The University of Manchester Research

19F NMR as a Tool for Monitoring Individual Differentially Labeled Proteins in Complex Mixtures

Document Version
Accepted author manuscript

Link to publication record in Manchester Research Explorer

OPEN ACCESS

^{19}F NMR as a Tool for Monitoring Individual Differentially Labeled Proteins in Complex Mixtures

John M. Edwards,[†] Jeremy P. Derrick,[‡] Christopher F. van der Walle,[§] Alexander P. Golovanov[*,†]

[†] Manchester Institute of Biotechnology, School of Chemistry, Faculty of Science and Engineering, University of Manchester, Manchester M1 7DN, U.K.

[‡] School of Biological Sciences, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, University of Manchester, Manchester M13 9PL, U.K.

[§] Formulation Sciences, MedImmune Ltd, Granta Park, Cambridge CB21 6GH, U.K.

ABSTRACT: The ability to monitor the behavior of individual proteins in complex mixtures has many potential uses, ranging from analysis of protein interactions in highly-concentrated solutions, modelling biological fluids or the intracellular environment, to optimizing biopharmaceutical co-formulations. Differential labelling NMR approaches, which traditionally use ^{15}N or ^{13}C isotope incorporation during recombinant expression, are not always practical in cases when endogenous proteins are obtained from an organism, or where the expression system does not allow for efficient labelling, especially for larger proteins. This study proposes differential labelling of proteins by covalent attachment of ^{19}F groups with distinct chemical shifts, giving each protein a unique spectral signature which can be monitored by ^{19}F NMR without signal overlap, even in complex mixtures, and without any interfering signals from the buffer or other unlabeled components. Parameters such as signal intensities, translational diffusion coefficients and transverse relaxation rates, which report on the behaviour of individual proteins in the mixture, can be recorded even for proteins as large as antibodies at a wide range of concentrations.

INTRODUCTION

In general proteins do not function in isolation, instead acting as parts of larger, and sometimes transient, multi-component assemblies. Such systems are often reconstituted *in vitro* using defined mixtures of protein components. Proteins in mixtures frequently interact with themselves and with each other, forming stable or transient complexes, which can be remodeled in response to stimuli or changes in conditions.[1-2] Understanding the behavior of individual proteins in a complex mixture is required in a wide range of situations. Examples include determination of optimal conditions for co-crystallization[3], structural and biophysical studies by NMR, SAXS and other methods[4], and optimization of co-formulations for medicine and/or biotechnology applications.[5] Protein-based biopharmaceuticals, in particular monoclonal antibodies (mAbs), can treat a wide range of diseases and constitute the fastest growing segment of pharmaceutical market.[5-7] There are several reasons why protein-based biopharmaceuticals may need to be prepared as mixtures. Firstly, some are formulated with additional proteins, such as human albumin, to improve their overall stability and shelf life.[8] Secondly, there may be therapeutic advantages in administering two biopharmaceuticals in combination (e.g. in cancer immunotherapy).[9] Similarly, synthetic biology applications, catalysis using mixtures of enzymes, and other applications may benefit from using protein mixtures.[10-11] It is therefore necessary to characterize the interactions within the mixtures, as well as to detect is-

sues such as aggregation and self-association, in order to optimise such formulations.

Proteins which are expressed recombinantly or in cell-free systems often can be uniformly labelled or incorporate site-specific NMR-visible labels (e.g. ^{13}C-labelled methyl groups in Met/Ile sidechains), natural amino acids individually labelled with ^{15}N and/or ^{13}C, or unnatural amino acids such as 5-fluorotryptophan.[12-14] For proteins which can be labelled isotopically with ^{15}N and ^{13}C via recombinant expression, different mixture components can be tracked simultaneously using differential-labelling strategies and isotopically-discriminated NMR experiments.[15-18] These strategies result in defined NMR signals representative of each specific protein in the mixture. This approach is, however, not practical to implement for endogenous proteins produced from natural sources (e.g. purified from blood, tissues or plants), or those expressed in mammalian cell cultures (e.g. Chinese hamster ovary cells).[19] A more practical option for mAbs and other proteins which are difficult or expensive to label biosynthetically, is to attach the label post-translation by chemical modification. Here we describe a method of differential labelling based on chemical modification of the protein using fluorine containing groups.

The ^{19}F nucleus is useful for NMR, with spin 1/2, high gyromagnetic ratio and 100% natural abundance, which gives rise to strong NMR signals.[20-21] The ^{19}F chemical shift is also extremely sensitive to the environment of the nucleus, resulting in a much wider dispersion of signals compared to proton NMR. Since ^{19}F does not occur naturally in proteins or typical buffer components there are no background signals to interfere with those of interest, and no signals that require suppression (such as water in ^{1}H NMR). The only signals in the ^{19}F spectrum will be those of the artificially added ^{19}F tags, greatly easing data interpretation.

Chemical modification of proteins with fluorinated tags - either ^{19}F or ^{18}F depending on the intended application - is typically achieved by targeting the nucleophilic side chains of cysteine and lysine residues. Compounds for covalent linking to cysteine residues generally consist of a trifluoromethyl moiety linked to a group reactive to nucleophilic substitution, such as the commonly used 3-bromo-1,1,1-trifluoroacetone.[22-24] This approach does require that the protein has accessible cysteine sulfhydryl groups so partial reduction may be needed in some cases to activate cysteines in disulphide bridges. The alternative approach of targeting lysine side chains typically allows labelling at a greater number of sites on the protein but

compounds capable of this, such as p-nitrophenyl trifluoroacetate, often require high pH (9 – 10) for efficient reaction, which may present problems for protein solubility and stability.[25] Modification of serine and threonine hydroxyl groups is another possibility.[20] Radical trifluoromethylation of tryptophan, tyrosine and phenylalanine residues has also been recently reported.[26-27] Although acquisition of site-specific information requires addition of the fluorine label at a specific site, non-site specific molecular labelling can be invaluable for describing the general behaviour of the protein as a whole, its interactions with other proteins, self-association and aggregation. Fluorine labelling of single proteins has led to important structural, [28-29] mechanistic and functional insights.[14, 21, 30-32] It has also been used previously to investigate binding of small molecules to a pair of 5-fluorotryptophan selectively-labelled proteins.[33] To our knowledge chemical modification using differential fluorine tags in protein mixtures has not previously been applied to characterization of protein-protein interactions.

In this study we demonstrate that labelling proteins differentially with fluorine tags allows monitoring of their individual behavior in complex mixtures using solution NMR spectroscopy. Self-association and binding between the components can be detected even in cases where the interactions are transient. This approach opens the way to optimize solution conditions or formulations for complex protein-based systems in a wide range of applications, from structural biology to biotechnology to pharmaceutical development.

MATERIALS AND METHODS

[19]F Labelling

Two monoclonal IgG antibody samples, denoted as mAb2 (MW 144.8 kDa, pI = 8.44, as used previously in Refs [34-35]) and COE-19 (MW 148 kDa, pI=7.4) were supplied by MedImmune Ltd., Cambridge, UK. Bovine serum albumin (BSA, MW 66.5 kDa) and Cytochrome C (CytC, MW 12 kDa) were purchased from Sigma-Aldrich (ca. no. A3733 and C2506, respectively).

For labelling with 3-bromo-1,1,1-trifluoroacetone (BTFA – supplied by Sigma-Aldrich, catalogue number 374059), protein solutions at 250 μM or less were prepared in 100 mM borate, 400 mM L-Arg (Sigma –Aldrich Ca. No. W381918) buffer, adjusted to pH 9.5 with concentrated HCl. Typical reaction volumes were 10 – 25 ml. 1 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP, Sigma-Aldrich, ca. no. C4706) was added where reduction

of disulphide bridges was required. 5 mM BTFA was added and the reaction mixture then agitated gently on a rocker platform for 16 hours. The reaction mixture was then dialyzed against the NMR buffer (100 mM pH 7.2 phosphate for the BSA/CytC system, 100 mM pH 5.5 acetate for the mAbs) to remove unreacted BTFA.

For labelling with 1-(4-(trifluoromethyl)benzyl)-1H-pyrrole-2,5-dione (TFBPD – custom-synthesised and supplied by Charnwood Molecular Ltd., Loughborough, UK) the proteins were first transferred into pH 7.2 100 mM sodium phosphate buffer and diluted to a concentration of 5 - 10 mg/ml. TFBPD is only minimally soluble under aqueous conditions so when 10 mg of crystalline TFBPD was added to a 25 ml reaction volume the majority of the crystals remained undissolved. The reaction mixture was agitated gently for 16 hours, with some further solubilisation of the TFBPD crystals as the reaction progressed. After 16 hours the mixture was spun down to remove any remaining insoluble material, and then dialysed against the NMR buffer to remove any remaining soluble unreacted small molecule ^{19}F species.

For N-(ε-trifluoroacetylcaproyloxy) succinimide ester (TFCS – supplied by Fisher Scientific, Ca. no. 22299) labelling the proteins were transferred into 100 mM pH 7.2 phosphate buffer. In the case of BSA and CytC, the lyophilized proteins were dissolved in the phosphate buffer. In the case of the mAbs, which were supplied as 45 mg/ml stocks in citrate buffer, they were diluted to 5 mg/ml using the phosphate buffer. Typical reaction volumes were 10 – 50 ml. 10 mg of TFCS was dissolved in 200 µl of DMSO and added to the reaction, mixing thoroughly during addition. The reaction mixture was then gently agitated overnight at room temperature. Unreacted TFCS was removed by dialysis against the NMR buffer.

Protein concentration was measured by optical density at 280 nm using extinction coefficients of 1.435 ml mg^{-1} cm^{-1} for mAb2 and 1.780 ml mg^{-1} cm^{-1} for COE-19. For all tags the degree of labelling achieved in each reaction batch was measured by comparison of the ^{19}F signal integral from labelled protein with that of a 1 mM reference sample of the unreacted small molecule tag. The labelling efficiency as a percentage (L^{eff}) was defined as:

$$L^{eff} = 100\% \times (C_{ref} \times I_{prot})/(C_{prot} \times I_{ref})$$

where I_{prot} and I_{ref} are the ^{19}F signal integrals, and C_{prot} and C_{ref} are the concentrations of the labelled protein and the reference sample, respectively. Since the protein substrates used typically have more than one potential reactive site, the reaction efficiency in many cases is greater

than 100%, indicating more than one tag attached per protein molecule on average.

For size exclusion chromatography coupled with multi angle light scattering (SEC-MALS) analysis, samples (0.5 mL at 1-2 mg/mL) were loaded onto a Superdex 200 10/300GL column (GE life-sciences, 0.75 mL/min in NMR sample buffer) and passed through a Wyatt DAWN Heleos II EOS 18-angle laser photometer coupled to a Wyatt Optilab rEX refractive index detector. Data were analyzed using Astra 6 software (Wyatt Technology Corp., CA, USA).

NMR Experimental Details

All NMR experiments were conducted using a 500 MHz Bruker Avance III spectrometer equipped with QCI-F cryoprobe (with cooled ^{1}H and ^{19}F channels) in 5 mm NMR tubes, at 298K in the case of the BSA/CytC system, and 313K in the case of mAbs. 5% D$_2$O was added to the NMR buffer to provide the NMR lock. Control experiments conducted in 3 mm NMR tubes gave very similar values of diffusion coefficient D_L suggesting that the influence of thermal convection in the sample tubes was negligible for the protein solutions used. Trifluoroethanol (TFE), which at low concentrations was found not to interact with proteins, was added to all samples as an internal reference, to check the accuracy of dilution and for calculating the ^{19}F signal loss for proteins.

To prepare the sample for the mAb dilution series, equimolar amounts of labelled mAb2 and COE-19 were co-concentrated by ultrafiltration, using a Vivaspin centrifugal concentrator with a 30 kDa molecular weight cutoff, to a total mAb concentration of 200 mg/ml. Dilution was carried out using the flow-through from the centrifugal concentrator to ensure an accurate buffer match. Changes in the translational diffusion coefficient (D_L) in Diffusion Ordered Spectroscopy (DOSY) were monitored using SE-PFG (stimulated echo- pulsed-field gradient) with bipolar gradients pulses with water suppression (Bruker's standard pulse program stebpgp1s19 set up for ^{19}F). The diffusion time (Δ) and the gradient length (δ) were set to 100 ms and 6.0 ms respectively for experiments with the mAb2/COE-19 system, and 200 ms and 2.0 ms respectively for the BSA/CytC system. The acquisition time and relaxation delay were 217 ms and 3.0 s respectively, with a gradient pulse of 45 G/cm. 16 scans were collected per experiment, except where otherwise stated, across a spectral width of 40 ppm with 32 linear gradient steps with 2–98% gradient intensity. Transverse relaxation rates (R_2) were measured with a modified Carr-

Purcell-Meiboom-Gill sequence (CPMG)[36] using a relaxation delay of 5 seconds. Numbers of spin-echoes ranging from 2 to 512 (corresponding to times of 2.8 ms to 730 ms) were applied. No ^{1}H-^{19}F decoupling was required in any of the experiments. Spectra were processed and analyzed using Topspin 3.5 and Dynamics Center 2.4.8.

For the BSA/CytC + NaCl titration concentrated 5M NaCl in pH 7.2 phosphate buffer was added in order to minimize the dilution of the sample. For the mAb titration with arginine glutamate (Arg·Glu), to ensure accuracy and avoid sample dilution, a stock solution of equimolar L-Arg and L-Glu was prepared with the same pH as the sample, split into aliquots, and lyophilized. The NMR sample was then successively reconstituted with the next Arg·Glu aliquot to raise the Arg·Glu concentration by 100 mM at each data point step.

For the titration of mAb2 into COE-19 a two NMR sample tube system was used. Initially tube 1 contained 200 mg/ml COE-19 only, while tube 2 contained a mixture of 200 mg/ml mAb2 and 200 mg/ml COE-19. These defined the start and end points of the titration. By successively exchanging defined volumes of solution between the two tubes, a range of intermediate mAb2 concentrations could be sampled, without changing the COE-19 concentration or the overall solution volume.

The effective radius of a protein clusters R_{eff} was derived from the apparent diffusion coefficient D_L measured using DOSY, by means of the Stokes-Einstein equation (Eq. 1) which links the diffusion coefficient D_o at infinite dilution with the apparent molecular radius R_h, and Eq. 2 which describes how the apparent diffusion coefficient D_L in concentrated solution, where the molecular crowding effects become dominant, relates to D_o, using the Medina-Noyola model.[37-38]

Equation 1: $\qquad D_0 = \frac{k_B T}{6\pi\eta R_h}$

where k_B is the Boltzmann constant, T is absolute temperature and η is the viscosity of the solution.

Equation 2: $\qquad D_L = \frac{D_0 * (1-\varphi)^3}{\left(1 + \frac{3}{2}\varphi + 2\varphi^2 + 3\varphi^3\right)}$

where φ is the total volume fraction of the proteins in the solution, which was calculated as:

Equation 3: $\qquad \varphi = c/(1000 * f)$

where c is the total concentration of protein in mg/ml and f is the protein density factor, taken as 1.25 g/ml.[39]

Combining Eqs 1 and 2 yields Eq. 4 which links the effective radius of the diffusing protein or cluster R_{eff} with

D_L by compensating for the expected effects of protein crowding at high concentrations:

Equation 4: $\qquad R_{eff} = \frac{k_B T}{6\pi\eta D_L} \frac{(1-\varphi)^3}{\left(1 + \frac{3}{2}\varphi + 2\varphi^2 + 3\varphi^3\right)}$

The transverse relaxation rate R_2 is expected to increase if chemical exchange is present. In the absence of chemical exchange, assuming that for large proteins R_2 is proportional to correlation time τ_c, this relaxation rate is also proportional to the effective spherical volume of the protein with the effective radius R_h [35].

Equation 5: $\quad R_2 \;\propto\; \tau_c = \frac{4\pi R_h^3 \eta}{3 k_B T} \;\propto\; R_h^3 \eta$

The signal loss in ^{1}H and ^{19}F NMR spectra (relative to the expected value) for protein concentration [P] was calculated as:

Equation 6: $\quad \text{Signal Loss} = 100\% \times \left(1 - \left(\frac{I^{[P]} \times I_{TFE}^{Ref}}{I_{TFE}^{[P]} \times I^{Ref}}\right)\right)$

where $I^{[P]}$ is the intensity of the labelled protein signal at a given concentration, $I_{TFE}^{[P]}$ is the intensity of the TFE signal at the same point. As the reference state we chose the most diluted sample in the series, with I^{REF} and I_{TFE}^{Ref} being the intensity of the protein and TFE signals in the reference sample, respectively. For the titration experiment presented in Figure 6, I^{REF} at higher concentrations was extrapolated from the intensity observed at 20 mg/ml by multiplying by the nominal concentration increase. No visible precipitation was observed in any of the samples, meaning that all aggregates or clusters present in the samples were soluble.

RESULTS

Chemical Modification of proteins with ^{19}F labels

Several previously reported fluorine tags for chemical modifications of proteins were investigated in this study, with a focus on maximizing labelling efficiency, and obtaining a set of tags with a range of characteristic ^{19}F chemical shifts. Examples assessed included 3-bromo-1,1,1-trifluoroacetone (BTFA) [40], 2,2,2-trifluoroethanthiol (TFET) [41-42] and S-ethyl-trifluorothioacetate (SETFA) [25]. Primary amines were also investigated as sites for ^{19}F labelling, with reagents reported to target both the lysine side chains (p-nitro trifluoroacetate – PNTFA) [25] and the N-terminus of the protein (N-succinimidyl trifluoroacetate - NSTFA) [43]. In order to assess the efficiency of the ^{19}F labelling process, each potential ^{19}F tag was initially tested using BSA as a protein substrate, followed by collection of

1D ^{19}F NMR spectra. ^{1}H spectra were also collected to check for any denaturation or other adverse effects on the protein substrate from the labelling process.

Of the fluorine tags investigated that had previously been reported in the literature, BTFA was found to be the most effective, with labelling efficiencies of up to 200% (i.e. an average of two tags with trifluoromethyl groups covalently linked to each protein molecule after reaction). This resulted in good experimental NMR sensitivity – the experimental data for measurement of relaxation and diffusion could typically be collected in less than 60 minutes per sample. To obtain these efficiencies when labelling BSA and antibodies, it was however necessary to add 1 mM TCEP to the reaction mixture to prime some of the cysteine side chains for reaction with BFTA. All endogenous cysteine residues in the mAb samples investigated were involved in disulphide bridges and, while BSA has one free sulfhydryl group, labelling this without reduction also resulted in low efficiency. A similar approach of partial cysteine reduction is routinely used for preparing antibody-drug conjugates without adverse effects on mAb structure or function.[44-46] Five further ^{19}F tags (PNTFA, SETFA, NSTFA, TFASAN and TFET), previously described in the literature and detailed in Supporting Information (SI) Figure S1, were found to have lower, or in one case zero, labelling efficiency under reaction conditions that the proteins in our investigation could tolerate without denaturation or precipitation.

To increase the repertoire of available tags, two further molecules were investigated that, to our knowledge, have not previously been described for the purpose of ^{19}F labelling of proteins: TFBPD and TFCS, the structures of which are shown in Figure 1. These differ in chemical moieties adjacent to the CF$_3$ group, and so have different ^{19}F chemical shifts. TFBPD was designed de novo based on known maleimide – thiol reaction chemistry to target cysteine side chains, and was custom-synthesised by Charnwood Molecular Ltd.

TFCS is a commercially available protein cross-linking agent,[47-48] which was selected due to its known reactivity to the primary amines of lysine side chains under mild reaction conditions. When TFCS is used for cross-linking purposes, the trifluoroacetyl moiety serves as a protecting group, which is then normally removed in order to allow formation of the second covalent link to the other protein at a pH of 7.8 or higher. Maintaining a pH lower than 7.8 essentially prevents the second cross-linking reaction, and allows the ^{19}F group to remain covalently attached to a protein. The length of the alkyl chain provides consider-

able mobility to the ^{19}F group even when attached to a large slowly-tumbling protein, giving rise to a strong and relatively slowly relaxing NMR signal.

1-(4-(Trifluoromethyl)benzyl)-1H-pyrrole-2,5-dione
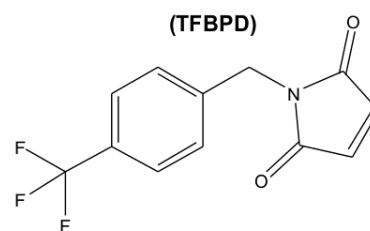
(TFBPD)

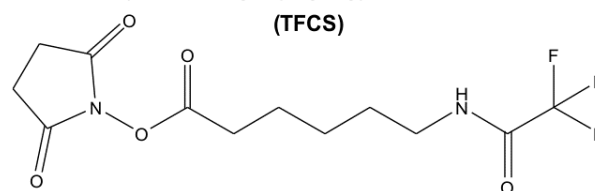N-(trifluoroacetylcaproyloxy) succinimide ester

(TFCS)

FIGURE 1. Structures of the TFBPD and TFCS tags used in this study for ^{19}F labelling. Structures of BTFA and other tags investigated are shown in SI Figure S1.

TFCS was successfully used to label BSA and both mAbs with efficiencies in excess of 100%, which was sufficient to collect high quality relaxation and diffusion data in 60 minutes per titration point. This tag has the advantages of milder reaction conditions (pH 7.2, room temperature) than most others previously investigated (pH >8.0 required for BTFA reaction, pH >9 for tags such as PNTFA), and so is likely to be tolerated by a wider range of proteins. Since attachment is via lysine side chains, this also avoids any need to reduce disulphide bridges. TFBPD was also found to be effective in labelling both BSA and mAbs but, since it targets cysteine sites, partial reduction with TCEP was still necessary for high efficiency antibody labelling. Overall labelling efficiencies of 160-200% were achieved, depending on the protein substrate, allowing collection of high quality ^{19}F NMR data with sufficiently diverse characteristic chemical shifts to distinguish them in mixtures.

Exploring mixtures of differentially-labelled BSA and CytC

With three high efficiency ^{19}F labels available, we were then able to proceed to testing the differential labelling strategy on model mixtures of proteins to confirm we could monitor the different components via their ^{19}F NMR signals. A system consisting of a mixture of BSA tagged

5

with BTFA (BTFA-BSA) and CytC tagged with TFBPD (TFBPD-CytC) was investigated first. Given that BSA has a pI of 4.7[49] while CytC has a pI of between 10 and 10.5,[50] it was expected that the two oppositely charged proteins would show some electrostatic attraction in solution at pH 7.2, potentially forming mixed transient complexes and clusters which could then be perturbed at higher ionic strengths. The apparent diffusion coefficients D_L measured for diluted (≤250 µM) CytC and BSA were 14.7×10^{-11} $\pm 0.05 \times 10^{-11}$ $m^2 s^{-1}$ and 6.07×10^{-11} $\pm 0.19 \times 10^{-11}$ $m^2 s^{-1}$, respectively, corresponding to R_{eff} of 1.65 ± 0.01 nm and 3.81 ± 0.12 nm, respectively. These values are very close to the values reported in the literature for CytC and BSA monomers.[51-54]
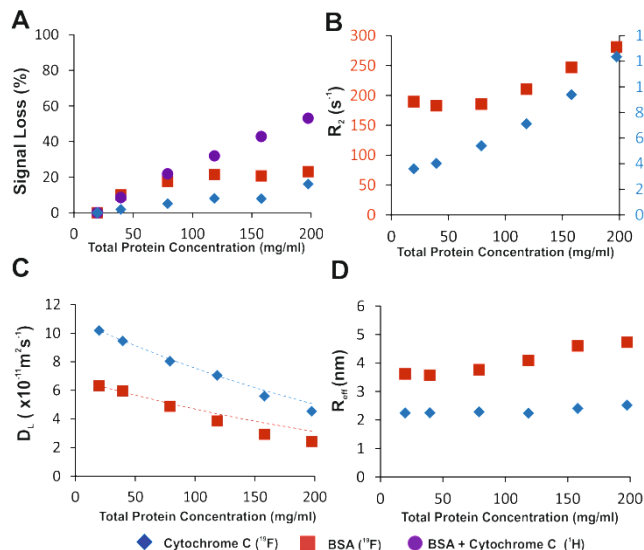

FIGURE 2. Concentration-dependent behaviour of an equimolar mixture of TFBPD-BSA and TFCS-CytC diluted from 2.5 mM to 250µM (197.5 mg/ml total protein concentration to 19.75 mg/ml). TFE (2 mM initial concentration) was added as internal reference to check dilution factor and calculate relative loss of intensity for signals from proteins. (A) Loss in expected signal intensity, based on extrapolated signal per unit protein from lowest concentration point. Data are shown in red for ^{19}F BSA, blue for ^{19}F CytC and purple for the combined ^{1}H envelope signals. (B) Transverse relaxation rates (R_2) measured for each protein in the mixture, with the scales for TFBPD-BSA and TFCS-CytC shown separately in red and blue, respectively, due to the difference in absolute relaxation rates for the two different tags. (C) Diffusion coefficients (D_L) for each protein in the mixture measured using DOSY. The expected changes in D_L based solely on the difference in molecular crowding at each concentration using the Medina-Noyola model [37-38] are shown as dashed lines for comparison. (D) R_{eff} derived from D_L, with correction for molecular crowding. Error bars on this and other figures were calculated in Dynamics Centre 2.4.8, using Monte Carlo

analysis and error propagation, but in all cases are smaller than the markers in this figure.

An equimolar mixture of TFBPD-CytC and BFTA-BSA at a total protein concentration of 197.5 mg/ml (corresponding to 2.5 mM of each protein) was prepared and then diluted stepwise, collecting NMR spectra at each point. TFE was added to the mixture as an inert reference component. All three components of the mixture are clearly resolved in the ^{19}F NMR spectrum, as shown in SI Figure S2, with TFBPD-CytC signals centred at -61.8 ppm, BTFA-BSA at -83.5 ppm and TFE at -75.4 ppm. For the dilution series, the concentration-dependent intensities of ^{19}F signals from each protein were measured separately. The overall ^{1}H signal envelope height for the mixture was also measured for comparison. For non-interacting molecules the signal intensity is expected to be proportional to concentration, with the observed values matching those extrapolated from the lowest concentration measurements where protein association is minimal. The deviation from the expected proportionality at higher concentrations, expressed as a "Signal Loss" (Figure 2A), is a signature of protein association and formation of larger "NMR-invisible" clusters.[55]

The observed ^{19}F signal intensities for both BSA and CytC at higher concentrations do not increase in proportion with the increasing concentration, which is reflected in increased signal losses shown in Figure 2A. Since the signal loss parameter is additionally normalized against the internal TFE reference intensity, this discrepancy cannot be explained by inaccuracies in the dilution process itself, and therefore reflects differential aggregation for BSA and CytC. For BSA the signal loss is saturated at around 20% at the higher concentrations, with CytC less affected with around 10% signal loss. In the ^{1}H spectrum the two proteins are not resolvable but the combined signal loss is much greater; almost 60% at the highest concentration. This suggests that at higher protein concentrations BSA and CytC increasingly enter the NMR-invisible "dark state".[56] Since the signal losses are greater in ^{1}H than the combined ^{19}F losses, it appears that ^{19}F is still able to detect some fraction of the larger protein cluster species which are lost in ^{1}H spectra. The transverse relaxation rates R_2 generally increase for both BSA and CytC with concentration (Figure 2B), however BSA shows bi-phasic behavior, with relaxation rates staying largely flat up to around 80 mg/ml, and linearly increasing afterwards. Both BSA and CytC show significant decreases in diffusion rates D_L as concentration increases (Fig. 2C), however most of this decrease can be explained by in-

crease in the molecular crowding which slows down translational diffusion due to excluded volume effects.[57-58] Once this crowding is taken into account the parameter R_{eff} (Figure 2D) reports on the residual effect - the change in the apparent size of NMR-visible clusters involving CytC and BSA, depending on concentration.

In low concentration equimolar BSA/CytC mixtures the diffusion rate of BSA indicates that it is monomeric. At higher concentrations there is a moderate increase in R_{eff}, suggesting an increasing fraction of the protein is forming dimers or higher oliogmers under these conditions. CytC diffuses with a rate indicating the presence of a significant dimer population even at the lowest point of the BSA/CytC dilution series, with only a very marginal further increase in R_{eff} at higher concentrations. Analysis of CytC in isolation (data shown in SI Figure S3) shows similar evidence of dimer formation at concentrations above 5 mg/ml, with CytC only matching the expected diffusion rate for the pure monomer at very low concentrations. BSA in isolation shows a gradual linear increase in R_{eff} at concentrations up to at least 200 mg/ml (SI Figure S4), consistent with a gradual increase in the size of the BSA clusters. This linear increasing trend is, however, different from the sigmoidal behavior seen in the concentration-dependence of the equimolar BSA/CytC mixture.

Overall the data reveal a complex concentration-dependent behavior, even in an ostensibly simple system. Contrary to expectations, the oppositely-charged BSA and CytC do not appear to form heterodimers even at high concentrations. If a BSA-CytC heterodimer were to form, it would represent a very substantial change in effective size for the tagged CytC (from 12 kDa to 79 kDa, assuming a 1:1 complex), which would dramatically alter the diffusion rate. Since no such change is seen, we can conclude that no significant heterodimer population is present. All changes observed appear to be from self-association of the proteins which is, however, modulated by the presence of a second protein, judging by the altered shape of the concentration dependencies. Despite being considered as one of the most soluble proteins, up to 20% of the BSA ^{19}F signal intensity is lost at high concentration, likely due to formation of large soluble clusters invisible to NMR. Losses for ^{1}H signals are even larger, reaching >80% for concentrations over 300 mg/ml (SI Figure S4).

To investigate if the observed self-association of CytC and BSA at higher concentrations is electrostatic in nature, NaCl was titrated into the 2.5 mM equimolar mixture of BSA and CytC (Figure 3). The reduction in cryoprobe sensitivity upon addition of salt is not a contrib-

uting factor to the signal loss parameter presented, due to the normalization against the internal TFE reference. Due to slight sample dilution upon addition of NaCl stock, a small decrease in overall molecular crowding was observed, which was taken into account while calculating R_{eff}. A slight increase in diffusion rate above that expected from dilution is still seen for both proteins with increasing NaCl concentration up to 150 mM. A decrease in R_2 is seen for BSA, matching the reduction in the apparent size of its clusters R_{eff}. R_2 for CytC is increased however, most likely due to change in the chemical exchange regime. The size of the clusters both for BSA and CytC reach a minimum at 150 mM NaCl (Figure 3D).
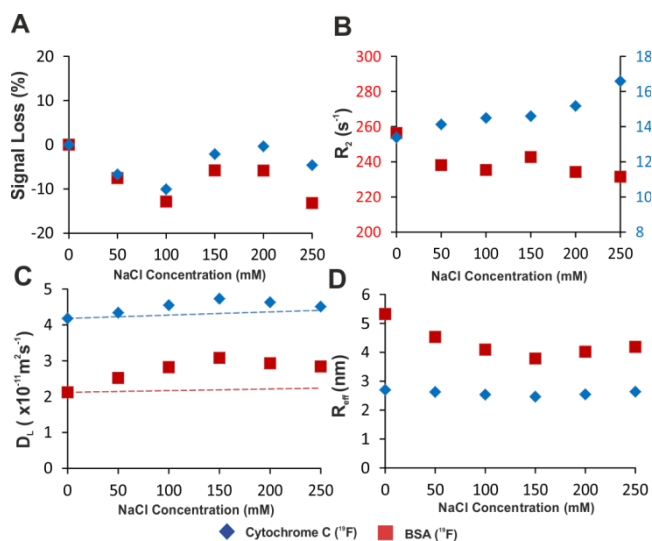

FIGURE 3. Behavior of 2.5 mM equimolar mixture of TFBPD-BSA and TFCS-CytC (total protein concentration 197.5 mg/ml) at different ionic strengths. (A) Relative loss in signal intensity, compared to expected values. Data is shown in red for ^{19}F BSA, blue for [19]F CytC. (B) Transverse relaxation rates R_2, with the scales for TFBPD-BSA and TFCS-CytC shown separately in red and blue, respectively. (C) Diffusion Coefficient (D_L) for each protein. The expected changes in D_L based solely on the difference in molecular crowding (due to the minor dilution effect of 5M NaCl stock addition) are shown as dashed lines for comparison. (D) R_{eff} derived from D_L, with correction for molecular crowding.

Together, the experiments with BSA and CytC mixtures revealed an underlying complex behavior of these proteins when mixed together, which is dependent on their total concentrations and ionic strength. Contrary to expectations, these overall oppositely charged molecules did not form equimolar hetero-protein complexes, even at low ionic strength. Instead BSA shows signs of concentration-dependent self-association, which is modulated by

the presence of CytC and somewhat disrupted by addition of NaCl, revealing there is an electrostatic component to this interaction. CytC appears to have a significant dimer population, but NaCl has little if any disruptive effect on this. Since there is loss of both ^{19}F and ^{1}H signal intensities at higher concentrations (to differing degrees), we can deduce that fractions of the proteins exist as large soluble clusters which are invisible to NMR. Application of ^{19}F NMR spectroscopy therefore uncovers the distinct concentration-dependent and condition-dependent behavior of the two differentially-labelled proteins in their co-formulations, while allowing us to track them independently.

Tracking individual mAbs in differentially-labelled mixtures

To demonstrate the utility of this approach for monitoring mixtures of even larger proteins, we tested a mixture of two ^{19}F labelled IgG mAbs, mAb2 labelled with BTFA and COE-19 labelled with TFCS (NB: labelling of COE-19 also worked well using TFBPD with 170% efficiency). These two antibodies have very similar molecular masses but display significant differences in their behavior. MAb2 is much more soluble whereas COE-19 is known to self-associate and is notably very viscous when handled at high concentrations. NMR data was collected on an equimolar mixture of the two mAbs at a range of concentrations, from 20 mg/ml to 200 mg/ml total (i.e. from 10 mg/ml to 100 mg/ml of each mAb). These experiments were conducted at 313K, as the increased tumbling rate and reduced water viscosity significantly improve the NMR spectrum quality, and mAbs can tolerate this temperature well.[34-35, 59] The differentially tagged mAbs give two clearly separated clusters of peaks, which can easily be tracked throughout the dilution series as shown in SI Figure S5.

As shown in Figure 4, at higher concentrations COE-19 ^{19}F signal losses are much greater than for mAb2, suggesting that more of the COE-19 is becoming part of larger soluble clusters invisible to NMR. In the ^{1}H spectra more than 60% of the overall signal intensity for the combined proteins is lost, revealing a significant population in the "dark state". Both mAbs also show greater slowdown in diffusion rate at higher concentration than can be explained by the increase in molecular crowding, again with COE-19 significantly more affected than mAb2. The transverse relaxation rate R_2 also displays differences in behavior for the two mAbs, with more than a 5-fold increase for COE-19, vs 1.3-fold for mAb2. For large, slowly-tumbling

proteins in the absence of chemical exchange R_2 is dependent on the size of the protein cluster (see Equation 5), and these R_2 trends match increases in R_{eff} derived from D_L, with COE-19 showing a dramatic increase in the apparent radii of observable clusters. These experiments clearly reveal differential behavior of mAb2 and COE-19 in the mixture.
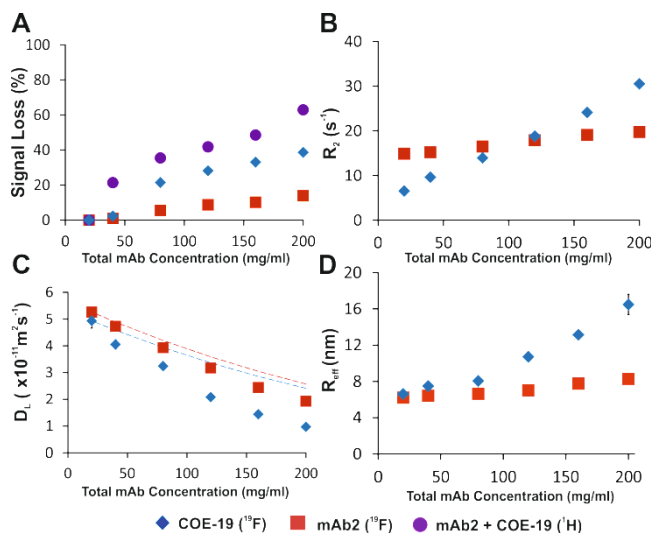

FIGURE 4. Behaviour of an equimolar mixture of BTFA-mAb2 and TFCS-COE-19 over a range of concentrations. (A) Loss in expected signal intensity, based on extrapolated signal per unit protein from lowest concentration point. Data is shown in red for ^{19}F mAb2, blue for ^{19}F COE-19 and purple for the combined ^{1}H signals. (B) Transverse relaxation rates (R_2) measured for each protein in the mixture. (C) Diffusion coefficients (D_L) for each protein in the mixture measured using DOSY. The expected changes in D_L based solely on the difference in molecular crowding at each concentration are shown as dashed lines for comparison. (D) R_{eff} derived from D_L, with correction for molecular crowding.

As a control the experiment was repeated for each mAb in isolation, to confirm whether the observed effects were from cross-interaction between mAb2 and COE-19, or from self-interaction by one or both of them. Reaction buffer, internal TFE reference, and titration conditions throughout were the same as in the admixture experiment. NMR signal intensities, diffusion coefficients and R_2 rates were measured for two proteins in separate samples, but overlaid on the same graph, as shown in Figure 5.

The general trends for isolated mAbs at concentrations up to 100 mg/ml appear similar as in the mixture (Figure 5). While for mAb2 the ^{19}F signal losses are as small as in the mixture, for COE-19 these losses are significantly less (a maximum of 17%) than in the mixture (39%). In ^{1}H the signal loss is even greater, 86% at maximum concentra-

tion. This suggests that in isolation the vast majority of COE-19 exists as larger ¹H-NMR-invisible soluble clusters. The relaxation rate R_2 and effective cluster radius R_{eff} for mAb2 increase to the same extent as in the mixture, however for isolated COE-19 the correspondent increases are less pronounced, suggesting that in the mixture COE-19 behavior is significantly influenced by the presence of mAb2. MAb2 promotes formation of smaller clusters of COE-19 while reducing the population of very large NMR-invisible clusters.
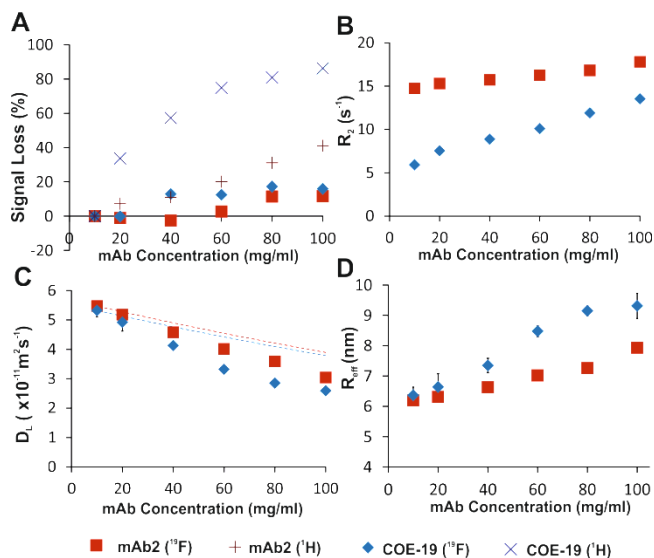

FIGURE 5. Comparison of relaxation and diffusion data for BTFA-mAb2 and TFCS-COE-19 in isolation over a range of concentrations. Each mAb was diluted from 100 mg/ml to 10 mg/ml in separate samples, with data shown as overlay. (A) Loss in expected signal intensity, based on extrapolated signal per unit protein from lowest concentration point. Data is shown in red for ¹⁹F mAb2, dark red (+) for ¹H mAb2, blue for ¹⁹F COE-19 and dark blue (x) for ¹H COE-19. (B) Transverse relaxation rates (R_2) measured for each protein in the mixture. (C) Diffusion coefficients (D_L) for each protein in the mixture measured using DOSY. The expected changes in D_L based solely on the difference in molecular crowding at each concentration are shown as dashed lines for comparison. (D) R_{eff} derived from D_L, with correction for molecular crowding.

To explore the effect of mAb2 on COE-19 clustering at even higher concentrations, in the following experiment the COE-19 concentration was held constant at 200 mg/ml, while the mAb2 concentration was raised from zero to 200 mg/ml (Figure 6). This titration used a two tube method (as further explained in the Methods section) with the two halves of the titrations meeting in the middle, with a final concentration of 100 mg/ml and 200 mg/ml of mAb2 and COE-19, respectively.
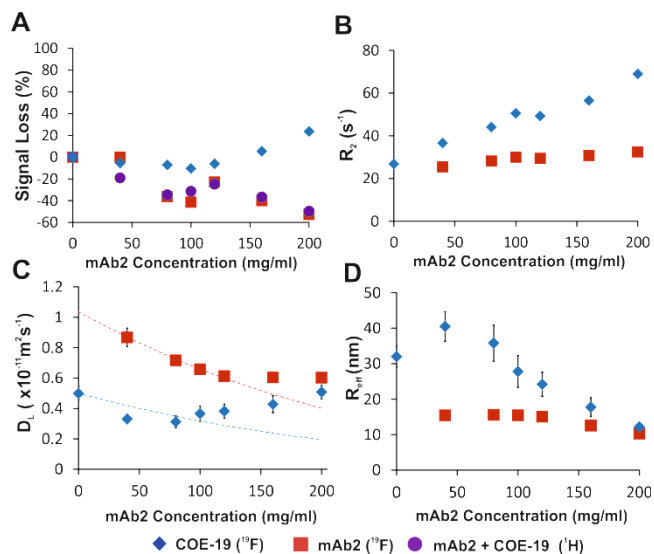

FIGURE 6. Titration of 200 mg/ml TFCS-COE-19 with BTFA-mAb2, from zero to 200 mg/ml. (A) Loss in expected signal intensity, based on extrapolated signal per unit protein from lowest concentration point. In this case at many points a signal gain was observed when mAb2 was added. Data is shown in red for ¹⁹F mAb2, blue for ¹⁹F COE-19 and purple for the combined ¹H signals. (B) Transverse relaxation rates (R_2) measured for each protein in the mixture. (C) Diffusion coefficients (D_L) for each protein in the mixture measured using DOSY. The expected changes in D_L based solely on the difference in molecular crowding at each concentration are shown as dashed lines for comparison. (D) R_{eff} derived from D_L, with correction for molecular crowding.

Interestingly, as the mAb2 concentration is raised, the combined ¹H signal loss decreases by up to 50%, suggesting that addition of mAb2 increases the total ¹H NMR-visible population of antibodies. ¹⁹F signal loss for COE-19 also slightly decreases as mAb2 is added up to 100 mg/ml, before increasing again at higher concentrations, even though the concentration of COE-19 remains constant throughout. For mAb2 the ¹⁹F signal appears to be lost to the "dark state" more at lower concentration, when COE-19 is in relative excess, whereas at high concentration the relative population of mAb2 in invisible states decreases. This indicates that the initial low concentrations of mAb2 are bound up in the large clusters with COE-19, but at higher mAb2 concentrations the composition of these clusters change, with a greater fraction of mAb2 not involved in them. This observation is supplemented by the dependencies shown by D_L and R_{eff} as mAb2 concentration is increased. The diffusion coefficients show an initial decline for both mAbs, but then plateau in the latter half of the titration for mAb2, and increase for COE-19. This

clearly goes against the expectation that an overall increase in molecular crowding should decrease D_L throughout the experiment. Once the effect of crowding is taken into account, the effective radius (R_{eff}) of the observable COE-19 protein clusters is actually decreasing upon addition of mAb2 for most of the titration. MAb2 at lower concentrations (0-100 mg/ml) in the presence of excess COE-19 forms oligomeric clusters with R_{eff} of ca. 15 nm, which reduce to around 10 nm when mAb2 and COE-19 become equimolar at very high concentrations (200 mg/ml each). The values of R_{eff} for COE-19 at 200 mg/ml are very large (>30 nm) when mAb2 is absent or only present at low concentration. The effective radius significantly reduces as the ratio of components approaches equimolarity, with the values of R_{eff} for COE-19 and mAb2 merging at the highest concentrations. Together this data shows that at extremely high concentrations (ca 400 mg/ml total) the addition of equimolar mAb2 to otherwise highly-self-associating COE-19 significantly decreases the apparent size of COE-19 clusters, forming smaller mixed mAb2/COE-19 clusters instead. The analysis of relaxation rates R_2 shows a marginal increase for mAb2, with a much greater increase for COE-19, suggesting that the chemical exchange increases, probably due to more fluid clusters and more transient interactions when proteins are equimolar at high concentrations. The ability to make such a highly-concentrated solution of COE-19, which self-associates even at low concentration, by using another protein (mAb2), supports the general idea of using protein mixtures to increase solubility of its individual components.
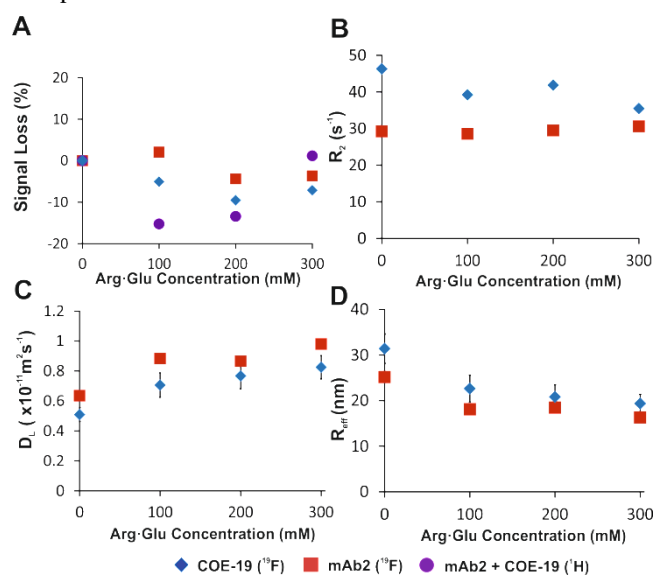
in expected signal intensity, based on extrapolated signal per unit protein from lowest concentration point. For some points signal gain was observed instead of loss. Data is shown in red for ^{19}F mAb2, blue for ^{19}F COE-19 and purple for the combined ^{1}H signals. (B) Transverse relaxation rates (R_2) measured for each protein in the mixture. (C) Diffusion coefficients (D_L) for each protein in the mixture measured using DOSY. At constant protein concentration no changes in molecular crowding occurred in this experiment. (D) R_{eff} derived from D_L.

To investigate how the addition of excipients, such as arginine glutamate (Arg·Glu)[34, 60], would affect the behavior and interactions of the mixture of 100 mg/ml mAb2 and 200 mg/ml COE-19, Arg·Glu was titrated into the sample (Figure 7). Addition of Arg·Glu resulted in a slight increase in combined mAb ^{1}H signal intensity and drop in "signal loss", implying a small reduction in the population of ^{1}H-invisible mAb species. The ^{19}F signal loss decreased by ca 10% and 5% for COE-19 and mAb2, respectively, when 200 mM Arg·Glu was added. The relaxation rates R_2 stayed relatively constant for mAb2 but showed a slight decreasing trend for COE-19. Arg·Glu noticeably increased the apparent diffusion coefficients D_L of both mAb2 and COE-19, which (at fixed protein crowding) manifests as a significant reduction of the apparent effective radii R_{eff} for both proteins. As addition of Arg·Glu itself causes a noticeable increase in the underlying buffer viscosity [35], the true decrease in the size of the protein clusters should be even more prominent. This experiment shows that at these extremely high protein concentrations Arg·Glu is noticeably reducing the effective size of the clusters, and that such an excipient addition effect can be monitored using the differential ^{19}F labelling approach. Altogether, the examples studied show a complex behavior of the mixed transiently-interacting protein systems, depending on the concentration, ratio of components, ionic strength and presence of excipients such as Arg·Glu. The ability of ^{19}F differential labelling and NMR spectroscopy to monitor such species separately makes possible to study and optimize complex protein mixtures.

FIGURE 7. Effects of addition of Arg·Glu to a mixture of 100 mg/ml BTFA-mAb2 and 200 mg/ml TFCS -COE-19. (A) Loss

DISCUSSION

Chemical Modification of mAbs with ^{19}F labels

Here we have established reliable methods for labelling proteins, and specifically mAbs, with ^{19}F using BTFA, TFBPD, and TFCS tags. These methods use mild reaction conditions, and label efficiently enough for rapid collection of ^{19}F NMR data. Since all three have distinct ^{19}F chemical shifts (-82.3 ppm, -62.0 ppm, and -75.4 ppm respectively) this already allows mixtures with up to three

components to be investigated by this differential label-ling method. The collection of available tags could be expanded further by using modified versions of these tags retaining the reactive group for covalent attachment, but modifying the ^{19}F containing moiety or neighboring chemical groups to produce other tags with similar reactivity, but a variety of characteristic chemical shifts.

In our experiments, when using relatively small size tags for chemical modifications, we have not observed any alterations to the protein behavior in solution, compared to the unlabeled form. For example, mAb2 when analyzed by SEC-MALS was found to show only negligible changes in apparent molecular mass and apparent hydrodynamic radius on BTFA labelling (R_h = 5.59 nm and 5.69 nm, and M_w = 147 kDa and 150 kDa for unlabeled and labelled, respectively, with estimated error margins of 0.45% for R_h values, 0.5% for M_w values). Extrinsic chemical labelling of proteins, e.g. with fluorescent dyes, is used, although typically at lower concentrations, in other quantitative protein binding characterization techniques such as microscale thermophoresis.[61] In some protein systems extensive labelling may in principle change association properties of proteins, therefore control experiments comparing modified *vs* non-modified variants would be advised.

We have demonstrated that a number of measurable NMR parameters such as ^1H, ^{19}F signal intensities, ^{19}F-derived diffusion coefficients (D_L), and transverse relaxation rates (R_2) can be recorded within reasonable experimental times even for large proteins such as ~150 kDa mAbs, and at total sample concentrations up to 400 mg/ml. Importantly, all of these parameters are sensitive to the changes in solution conditions and report on the apparent sizes of observable protein species. In addition this allows estimation of the amount of invisible "dark state" species which do not contribute to observable spectra but are still present in the sample, likely as very large soluble protein clusters. Although longitudinal relaxation rate (R_1) was also measured on the same samples, this parameter was not found to be informative as it did not change significantly with changing solution conditions, which agrees with other observations that T_1 is insensitive to transient interactions.[62] We also found this parameter largely dependent on the nature of the tag itself, with T_1 of 1.0 s for TFBPD, 1.1 s for TFCS and 0.4 s for BTFA tags attached to different proteins. Further example T_1 values are tabulated in SI Table S6. The relatively slow longitudinal relaxation of ^{19}F signals from the tags used allows efficient usage of stimulated echo- pulsed-field gradient experiments relying on $T_1 \gg T_2$ condition,[63] and capable of

recording diffusion of very slowly-diffusing molecules under extreme crowding environments. Transverse relaxation rate R_2 displayed far greater variation throughout the titrations, making it very useful for monitoring behavior of the protein components. The interpretation of this parameter may be however ambiguous as it depends on both the apparent size of protein and chemical exchange processes. Even for proteins the size of mAbs, the ^{19}F signal relaxation rate was sufficiently slow for straightforward measurement of R_2 using a CPMG experiment. Major differences in the absolute R_2 values were however observed for different ^{19}F tags, e.g. TFCS and BFTA, linked to the same protein substrate under the same conditions. This can be explained by differences in linker length between the protein attachment point and the trifluoromethyl group, resulting in differences in flexibility. The diffusion coefficient (D) provides the most direct method of detecting changes in the apparent size of the protein, via the Stokes-Einstein equation. In cases where the protein concentration is changing, it must be remembered that the change in molecular crowding will also alter the diffusion coefficient. This effect can be calculated using the Medina-Noyola model[37] and compensated for, allowing any residual effects from aggregation or protein-protein interactions to be observed, reflected in the value of apparent radius of protein cluster R_{eff}.

Finally, the basic intensities of the signals in the NMR spectra provide an indirect indicator of protein aggregation into larger clusters which, due to very slow molecular tumbling and extreme signal broadening, become invisible to NMR. As no precipitation is observed, these large clusters are soluble. The presence of such clusters can be inferred by measuring loss of signal intensity in ^1H and ^{19}F spectra relative to the expected values as extrapolated from dilute conditions. It is noticeable throughout these experiments than the ^{19}F spectra show significantly less signal loss than the ^1H spectra in samples with a high degree of oligomerization. This suggests fluorine NMR can continue to provide information on larger oligomers than proton NMR.

The NMR parameters measured for multiple titration experiments for BSA/CytC and mAb2/COE-19 mixtures generally reveal a complex behavior stemming from the balance between non-specific and transient self-association and hetero-association, depending on sample conditions. Overall, the possible species present in the sample, and relation to the measurable NMR parameters, are schematically represented on Figure 8. The sizes of monomers or lower-oligomers can be inferred from ^{19}F-

derived D_L and R_2 parameters, whereas the population size of presumably higher-oligomer species can be inferred from ${}^1H$ signal loss, and population of even larger higher-oligomer species from ${}^{19}F$ signal loss. Unfortunately the complexity of the system shown on Figure 8 and lack of sufficient number of measurable parameters does not at present allow quantitative fitting of equilibrium parameters characterizing exchange between these species. However, running experiments as titrations would allow optimization of the buffer, excipients and other sample parameters to achieve the required balance between self-association and hetero-association.
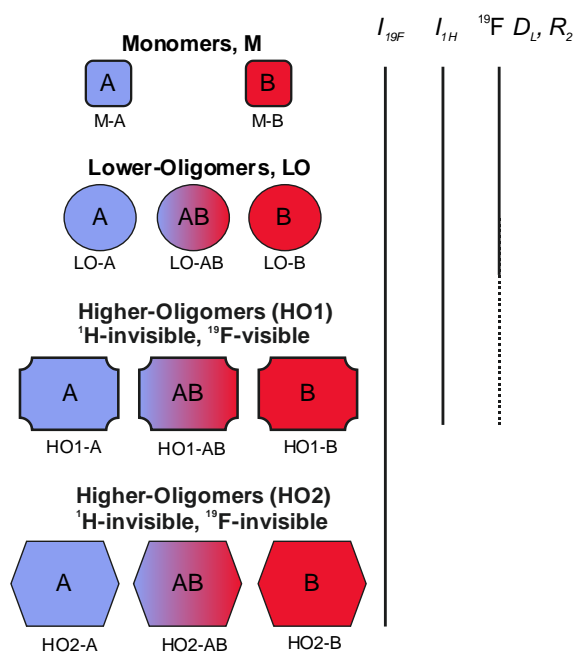

FIGURE 8. Schematic representation of hypothetical differentially-labelled protein species present in mixed samples. Proteins A and B can exist as monomers, or either homo-oligomers (resulting from self-association) or hetero-oligomers (resulting from hetero-association) of different size. Soluble assemblies of larger size (higher-oligomers) are not be visible by ${}^1H$ NMR yet still contribute to ${}^{19}F$ spectra, whereas very large soluble higher-oligomers may be invisible in both ${}^1H$ and ${}^{19}F$ spectra, due to extreme signal broadening. The change in the balance of different species in the course of titration series is assessed here by complementary NMR parameters. Deviation from the expected values of signal intensities I_{19F} and I_{1H} in ${}^{19}F$ and ${}^1H$ spectra, respectively, enables to monitor the balance between monomers or lower-oligomers and invisible species. Changes in the apparent diffusion coefficients D_L measured for ${}^{19}F$ covalently-attached tags allow us to follow the change in the apparent hydrodynamic radius of LO species, detecting early stages of M→LO transition for homo- and hetero-oligomers, whereas the transverse relaxation rate R_2 reflects both a change in molec-

ular size and involvement of protein in transient chemical exchange processes.

The titration experiments presented here for pairs of proteins which do not undergo specific interactions show that at high concentrations they do influence each other's behavior in solution through non-specific hetero-protein associations. A surprising observation is that mixing the self-associating COE-19 with mAb2 enabled us to achieve a total concentration of 400 mg/ml, albeit with the formation of soluble mAb2-COE-19 clusters, sufficiently small for the majority to remain observable by ${}^{19}F$ NMR. The ${}^{19}F$ differential labelling strategy may be useful for the optimization of protein co-formulations, but also could be applied in other natively-crowded systems reconstituted *in vitro*, such as those modelling phase-separation in membraneless organelles, the cell environment, and others.[64-65]

ASSOCIATED CONTENT

Supporting Information (SI)

Figure S1: Structures of ${}^{19}F$ reagents investigated for differential labelling. Figure S2: Example ${}^{19}F$ 1D spectrum of a differentially labelled protein mixture. Figure S3: Signal loss, relaxation and diffusion data for dilution of CytC. Figure S4: Signal loss, relaxation and diffusion data for dilution of BSA. Figure S5: 1D ${}^{19}F$ spectra over the course of a mAb mixture dilution series. Table S6: Example T_1 values for different protein-tag combinations

AUTHOR INFORMATION

Corresponding Author

A.Golovanov@manchester.ac.uk

ORCID 0000-0002-8592-3984

Notes

The authors declare no competing financial interest.

for the SEC-MALS analysis, and to Prof. Alan Dickson for general discussions.

ABBREVIATIONS

BSA: Bovine Serum Albumin

BTFA: 3-bromo-1,1,1-trifluoroacetone

CytC: Cytochrome C

DOSY: Diffusion Ordered Spectroscopy

IgG: Immunoglobulin G

mAb: Monoclonal Antibody

NMR: Nuclear Magnetic Resonance

NSTFA: N-succinimidyl trifluoroacetate

PNTFA: p-nitrophenyl trifluoroacetate

SAXS: Small Angle X-Ray Scattering

SEC-MALS: Size Exclusion Chromatography Multiangle Light Scattering

SETFA: S-ethyl-trifluorothioacetate

TFASAN: Trifluoroacetamidosuccinic anhydride

TFBPD: 1-(4-(Trifluoromethyl)benzyl)-1H-pyrrole-2,5-dione

TFCS: N-(ε-trifluoroacetylcaproyloxy) succinimide ester

TFET: 2,2,2-trifluoroethanthiol

REFERENCES

1.   Yearley, E. J.; Godfrin, P. D.; Perevozchikova, T.; Zhang, H.; Falus, P.; Porcar, L.; Nagao, M.; Curtis, J. E.; Gawande, P.; Taing, R.; Zarraga, I. E.; Wagner, N. J.; Liu, Y., Observation of small cluster formation in concentrated monoclonal antibody solutions and its implications to solution viscosity. *Biophys J* **2014**, *106* (8), 1763-70.

2.   Stradner, A.; Sedgwick, H.; Cardinaux, F.; Poon, W. C.; Egelhaaf, S. U.; Schurtenberger, P., Equilibrium cluster formation in concentrated protein solutions and colloids. *Nature* **2004**, *432* (7016), 492-5.

3.   McPherson, A.; Cudney, B., Optimization of crystallization conditions for biological macromolecules. *Acta Crystallogr F Struct Biol Commun* **2014**, *70* (Pt 11), 1445-67.

4.   Sugiki, T.; Yoshiura, C.; Kofuku, Y.; Ueda, T.; Shimada, I.; Takahashi, H., High-throughput screening of optimal solution conditions for structural biological studies by fluorescence correlation spectroscopy. *Protein Sci* **2009**, *18* (5), 1115-20.

5.   Shpilberg, O.; Jackisch, C., Subcutaneous administration of rituximab (MabThera) and trastuzumab (Herceptin) using hyaluronidase. *Br J Cancer* **2013**, *109* (6), 1556-61.

6.   Ecker, D. M.; Jones, S. D.; Levine, H. L., The therapeutic monoclonal antibody market. *MAbs* **2015**, *7* (1), 9-14.

7.   Elvin, J. G.; Couston, R. G.; van der Walle, C. F., Therapeutic antibodies: market considerations, disease targets and bioprocessing. *Int J Pharm* **2013**, *440* (1), 83-98.

8.   Hawe, A.; Friess, W., Formulation development for hydrophobic therapeutic proteins. *Pharm Dev Technol* **2007**, *12* (3), 223-37.

9.   Rotte, A.; Jin, J. Y.; Lemaire, V., Mechanistic overview of immune checkpoints to support the rational design of their combinations in cancer immunotherapy. *Ann Oncol* **2018**, *29* (1), 71-83.

10.   Santacoloma, P. A.; Sin, G.; Gernaey, K. V.; Woodley, J. M., Multienzyme-Catalyzed Processes: Next-Generation Biocatalysis. *Org Process Res Dev* **2011**, *15* (1), 203-212.

11.   You, C.; Zhang, Y. H. P., Biomanufacturing by in vitro biosystems containing complex enzyme mixtures. *Process Biochem* **2017**, *52*, 106-114.

12.   Ohki, S. Y.; Kainosho, M., Stable isotope labeling methods for protein NMR spectroscopy. *Prog Nucl Mag Res Sp* **2008**, *53* (4), 208-226.

13.   Verardi, R.; Traaseth, N. J.; Masterson, L. R.; Vostrikov, V. V.; Veglia, G., Isotope labeling for solution and solid-state NMR spectroscopy of membrane proteins. *Adv Exp Med Biol* **2012**, *992*, 35-62.

14.   Arntson, K. E.; Pomerantz, W. C., Protein-Observed Fluorine NMR: A Bioorthogonal Approach for Small Molecule Discovery. *J Med Chem* **2016**, *59* (11), 5158-71.

15.   Golovanov, A. P.; Blankley, R. T.; Avis, J. M.; Bermel, W., Isotopically discriminated NMR spectroscopy: a tool for investigating complex protein interactions in vitro. *J Am Chem Soc* **2007**, *129* (20), 6528-35.

16.   Bermel, W.; Tkach, E. N.; Sobol, A. G.; Golovanov, A. P., Simultaneous measurement of residual dipolar couplings for proteins in complex using the isotopically discriminated NMR approach. *J Am Chem Soc* **2009**, *131* (24), 8564-70.

17.   Masterson, L. R.; Tonelli, M.; Markley, J. L.; Veglia, G., Simultaneous detection and deconvolution of congested NMR spectra containing three isotopically labeled species. *J Am Chem Soc* **2008**, *130* (25), 7818-9.

18.   Tunnicliffe, R. B.; Hautbergue, G. M.; Wilson, S. A.; Kalra, P.; Golovanov, A. P., Competitive and cooperative interactions mediate RNA transfer from herpesvirus saimiri ORF57 to the mammalian export adaptor ALYREF. *Plos Pathog* **2014**, *10* (2), e1003907.

19.   Hossler, P.; Khattak, S. F.; Li, Z. J., Optimal and consistent protein glycosylation in mammalian cell culture. *Glycobiology* **2009**, *19* (9), 936-49.

20.   Gerig, J. T., Fluorine Nmr of Proteins. *Prog Nucl Mag Res Sp* **1994**, *26*, 293-370.

21.   Kitevski-LeBlanc, J. L.; Prosser, R. S., Current applications of 19F NMR to studies of protein structure and dynamics. *Prog Nucl Magn Reson Spectrosc* **2012**, *62*, 1-33.

22.   Kim, T. H.; Chung, K. Y.; Manglik, A.; Hansen, A. L.; Dror, R. O.; Mildorf, T. J.; Shaw, D. E.; Kobilka, B. K.; Prosser, R. S., The role of ligands on the equilibria between functional states of a G protein-coupled receptor. *J Am Chem Soc* **2013**, *135* (25), 9465-74.

23.   Bouchard, M.; Pare, C.; Dutasta, J. P.; Chauvet, J. P.; Gicquaud, C.; Auger, M., Interaction between G-actin and vari-

ous types of liposomes: A 19F, 31P, and 2H nuclear magnetic resonance study. *Biochemistry* **1998,** *37* (9), 3149-55.

24.   Huestis, W. H.; Raftery, M. A., Bromotrifluoroacetone Alkylates Hemoglobin at Cysteine Beta-93. *Biochem Bioph Res Co* **1978,** *81* (3), 892-899.

25.   Adriaensens, P.; Box, M. E.; Martens, H. J.; Onkelinx, E.; Put, J.; Gelan, J., Investigation of Protein-Structure by Means of F-19-Nmr - a Study of Hen Egg-White Lysozyme. *Eur J Biochem* **1988,** *177* (2), 383-394.

26.   Imiolek, M.; Karunanithy, G.; Ng, W. L.; Baldwin, A. J.; Gouverneur, V.; Davis, B. G., Selective Radical Trifluoromethylation of Native Residues in Proteins. *J Am Chem Soc* **2018,** *140* (5), 1568-1571.

27.   Ichiishi, N.; Caldwell, J.; Lin, M.; Zhong, W.; Zhu, X.; Streckfuss, E.; Kim, H.; Parish, C.; Krska, S., Protecting group free radical C-H trifluoromethylation of peptides. *Chem Sci* **2018,** *Advance Article, DOI: 10.1039/C8SC00368H*.

28.   Li, C.; Lutz, E. A.; Slade, K. M.; Ruf, R. A.; Wang, G. F.; Pielak, G. J., 19F NMR studies of alpha-synuclein conformation and fibrillation. *Biochemistry* **2009,** *48* (36), 8578-84.

29.   Wang, G. F.; Li, C.; Pielak, G. J., Probing the micelle-bound aggregation-prone state of alpha-synuclein with (19)F NMR spectroscopy. *Chembiochem* **2010,** *11* (14), 1993-6.

30.   Arseniev, A. S.; Utkin, Y. N.; Pashkov, V. S.; Tsetlin, V. I.; Ivanov, V. T.; Bystrov, V. F.; Ovchinnikov, Y. A., F-19 Nmr Determination of Intramolecular Distances in Spin-Labeled and Fluorine-Labeled Proteins - Neurotoxin-Ii Naja-Naja-Oxiana. *Febs Lett* **1981,** *136* (2), 269-274.

31.   Danielson, M. A.; Falke, J. J., Use of 19F NMR to probe protein structure and conformational changes. *Annu Rev Biophys Biomol Struct* **1996,** *25*, 163-95.

32.   Suzuki, Y.; Brender, J. R.; Soper, M. T.; Krishnamoorthy, J.; Zhou, Y.; Ruotolo, B. T.; Kotov, N. A.; Ramamoorthy, A.; Marsh, E. N., Resolution of oligomeric species during the aggregation of Abeta1-40 using (19)F NMR. *Biochemistry* **2013,** *52* (11), 1903-12.

33.   Urick, A. K.; Hawk, L. M.; Cassel, M. K.; Mishra, N. K.; Liu, S.; Adhikari, N.; Zhang, W.; dos Santos, C. O.; Hall, J. L.; Pomerantz, W. C., Dual Screening of BPTF and Brd4 Using Protein-Observed Fluorine NMR Uncovers New Bromodomain Probe Molecules. *ACS Chem Biol* **2015,** *10* (10), 2246-56.

34.   Kheddo, P.; Tracka, M.; Armer, J.; Dearman, R. J.; Uddin, S.; van der Walle, C. F.; Golovanov, A. P., The effect of arginine glutamate on the stability of monoclonal antibodies in solution. *Int J Pharm* **2014,** *473* (1-2), 126-33.

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

36.   Aguilar, J. A.; Nilsson, M.; Bodenhausen, G.; Morris, G. A., Spin echo NMR spectra without J modulation. *Chem Commun* **2012,** *48* (6), 811-813.

37.   Medina-Noyola, M., Long-time self-diffusion in concentrated colloidal dispersions. *Phys Rev Lett* **1988,** *60* (26), 2705-2708.

38.   Vanblaaderen, A.; Peetermans, J.; Maret, G.; Dhont, J. K. G., Long-Time Self-Diffusion of Spherical Colloidal Particles Measured with Fluorescence Recovery after Photobleaching. *J Chem Phys* **1992,** *96* (6), 4591-4603.

39.   Andersson, K. M.; Hovmoller, S., The average atomic volume and density of proteins. *Z Kristallogr* **1998,** *213* (7-8), 369-373.

40.   Luchette, P. A.; Prosser, R. S.; Sanders, C. R., Oxygen as a paramagnetic probe of membrane protein structure by cysteine mutagenesis and F-19 NMR spectroscopy. *J Am Chem Soc* **2002,** *124* (8), 1778-1781.

41.   Klein-Seetharaman, J.; Getmanova, E. V.; Loewen, M. C.; Reeves, P. J.; Khorana, H. G., NMR spectroscopy in studies of light-induced structural changes in mammalian rhodopsin: Applicability of solution F-19 NMR. *P Natl Acad Sci USA* **1999,** *96* (24), 13744-13749.

42.   Loewen, M. C.; Klein-Seetharaman, J.; Getmanova, E. V.; Reeves, P. J.; Schwalbe, H.; Khorana, H. G., Solution 19F nuclear Overhauser effects in structural studies of the cytoplasmic domain of mammalian rhodopsin. *Proc Natl Acad Sci U S A* **2001,** *98* (9), 4888-92.

43.   Garg, P. K.; Garg, S.; Zalutsky, M. R., F-18 Labeling of Monoclonal-Antibodies and Fragments with Preservation of Immunoreactivity. *Bioconjugate Chem* **1991,** *2* (1), 44-49.

44.   Sun, M. M.; Beam, K. S.; Cerveny, C. G.; Hamblett, K. J.; Blackmore, R. S.; Torgov, M. Y.; Handley, F. G.; Ihle, N. C.; Senter, P. D.; Alley, S. C., Reduction-alkylation strategies for the modification of specific monoclonal antibody disulfides. *Bioconjug Chem* **2005,** *16* (5), 1282-90.

45.   Wiggins, B.; Liu-Shin, L.; Yamaguchi, H.; Ratnaswamy, G., Characterization of cysteine-linked conjugation profiles of immunoglobulin G1 and immunoglobulin G2 antibody-drug conjugates. *J Pharm Sci* **2015,** *104* (4), 1362-72.

46.   Lyon, R. P.; Meyer, D. L.; Setter, J. R.; Senter, P. D., Conjugation of anticancer drugs through endogenous monoclonal antibody cysteine residues. *Methods Enzymol* **2012,** *502*, 123-38.

47.   Ranganathan, A.; Gee, S. J.; Hammock, B. D., An immunoassay for the detection of triclosan-O-glucuronide, a primary human urinary metabolite of triclosan. *Anal Bioanal Chem* **2015,** *407* (24), 7263-73.

48.   Ding, Y. H.; Fan, S. B.; Li, S.; Feng, B. Y.; Gao, N.; Ye, K.; He, S. M.; Dong, M. Q., Increasing the Depth of Mass-Spectrometry-Based Structural Analysis of Protein Complexes through the Use of Multiple Cross-Linkers. *Anal Chem* **2016,** *88* (8), 4461-9.

49.   Ge, S.; Kojio, K.; Takahara, A.; Kajiyama, T., Bovine serum albumin adsorption onto immobilized organotrichlorosilane surface: influence of the phase separation on protein adsorption patterns. *J Biomater Sci Polym Ed* **1998,** *9* (2), 131-50.

50.   Vangelder, B. F.; Slater, E. C., Extinction Coefficient of Cytochrome C. *Biochim Biophys Acta* **1962,** *58* (3), 593-&.

51.   Raj, T.; Flygare, W. H., Diffusion Studies of Bovine Serum-Albumin by Quasi-Elastic Light-Scattering. *Biochemistry* **1974,** *13* (16), 3336-3340.

52.   Axelsson, I., Characterization of Proteins and Other Macromolecules by Agarose-Gel Chromatography. *J Chromatogr* **1978,** *152* (1), 21-32.

53.   Kontturi, A. K.; Kontturi, K.; Niinikoski, P.; Savonen, A.; Vuoristo, M., The effective charge number and diffusion coefficient of cationic cytochrome c in aqueous solution. *Acta Chem Scand* **1992,** *46* (4), 348-53.

54.   Wilkins, D. K.; Grimshaw, S. B.; Receveur, V.; Dobson, C. M.; Jones, J. A.; Smith, L. J., Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* **1999,** *38* (50), 16424-31.

55.   Kheddo, P.; Bramham, J. E.; Dearman, R. J.; Uddin, S.; van der Walle, C. F.; Golovanov, A. P., Investigating Liquid-Liquid Phase Separation of a Monoclonal Antibody Using Solution-State NMR Spectroscopy: Effect of Arg.Glu and Arg.HCl. *Mol Pharm* **2017,** *14* (8), 2852-2860.

56.   Anthis, N. J.; Clore, G. M., Visualizing transient dark states by NMR spectroscopy. *Q Rev Biophys* **2015,** *48* (1), 35-116.

57.   Bernado, P.; Garcia de la Torre, J.; Pons, M., Macromolecular crowding in biological systems: hydrodynamics and NMR methods. *J Mol Recognit* **2004,** *17* (5), 397-407.

58.   Roos, M.; Link, S.; Balbach, J.; Krushelnitsky, A.; Saalwachter, K., NMR-detected brownian dynamics of alphaB-crystallin over a wide range of concentrations. *Biophys J* **2015,** *108* (1), 98-106.

59.   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.

60.   Golovanov, A. P.; Hautbergue, G. M.; Wilson, S. A.; Lian, L. Y., A simple method for improving protein solubility and long-term stability. *J Am Chem Soc* **2004,** *126* (29), 8933-9.

61.   Seidel, S. A. I.; Dijkman, P. M.; Lea, W. A.; van den Bogaart, G.; Jerabek-Willemsen, M.; Lazic, A.; Joseph, J. S.; Srinivasan, P.; Baaske, P.; Simeonov, A.; Katritch, I.; Melo, F. A.; Ladbury, J. E.; Schreiber, G.; Watts, A.; Braun, D.; Duhr, S., Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. *Methods* **2013,** *59* (3), 301-315.

62.   Ye, Y.; Wu, Q.; Zheng, W.; Jiang, B.; Pielak, G. J.; Liu, M.; Li, C., Quantification of size effect on protein rotational mobility in cells by (19)F NMR spectroscopy. *Anal Bioanal Chem* **2018,** *410* (3), 869-874.

63.   Tanner, J. E., Use of Stimulated Echo in NMR-Diffusion Studies. *J Chem Phys* **1970,** *52* (5), 2523-&.

64.   Mitrea, D. M.; Kriwacki, R. W., Phase separation in biology; functional organization of a higher order. *Cell Commun Signal* **2016,** *14*, 1.

65.   Chong, P. A.; Forman-Kay, J. D., Liquid-liquid phase separation in cellular signaling systems. *Curr Opin Struct Biol* **2016,** *41*, 180-186.

For Table of Contents Only.