHCl) in different proportions. The R_2 values were then determined from these samples using SOFAIR, revealing a linear correlation between these R_2 values and the amount of heat-stressed protein present. Notably, samples with higher levels of heat-stressed protein exhibited substantially slower relaxation, likely attributable to a combination of protein unfolding, chemical modifications, and degradation (Fig. 2A, Table S1). Remarkably, the same trend was observed with the values measured by CPMG.

Second, we compared the two NMR methods for the detection of self-association. Here, a series of samples with different concentrations of BSA (10, 20, 50, 100, 200 mg/mL) were prepared. BSA is known to exist in a monomer/dimer/trimer equilibrium. Higher-order aggregates have also been observed in a concentration-dependent manner[39]. Thus, faster relaxation rates (i.e. larger R_2) can be

anticipated at higher BSA concentrations due to increased self-association. The R_2 values were measured for both the Methyl and the H(N) regions, the same as in the previous example. In both cases, a linear response with a positive slope was observed, indicative of faster relaxation. This is consistent with an increase in self-interactions and protein-protein association at higher BSA concentrations (Fig. 2B, Table S2). Once again, a similar trend was observed using the CPMG method. However, we showed that the estimated uncertainties of R_2 measurements were much larger when using the CPMG experiment as compared to SOFAIR, despite using a longer acquisition time (~40 min CPMG vs. ~20 min SOFAIR for both Methyl and H(N) regions). The difference was particularly evident for the lower BSA concentration samples at 10 and 20 mg/mL. Of note, if CPMG was collected with the same number of scans as SOFAIR, the duration of the experiment would extend to approximately 2 hours, and the signal-to-noise ratio of CPMG would still be approximately 6 times lower due to the implementation of the diffusion filter (Fig. S2).

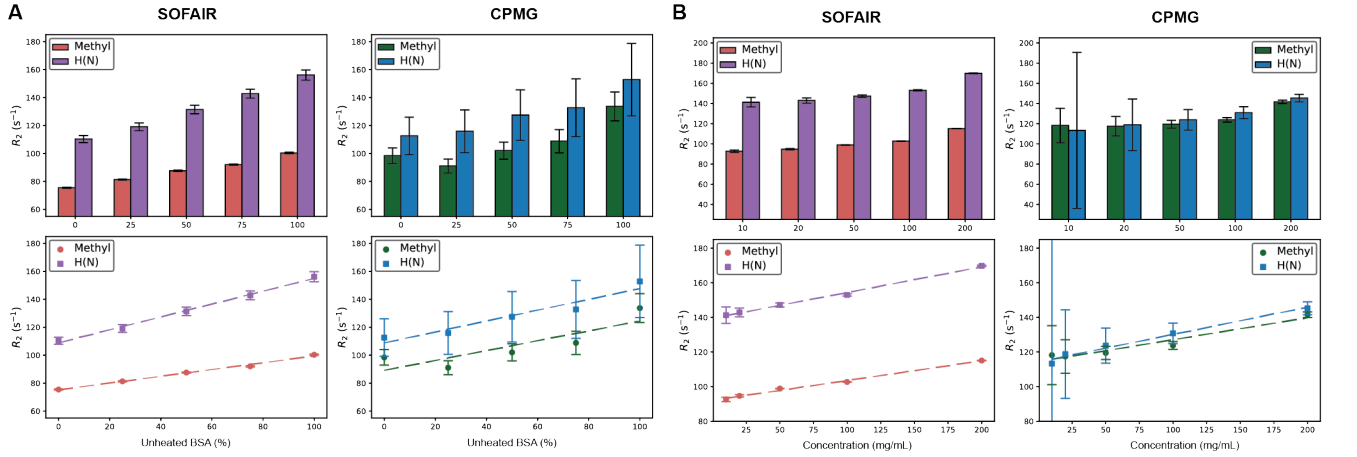

Figure 2. SOFAST-derived relaxation parameters are sensitive to protein structural stability and protein self-interactions. Comparison of apparent R_2 values measured from ^{1}H SOFAIR and R_2 from diffusion-filtered CPMG experiments for (A) mixture of unstressed and temperature-stressed BSA (0, 25, 50, 75, 100% of unstressed sample) both in the presence of 1M GdnHCl; and (B) Native BSA at various concentrations (10, 20, 50, 100, 200 mg/mL). Two different regions were integrated for R_2 measurements: the methyl proton region (Methyl, 1.2 to -0.8 ppm) and the amide proton region (H(N), 9.5 to 7.5 ppm). A) The relaxation parameters track linearly with the amount of heat-stressed protein mixed in the sample. Stressed proteins showed considerably slower relaxation than their unstressed counterparts. The outstanding sensitivity of SOFAIR compared to CPMG allows for highly accurate parameter extraction and quantification of the stressed species. B) Extracted R_2 values for both regions also exhibited a linear response to increased BSA concentration; this is consistent with an increase in self-interactions at higher protein concentrations, indicative of a shift in equilibrium of BSA from monomer to oligomer.

It is worth reiterating that the R_2 measured with SOFAIR has differential contributions for the H(N) and the Methyl region. Apart from local motions, for the H(N) region, the protons are only coupled with the Hα atoms, which are unaffected by the band-selective refocusing pulse, and, as a result, any unwanted *J*-modulation is effectively suppressed. The additional relaxation contribution compared to CPMG comes from the slow pulsing, which does not suppress the in-phase with anti-phase relaxation averaging[40]. This explains why the R_2 values for the amide proton region generally appear larger in the SOFAIR method compared to a fast pulsing CPMG. The differences in relaxation contributions between the two types of relaxation encoding are elegantly demonstrated in the work of Kiraly *et al.*[38], which served as inspiration for the development of SOFAIR.

In contrast, measuring R_2 for the Methyl region using SOFAIR presents additional complexity. Since the methyl protons may also have couplings with other protons within the selectively refocused region, the extracted R_2 is potentially influenced by *J*-evolution. Nevertheless, as demonstrated in the above examples, the general trends observed have consistently aligned with measurements from the CPMG method. In our opinion, the notable advantages of improved sensitivity and reduced acquisition time offered by SOFAIR, still hold substantial value even in the absence of absolute measurement matches between the two methods.

Lastly, we examined the potential influence of the short recycle delay on R_2 measurements as well as the potential effects related to differential heat deposition during the experiments. Our investigations revealed that these factors had a negligible impact on the R_2 values obtained here (Fig. S3 and S4).

***D* Measurements from SOFIT Are Comparable to Conventional DOSY Experiments.** To assess the performance of the SOFIT pulse sequence, diffusion coefficients of BSA at various concentrations were measured using both the SOFIT experiment and the standard DOSY experiment with stimulated echo pulsed field gradient (Fig. 3A).

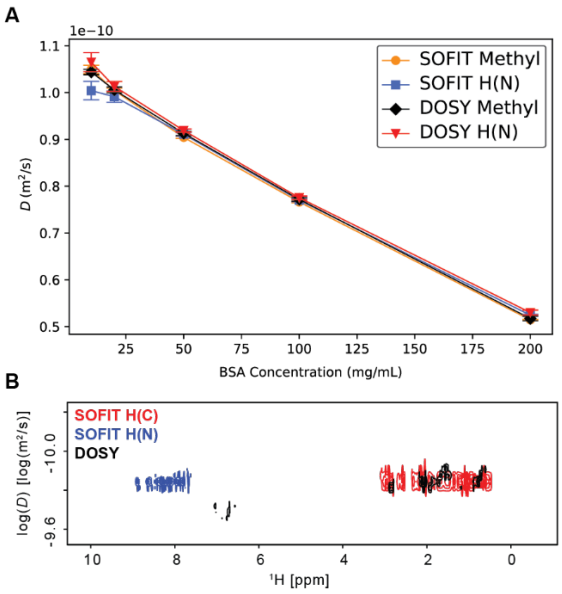

Figure 3. Evaluation of SOFIT using BSA at various concentrations and diphtheria toxin domain A. A) Diffusion measurements of both SOFIT and DOSY were obtained by fitting the integrals of either the methyl or H(N) regions to Eq. 2. SOFIT measurements revealed a decrease in *D* as the concentration of BSA increased. This trend aligns with the results obtained from the conventional DOSY experiment using stimulated echo pulse field gradient. The observed decrease in *D* is likely attributed to both BSA self-association and an increase in sample viscosity resulting from higher protein concentration, consistent with R_2 results. B) DOSY experiments of the diphtheria toxin catalytic domain were carried out at 14 mg/mL in 10 mM sodium phosphate buffer, pH 7. The SOFIT and regular DOSY experiments were acquired for 14 min and 292 min, respectively. Both experiments yielded comparable *D* measurements, while the SOFAST approach demonstrated better sensitivity in both methyl and amide proton regions.

Expectedly, both methods led to the extraction of nearly identical *D* values, with a discrepancy of less than 5% (Table S2). As noted above, BSA tends to self-interact at higher concentrations, resulting in a decreased diffusion coefficient. Both SOFIT and DOSY measurements consistently demonstrated a gradual reduction in the diffusion coefficient with increasing concentration, indicating an increase in the oligomeric states to monomer ratios of BSA, as expected. It is worth mentioning that the change in *D* is likely the result of both BSA self-association and also increased viscosity due to increased protein concentration.

Additionally, we note that the SOFIT method has a potential limitation due to incomplete signal attenuation. This limitation arises from the fact that in SOFIT the translational diffusion is encoded using a single echo scheme, and the relaxation characteristics of the biologic constrain the maximum diffusion encoding period used. The total decay in our SOFIT experiments of BSA ranged from 40 – 60%, whereas regular DOSY experiments typically achieved more than 90% signal decay. Consequently, SOFIT is primarily applicable to the study of smaller therapeutic proteins and biologics with favorable relaxation and diffusion properties, where sufficient gradient-induced decay can be achieved without compromising sensitivity. However, its effectiveness and sensitivity boost may be diminished for higher molecular weight species, like oligomers, which can be in the 100s kDa to MDa range.

In a separate example, the diffusion of a small protein domain (Fig. 3B), the catalytic domain of diphtheria toxin (~22 kDa), was examined. In this case, the methyl region of the DOSY spectra obtained from both SOFIT and conventional DOSY methods exhibited a nice overlap. However, the conventional DOSY method suffered from low sensitivity, making it difficult to observe any signals in the amide proton region of the DOSY spectrum. Conversely, the SOFIT experiment showed decent sensitivity when the amide protons were selectively excited, enabling the extraction of diffusion coefficients from that region. Although not necessary in this case, the good sensitivity in the H(N) region can be useful in characterizing therapeutic proteins, especially when excessive excipient signals

obscure the methyl region of the spectrum. Analyzing data from such regions with multiexponential fittings is possible, but these efforts often result in poor fits due to the complexity caused by high degrees of overlap and differential line broadening[25].

Assessment of Biologic Formulation Suitability through the Combination of SOFAIR and SOFIT. To demonstrate the potential of SOFAIR and SOFIT experiments in the developability of biologic drugs, we applied the methods presented here to a proprietary therapeutic protein (biologic A). For this study, we compared the behavior of the biologic A in three different formulations. To minimize interference from excipient peaks, the amide protons were selected for analysis rather than the methyl proton region.

The SOFAIR measurements revealed that R_2 was the largest in formulation 1 and the smallest in formulation 3, as shown in Fig. 4A. On the other hand, SOFIT measurements provided information about D, which was also found to be the largest in formulation 1 and the smallest in formulation 3, as depicted in Fig. 4B. The simultaneous decrease in both the transverse relaxation rate and the diffusion coefficient (which corresponds to an increase in the average hydrodynamic radius, R_h, calculated from the Stokes–Einstein equation, Eq. 3) suggests the molecules appear to have larger average size but relax slower in formulation 2 and 3, indicative of a loss of conformational stability of the biologic in those two formulations[26]. In another word, the biologic is the most stable in formulation 1, where R_2 and R_h values are more consistent with that of a globular protein[26, 41] (Fig. 4C). All relevant parameters are summarized in Table S3.

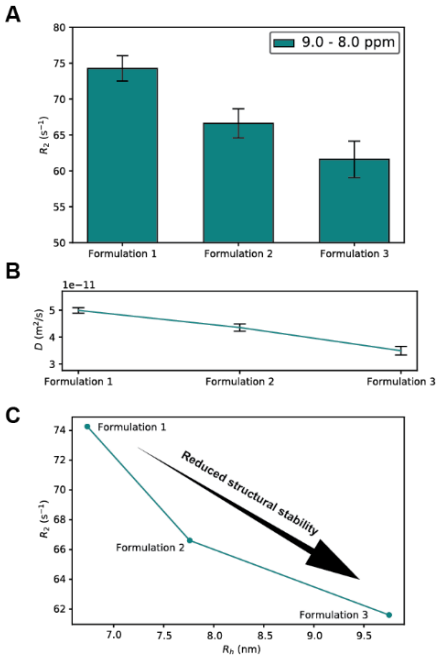

Figure 4. Assessment of the structural stability and formulation suitability of Biologic A using SOFAIR and SOFIT. The behavior of Biologic A in three different formulations was evaluated using (A) SOFAIR and (B) SOFIT. The concomitant decrease in transverse relaxation rate and diffusion coefficient (and correspondingly an increase in the average hydrodynamic radius) indicates a decrease in the biologic's structural stability, as shown in (C).

We would like to stress that this conclusion was derived from experiments with a total data acquisition time of fewer than 20 minutes for each sample. In contrast, the fitting uncertainty associated with employing CPMG and DOSY experiments for a low concentration sample, such as the one studied here, and in particular for analyzing the low sensitivity amide proton region, could take more than 10 hours to obtain reliable results.

CONCLUSIONS

In conclusion, we have introduced a pair of simple yet highly effective NMR experiments for rapid measurement of proton transverse relaxation rate (R_2) and translational diffusion coefficient (D), two parameters that serve as sensitive indicators of protein

behavioral changes in solution. By combining these two complementary methods, valuable information can be obtained rapidly regarding the protein's structural stability and association states. This capability makes the SOFAIR and SOFIT techniques well-suited for screening and characterizing low concentration and/or fully formulated samples.

These newly proposed experiments offer significant advantages over CPMG and DOSY for measuring R_2 and D: (1) Due to the very short recycle delay required for SOFAST experiments, measurements can be performed with significantly reduced experimental duration, achieving high sensitivity per time unit. Furthermore, additional increments can be acquired within each experiment without adding too much experiment time. This could improve multiexponential fits to observe differential changes in the behavior of biologics. (2) The band-selective nature of the SOFAST experiments allows for the selective excitation of regions devoid of excipient or other undesired signals, providing accurate measurements that correspond only to the biological target of interest. This also eliminates the need for water suppression, as R_2 and D measurements are typically performed on the methyl and/or amide proton regions. For low-concentration samples (*e.g.*, less than 5 mg/mL), the experiment duration can improve from 10 – 20 hours to less than 15 minutes for a comparable signal-to-noise ratio, saving time while avoiding any potential property changes of the sample during data acquisition. Furthermore, rapid R_2 and D measurements can also be interpreted alongside results from other ^{1}H 1D NMR techniques, such as comparability studies like PROFILE[24], to provide comprehensive assessments of large number of samples without sacrificing throughput. (3) By fixing the relaxation or diffusion encoding parameters, SOFAIR and SOFIT can be effectively utilized as highly sensitive relaxation and diffusion-filtered experiments. Together, these advancements enable the proposed NMR experiments to serve as *bona fide* screening and characterization tools to influence molecular design in drug discovery, particularly in situations where sample quantities are scarce. We anticipate that these novel methods will strongly impact the analysis of biologics, especially in the context of pharmaceutical research.

AUTHOR INFORMATION

Corresponding Authors

* Xingjian Xu – *Analytical Research & Development, MRL, Merck & Co., Inc., Rahway, NJ, USA*;

Email: xingjian.xu@merck.com

* Guilherme Dal Poggetto – *Analytical Research & Development, MRL, Merck & Co., Inc., Rahway, NJ, USA*;

Email: guilherme.dal.poggetto@merck.com

* Pablo Trigo-Mourino – *Analytical Research & Development, MRL, Merck & Co., Inc., Boston, MA, USA*;

Email: pablo.trigo.mourino@merck.com

Author Contributions

The manuscript was written through the contributions of all authors. All authors have given approval to the final version of the manuscript.

ACKNOWLEDGMENT

X.X. acknowledges support from the MRL Postdoctoral Research Program. The authors thank Rahul Upadhya for providing samples of Biologic A.

REFERENCES

1. Kesik-Brodacka, M., Progress in biopharmaceutical development. *Biotechnol Appl Biochem* **2018,** *65* (3), 306-322.
2. Crescioli, S.; Kaplon, H.; Chenoweth, A.; Wang, L.; Visweswaraiah, J.; Reichert, J. M., Antibodies to watch in 2024. *MAbs* **2024,** *16* (1), 2297450.
3. Surowka, M.; Klein, C., A pivotal decade for bispecific antibodies? *MAbs* **2024,** *16* (1), 2321635.
4. Hughes, J. P.; Rees, S.; Kalindjian, S. B.; Philpott, K. L., Principles of early drug discovery. *Br J Pharmacol* **2011,** *162* (6), 1239-49.
5. Sampathkumar, K.; Kerwin, B. A., Roadmap for Drug Product Development and Manufacturing of Biologics. *J Pharm Sci* **2024,** *113* (2), 314-331.

6. Fernandez-Quintero, M. L.; Ljungars, A.; Waibl, F.; Greiff, V.; Andersen, J. T.; Gjolberg, T. T.; Jenkins, T. P.; Voldborg, B. G.; Grav, L. M.; Kumar, S.; Georges, G.; Kettenberger, H.; Liedl, K. R.; Tessier, P. M.; McCafferty, J.; Laustsen, A. H., Assessing developability early in the discovery process for novel biologics. *MAbs* **2023,** *15* (1), 2171248.

7. Mieczkowski, C.; Zhang, X.; Lee, D.; Nguyen, K.; Lv, W.; Wang, Y.; Zhang, Y.; Way, J.; Gries, J. M., Blueprint for antibody biologics developability. *MAbs* **2023,** *15* (1), 2185924.

8. Kuzman, D.; Bunc, M.; Ravnik, M.; Reiter, F.; Zagar, L.; Boncina, M., Long-term stability predictions of therapeutic monoclonal antibodies in solution using Arrhenius-based kinetics. *Sci Rep* **2021,** *11* (1), 20534.

9. Zarzar, J.; Khan, T.; Bhagawati, M.; Weiche, B.; Sydow-Andersen, J.; Alavattam, S., High concentration formulation developability approaches and considerations. *MAbs* **2023,** *15* (1), 2211185.

10. Yang, X.; Xu, W.; Dukleska, S.; Benchaar, S.; Mengisen, S.; Antochshuk, V.; Cheung, J.; Mann, L.; Babadjanova, Z.; Rowand, J.; Gunawan, R.; McCampbell, A.; Beaumont, M.; Meininger, D.; Richardson, D.; Ambrogelly, A., Developability studies before initiation of process development: improving manufacturability of monoclonal antibodies. *MAbs* **2013,** *5* (5), 787-94.

11. Karplus, M.; Kuriyan, J., Molecular dynamics and protein function. *Proc Natl Acad Sci U S A* **2005,** *102* (19), 6679-85.

12. Houde, D. J.; Berkowitz, S. A., Chapter 4 - An Introduction and Hierarchical Organization of the Biophysical Tool in Section II. In *Biophysical Characterization of Proteins in Developing Biopharmaceuticals*, Houde, D. J.; Berkowitz, S. A., Eds. Elsevier: Amsterdam, 2015; pp 79-85.

13. Alderson, T. R.; Kay, L. E., NMR spectroscopy captures the essential role of dynamics in regulating biomolecular function. *Cell* **2021,** *184* (3), 577-595.

14. Ishima, R.; Torchia, D. A., Protein dynamics from NMR. *Nat Struct Biol* **2000,** *7* (9), 740-743.

15. Kovermann, M.; Rogne, P.; Wolf-Watz, M., Protein dynamics and function from solution state NMR spectroscopy. *Q Rev Biophys* **2016,** *49*, e6.

16. Dubey, A.; Takeuchi, K.; Reibarkh, M.; Arthanari, H., The role of NMR in leveraging dynamics and entropy in drug design. *J Biomol NMR* **2020,** *74* (10-11), 479-498.

17. Bramham, J. E.; Podmore, A.; Davies, S. A.; Golovanov, A. P., Comprehensive Assessment of Protein and Excipient Stability in Biopharmaceutical Formulations Using (1)H NMR Spectroscopy. *ACS Pharmacol Transl Sci* **2021,** *4* (1), 288-295.

18. Phyo, P.; Zhao, X.; Templeton, A. C.; Xu, W.; Cheung, J. K.; Su, Y., Understanding molecular mechanisms of biologics drug delivery and stability from NMR spectroscopy. *Adv Drug Deliv Rev* **2021,** *174*, 1-29.

19. Ma, J.; Pathirana, C.; Liu, D. Q.; Miller, S. A., NMR spectroscopy as a characterization tool enabling biologics formulation development. *J Pharm Biomed Anal* **2023,** *223*, 115110.

20. Zhang, C.; Gossert, S. T.; Williams, J.; Little, M.; Barros, M.; Dear, B.; Falk, B.; Kanthe, A. D.; Garmise, R.; Mueller, L.; Ilott, A.; Abraham, A., Ranking mAb-excipient interactions in biologics formulations by NMR spectroscopy and computational approaches. *MAbs* **2023,** *15* (1), 2212416.

21. Gardner, K. H.; Kay, L. E., The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins. *Annu Rev Biophys Biomol Struct* **1998,** *27*, 357-406.

22. Kanelis, V.; Forman-Kay, J. D.; Kay, L. E., Multidimensional NMR methods for protein structure determination. *IUBMB Life* **2001,** *52* (6), 291-302.

23. Rehm, T.; Huber, R.; Holak, T. A., Application of NMR in structural proteomics: screening for proteins amenable to structural analysis. *Structure* **2002,** *10* (12), 1613-8.

24. Poppe, L.; Jordan, J. B.; Lawson, K.; Jerums, M.; Apostol, I.; Schnier, P. D., Profiling formulated monoclonal antibodies by (1)H NMR spectroscopy. *Anal Chem* **2013,** *85* (20), 9623-9.

25. Falk, B. T.; Liang, Y.; McCoy, M. A., Diffusion Profiling of Therapeutic Proteins by Using Solution NMR Spectroscopy. *Chembiochem* **2019,** *20* (7), 896-899.

26. Falk, B. T.; Liang, Y.; Bailly, M.; Raoufi, F.; Kekec, A.; Pissarnitski, D.; Feng, D.; Yan, L.; Lin, S.; Fayadat-Dilman, L.; McCoy, M. A., NMR Assessment of Therapeutic Peptides and Proteins: Correlations That Reveal Interactions and Motions. *Chembiochem* **2020,** *21* (3), 315-319.

27. Kheddo, P.; Cliff, M. J.; Uddin, S.; van der Walle, C. F.; Golovanov, A. P., Characterizing monoclonal antibody formulations in arginine glutamate solutions using ^{1}H NMR spectroscopy. *Mabs* **2016,** *8* (7), 1245-1258.

28. Schanda, P.; Brutscher, B., 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* **2005,** *127* (22), 8014-5.

29. Schanda, P.; Kupce, E.; Brutscher, B., SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J Biomol NMR* **2005,** *33* (4), 199-211.

30. Schanda, P.; Forge, V.; Brutscher, B., HET-SOFAST NMR for fast detection of structural compactness and heterogeneity along polypeptide chains. *Magn Reson Chem* **2006,** *44 Spec No*, S177-84.

31. Sauve, S.; Gingras, G.; Aubin, Y., NMR study of mutations of glycine-52 of the catalytic domain of diphtheria toxin. *J Pharm Biomed Anal* **2018,** *150*, 72-79.

32. Panchal, J.; Falk, B. T.; Antochshuk, V.; McCoy, M. A., Investigating protein-excipient interactions of a multivalent V(HH) therapeutic protein using NMR spectroscopy. *MAbs* **2022,** *14* (1), 2124902.

33. Johnson, C. S., Diffusion ordered nuclear magnetic resonance spectroscopy: principles and applications. *Prog Nucl Mag Res Sp* **1999,** *34* (3-4), 203-256.

34. Carr, H. Y., Citation Classic - Effects of Diffusion on Free Precession in Nuclear Magnetic-Resonance Experiments. *Cc/Phys Chem Earth* **1983,** (20), 24-24.

35. Dal Poggetto, G.; Castanar, L.; Foroozandeh, M.; Kiraly, P.; Adams, R. W.; Morris, G. A.; Nilsson, M., Unexploited Dimension: New Software for Mixture Analysis by 3D Diffusion-Ordered NMR Spectroscopy. *Anal Chem* **2018,** *90* (22), 13695-13701.

36. Birlirakis, N.; Guittet, E., A new approach in the use of gradients for size-resolved 2D-NMR experiments. *Journal of the American Chemical Society* **1996,** *118* (51), 13083-13084.

37. Foroozandeh, M.; Castanar, L.; Martins, L. G.; Sinnaeve, D.; Dal Poggetto, G.; Tormena, C. F.; Adams, R. W.; Morris, G. A.; Nilsson, M., Ultrahigh-Resolution Diffusion-Ordered Spectroscopy. *Angew Chem Int Ed Engl* **2016,** *55* (50), 15579-15582.

38. Kiraly, P.; Dal Poggetto, G.; Castanar, L.; Nilsson, M.; Deak, A.; Morris, G. A., Broadband measurement of true transverse relaxation rates in systems with coupled protons: application to the study of conformational exchange. *Chem Sci* **2021,** *12* (34), 11538-11547.

39. Bhattacharya, A.; Prajapati, R.; Chatterjee, S.; Mukherjee, T. K., Concentration-Dependent Reversible Self-Oligomerization of Serum Albumins through Intermolecular β-Sheet Formation. *Langmuir* **2014,** *30* (49), 14894-14904.
40. Meersmann, T.; Bodenhausen, G., Relaxation-induced oscillations of spin-echo envelopes. *Chem Phys Lett* **1996,** *257* (3-4), 374-380.
41. La Verde, V.; Dominici, P.; Astegno, A., Determination of Hydrodynamic Radius of Proteins by Size Exclusion Chromatography. *Bio Protoc* **2017,** *7* (8), e2230.