

Selective Suppression of Excipient Signals in 2D ^1H - ^{13}C Methyl Spectra of Biopharmaceutical Products

Luke W. Arbogast¹, Frank Delaglio¹, Joel R. Tolman² and John P. Marino¹

1. National Institute of Standards and Technology, Institute for Bioscience and Biotechnology Research, 9600 Gudelsky Dr. Rockville, Md. 20850
2. Johns Hopkins University, Department of Chemistry, 3400 N. Charles St. Baltimore, Md. 21218

Contact: luke.arbogast@nist.gov ORCID, 0000-0001-6871-450X

Abstract

While the use of ^1H - ^{13}C methyl correlated NMR spectroscopy at natural isotopic abundance has been demonstrated as feasible on protein therapeutics as large as monoclonal antibodies, spectral interference from aliphatic excipients remains a significant obstacle to its widespread application. These signals can cause large baseline artifacts, obscure protein resonances, and cause dynamic range suppression of weak peaks in NUS applications, thus hampering both traditional peak-based spectral analyses as well as emerging chemometric methods of analysis. Here we detail modifications to the 2D ^1H - ^{13}C gradient-selected HSQC experiment that make use of selective pulsing techniques for targeted removal of interfering excipient signals in spectra of the NISTmAb prepared in several different formulations. This approach is demonstrated to selectively reduce interfering excipient signals while still yielding 2D spectra with only modest losses in protein signal. Furthermore, it is shown that spectral modeling based on the SMILE algorithm can be used to simulate and subtract any residual excipient signals and their attendant artifacts from the resulting 2D NMR spectra.

Keywords

biotherapeutics, excipients, selective cross-polarization, SMILE, SIERRA

Abbreviations

HOS = higher order structure

HSQC = heteronuclear single quantum coherence

RMS = root mean squared

SIERRA = selective excipient reduction and removal

SMILE = sparse multidimensional iterative lineshape enhancement

S/N = signal to noise

Introduction

Characterization of higher order structure (HOS) in protein therapeutics by two-dimensional (2D) NMR is receiving increased attention, due to the ability of 2D NMR to provide detailed structural information at atomic resolution (Aubin et al. 2008, 2010; Wishart 2013). Hardware and methodological developments now allow for acquisition of 2D ^1H - ^{13}C methyl spectra at natural isotopic abundance on protein therapeutics as large as monoclonal antibodies, with measurement times that are competitive with traditional HOS techniques (Arbogast et al. 2015, 2016; Singh et al. 2017). Indeed, recent multi-national interlaboratory studies have demonstrated interlaboratory precision, reproducibility, and reliability of 2D NMR applications to protein therapeutics, thus demonstrating its fit for purpose in the biopharmaceutical industry (Ghasriani et al. 2016; Brinson et al. 2017). However, applications of these techniques can be hampered by potential spectral interference from aliphatic excipients. This can be especially problematic for applications to formulation and stability development, where high resolution NMR could be leveraged to provide atomic level detail in screening applications and help elucidate molecular frameworks of protein-excipient interactions. Recent developments in chemometric analysis of 2D methyl spectra show promise for rapid and high resolution characterization of biotherapeutic spectral data, which would greatly aid this endeavor (Japelj et al. 2016; Chen et al. 2018; Arbogast et al. 2017), however, excipient signals, which are often orders of magnitude larger than that of the protein, can negatively influence chemometric analysis through introduction of baseline distortions or impacting the fidelity of picked peak parameters in the vicinity of the excipient signal.

One approach to circumvent this limitation is to substitute deuterated analogs for the true formulation components; however this is not practical as these compounds can be costly and often cannot be obtained commercially. Furthermore, for drug-product testing it is most desirable to analyze the product as manufactured and formulated, therefore buffer substitution is not ideal. Spectroscopic techniques to eliminate the excipient signals, such as gradient diffusion filtering, have been successfully employed to applications of one-dimensional (1D) proton NMR of protein therapeutics; however such techniques come at a cost in sensitivity due to attenuation of the protein signal (Pope et al. 2013; Franks et al. 2016). For 1D NMR applications, these losses in sensitivity can be compensated by increased signal averaging while still maintaining reasonable experimental durations, but for 2D NMR applications such losses cannot so easily be mitigated. However, an intrinsic advantage of the 2D NMR approach is dispersal of signals in two spectral dimensions, making it possible to reduce the spectral effects of formulation components through selective pulsing techniques that suppress signals specifically at desired ^1H - ^{13}C frequency positions. Here, we detail two modifications to a standard ^1H - ^{13}C gradient-selected, sensitivity-enhanced HSQC (gsHSQC) sequence that can be used to eliminate formulation signals with minimal loss to protein signals elsewhere in the spectrum. Furthermore, we demonstrate how the sparse multidimensional iterative lineshape enhancement (SMILE) method (Ying et al. 2017) for reconstruction of sparsely sampled data can be repurposed to further eliminate unwanted residual excipient signals. The 2D NMR methods are demonstrated on the NISTmAb (PS 8670) in various formulation conditions.

Materials and Methods

Certain commercial equipment, instruments, and materials are identified in this paper to specify the experimental procedure. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the material or equipment identified is necessarily the best available for the purpose.

Sample Preparation

All NMR samples were prepared from a stock solution of NISTmAb Primary Sample (PS) 8670, formulated at 100 mg/mL in 25 mmol/L L-histidine at pH 6.0. The 0.3% w/v polysorbate-80 and 50 mmol/L L-methionine samples, which remained in the L-histidine base buffer, were prepared by dilution of the PS-8670 stock solution to 40 mg/mL with addition of the secondary formulation component (i.e. polysorbate-80 or L-methionine)

from a 300 mmol/L stock solution in 25 mmol/L L-histidine at pH 6.0. For the sodium acetate sample, 120 mL of the PS 8670 stock solution was buffer exchanged into 10 mmol/L sodium acetate at pH 5.2 and concentrated to 300 μ L using a 100 kDa centrifugal filter. The sucrose sample was similarly prepared from a solution of 10 mmol/L sodium acetate buffer at pH 5.2 containing 10% w/v sucrose. All samples contained 3% D₂O.

NMR Data Acquisition and Processing

Unless otherwise noted in the text, all data was collected at 50°C on a Bruker Avance III 900 MHz spectrometer equipped with a z-axis gradient filter and a triple-resonance cryogenically cooled probe. Typical ¹H-¹³C gsHSQC datasets were collected with 64 scans per transient, a recycle delay of 1.5 s with 1262x60 (46) complex points, corresponding to acquisition times of 100 ms and 10 ms over spectral widths of 14 ppm and 26.5 (20) ppm for standard (methyl-selective) experiments in the F2 and F1 domains with an experimental duration of 208 (160) min. Data were processed using NMRPipe (Delaglio et al. 1995) with a shifted sine-squared bell apodization and zero filling in both dimensions prior to Fourier Transform. Signal to noise ratios (S/N) were determined from the root mean squared (RMS) amplitudes of signal and noise spectral intensity vectors as previously described (Arbogast et al. 2016). Principal component analysis of NMR spectra was performed as previously described (Arbogast et al. 2017) using NMRPipe and scatter plots visualized using Matlab R2015B.

Optimization of SIERRA filter

For each targeted excipient signal, the ¹H/¹³C frequency offsets and ¹J_{HC} coupling constant were determined from a 2D gsHSQC sequence run with 4 scans and without ¹³C decoupling during acquisition. From the measured coupling constant, SIERRA-filter parameters were estimated with τ_{cp} set according to eq. 3 and ¹H and ¹³C CP power levels estimated from the respective 90 ° hard pulse width (pw_{90}) values according to eq. 2 and

$$\Delta pwr(db) = 20 \log(4 \times pw_{90} \times \omega_{cp}) \quad (1)$$

where ω_{cp} is the desired nutation frequency from eq. 2 and Δpwr is the change in power in dB from the hard pulse values. ¹H and ¹³C CP power and frequency offsets were then iteratively optimized for the null crossing of the target signal.

SMILE Modeling

For SMILE modeling, spectra were first processed and phased as normal and if necessary a baseline correction was applied in the vicinity of the targeted resonance. Afterwards, the processed spectrum was inverse transformed in the indirect dimension. The resulting interferograms were presented for analysis by SMILE, with the corresponding report of identified signals saved as a peak table file. Time domain signals of peaks within a specified box around the targeted resonances were then simulated and Fourier transformed using identical processing parameters as the original data. Finally, the simulated frequency domain signal was scaled to match the intensity of the original measured data and then subtracted from it, to give the excipient subtracted spectrum.

Results and Discussion

Interference from formulation components in a 2D NMR experiment can be broadly categorized by three distinct classes; near-resonance interference, on-resonance interference and heteronuclear interference. Near resonance interference describes any excipient signal that is separated from the protein signal in at least one of the two dimensions, but is still close enough to detrimentally affect the spectrum due to baseline distortions or suppression of weak peaks from unfavorable sample dynamic range in NUS applications. On-resonance interference describes an excipient signal of the same correlation type as that of the protein signal (e.g. ¹H, ¹³C or ¹H, ¹⁵N) and which is contained within in the protein spectral window in both dimensions. Heteronuclear interference describes

an excipient signal that arises from a separate coherence other than that which is selected for in the employed pulse sequence but has overlap in the direct (^1H) dimension and gives rise to T_1 noise artifacts. For biotherapeutic applications, this type of interference is most often found in NH correlations experiments when gradient coherence selection is not employed (Arbogast et al. 2016) and will not be considered here further.

Near Resonance Excipients

For the near-resonance class, we consider the case of a formulation containing a disaccharide, such as the common cryogenic protecting excipient sucrose. These excipients give rise to set of strong signals downfield of the methyl protein signal in the directly detected ^1H dimension. Using a standard ^1H - ^{13}C gradient-selected, sensitivity-enhanced HSQC (gsHSQC) sequence with a typical indirect acquisition spectral width (ca. 25 ppm centered at 20 ppm), the sucrose signal is folded into the methyl window, overwhelming the spectrum and compromising the protein signal (Fig. 1a). As a large excess of signal averaging over that required for the minimum phase cycle is employed for mAb ^1H - ^{13}C gsHSQC spectra, it would be possible to separate the sucrose and protein signals using a ca. three times larger indirect spectral window while keeping the indirect acquisition time constant and using three times fewer scans, achieving the same signal to noise (S/N) and resolution per experimental time. However, this would still result in unfavorably increased sample dynamic range (Table S1). Furthermore, in consideration of chemometric analysis of resultant spectra, it would be advantageous to keep carrier frequencies and spectral widths matched among samples collected in various formulations.

A practical solution is to convert the standard gsHSQC sequence to a methyl selective sequence by replacing the hard ^{13}C refocusing pulse during the t_1 period with a band selective refocusing pulse (Fig. 2a). This results in significant attenuation of the carbohydrate signal, and depending on its concentration, may or may not be sufficient to record an artifact-free protein spectrum. To further attenuate the formulation signal, the proximity of the carbohydrate proton resonances to the water signal can be exploited by increasing the bandwidth of the water flip back pulse prior to the ^{13}C t_1 period, thus encompassing the carbohydrate signal. The effect of the flip-back pulse then is to restore transverse carbohydrate magnetization following the INEPT transfer, which is subsequently dephased by application of the Z-filter gradient prior to the beginning of the t_1 period. Together, these two modifications can substantially attenuate the carbohydrate signal by almost three orders of magnitude (Table S1) at 10% w/v (Fig. 1b). If necessary, the carbohydrate signal may be further attenuated by replacement of the ^{13}C excitation pulse at the beginning of the t_1 period with a band selective pulse. This results in near complete attenuation of the carbohydrate signal, allowing a high sensitivity, artifact-free protein spectrum to be acquired (Fig. 1c). This approach requires additional calibration of the evolution refocusing delay during the t_1 period to avoid introduction of 1st order phase errors in the ^{13}C dimension, which may be disadvantageous for robust analytical application of this methodology. While we demonstrate this method for carbohydrates, it should be suitable for suppressing any signal within ca. 1.5 ppm of water and outside of the carbon spectral region of interest.

On-resonance Excipients

For the case of on-resonance formulation signals, traditional band-selective methods will fail to selectively attenuate the interfering signal without affecting the protein signal as well. However, here too, the additional signal dispersal afforded by the multidimensional NMR experiment can be exploited by employing a selective double resonance technique to target the interfering signal. It has been previously demonstrated that weak field ($\omega_1 \leq J_{IS}$) heteronuclear cross polarization can be used to selectively transfer magnetization between, or invert spin-states of, coupled spin pairs in solution (Ferrage et al. 2000; Arbogast et al. 2013). Likewise, it can also be shown that starting from Zeeman magnetization in an I-S coupled spin system, spin I can be transformed entirely into multiple quantum coherence under the specific situation in which Hartmann-Hahn matched radio frequency (RF) fields are applied directly at the decoupled double-resonance position of a target resonance and with a nutation frequency of

$$\omega_{cp} = \frac{\sqrt{3}}{4} J_{IS} \quad (2)$$

and over a duration of

$$\tau_{CP} = \frac{1}{J_{IS}} \quad (3)$$

Specifically the system evolves as

$$I_z \xrightarrow{H_{CP}\tau_{CP}} -2I_xS_y \quad (4)$$

(See Appendix 1). This leads to the scheme shown in Fig. S1, which we refer to as the *Selective Excipient Reduction and Removal* (SIERRA) filter. The SIERRA filter consists of a pair of I-spin (^1H) and S-spin (^{13}C) contact pulses (CP) applied according to equations 1 and 2, followed by subsequent application of a hard 180° pulse and single contact pulse on the I-spin to refocus off target effects for signals that are on resonance in the I-spin dimension only. All non-targeted I-spins are left inverted at the end of the sequence. For spins near degenerate with the target S-spin frequency, however, nutation about the applied S-spin effective RF field will occur producing an offset dependent mixture of longitudinal and transverse S-spin magnetization. Any resulting effects on the experiment are expected to be second order as ^1H Zeeman magnetization is selected for by the subsequent HSQC experiment. Because the SIERRA filter only affects signals with doubly resonant degeneracy, it is possible to individually optimize SIERRA parameters for multiple excipient signals and then apply successive SIERRA filter elements targeting each excipient signal in a series. We note that the ability to apply an arbitrary number of SIERRA filter elements is not unlimited since resonances degenerate in ^1H will begin suffer loss of intensity due to T_2 relaxation effects (during application of CP pulses) and RF saturation (*vide infra*).

To demonstrate the performance of the SIERRA filter, we have inserted it prior to the initial ^1H polarizing pulse of the methyl-selective ^1H - ^{13}C gsHSQC experiment. The methyl-selective experiment was chosen over the standard gsHSQC since many excipients contain signals downfield of the methyl region in the ^{13}C dimension that can be adequately suppressed by band-selective ^{13}C pulsing. As a corollary, the SIERRA filter could similarly be combined with other $^1\text{H}/^{13}\text{C}$ correlation experiments, such as the SOFAST-HMQC, where selective pulsing in the ^1H dimension could likewise prove beneficial for near-resonance excipient suppression. However, for application to natural abundance proteins when experimental duration is not a limiting factor, the gsHSQC experiment sequence is preferred due to its superior suppression of heteronuclear interference artifacts arising from ^1H - ^{12}C spin-pairs. When used as part of a 2D $^1\text{H}/^{13}\text{C}$ correlation experiment, it is necessary to add an additional ^1H hard 180° pulse after the SIERRA filter to restore water magnetization along the +Z axis, followed by a Z-filter gradient to dephase transverse magnetization resulting from the element (Fig. 2b). As a first target, we consider the NISTmAb in 10 mM sodium acetate at pH 5.2. The acetyl CH_3 group of sodium acetate has a resonance in the vicinity of the N-acetyl glucosamine (GlcNAc) resonance of the NISTmAb glycan in the downfield region of the methyl spectrum as illustrated in the methyl-selective gsHSQC spectrum (Fig. 3a). After optimization of the SIERRA filter CP power and contact time, a second spectrum was recorded with the SIERRA filter. The sharp signal from acetate has been completely removed by initial conversion to $2\text{H}_x\text{C}_y$ magnetization and subsequent evolution during the pulse sequence which does not result in observable magnetization, while the majority of protein signals, including those that are degenerate along either the ^{13}C or ^1H dimension only are not appreciably perturbed by the element (Fig. 3b). This is not without any consequence to the spectrum however, as the upfield GlcNAc signals have also been attenuated. This is likely due to the partial degeneracy of GlcNAc ^1H - ^{13}C and acetyl ^1H - ^{13}C transitions, a limitation seen in previous applications of semi-selective cross polarization and selective inversion. (Arbogast et al. 2013; Ferrage et al. 2004)

Comparison of performance of pulse sequences.

Given off-target effects observed on the upfield GlcNAc signal, the assessment of the offset dependence of the ^1H and ^{13}C RF fields was determined. First, with the ^{13}C field set to the acetate resonance, the ^1H field was systematically varied by ± 800 Hz from the target resonance, and then the same was performed for the ^1H field on

resonance and the ^{13}C field varied (Fig. S2). In both dimensions, the offset dependence is quite steep and extends out no more than ± 300 Hz or ca. 2 times $^1\text{J}_{\text{HC}}$. As neighboring resonances are more likely to be offset in both ^1H and ^{13}C , we then systematically varied both fields along a grid of $\pm 3.5 \times ^1\text{J}_{\text{HC}}$ (437.5 Hz) in increments of $0.5 \times ^1\text{J}_{\text{HC}}$ (67.5 Hz) (Fig. S3). Severe attenuation is observed in the area bound by $\pm ^1\text{J}_{\text{HC}}$ in ^1H and $\pm 1.5 \times ^1\text{J}_{\text{HC}}$ in ^{13}C with lesser attenuation over the area from $\pm 1.5 \times ^1\text{J}_{\text{HC}}$ in ^1H and $\pm 2.0 \times ^1\text{J}_{\text{HC}}$ in ^{13}C . For reference, the attenuated GlcNAc peak is located at ca. $+0.25 \times ^1\text{J}_{\text{HC}}$ in ^1H and $-1.2 \times ^1\text{J}_{\text{HC}}$ in ^{13}C from the acetate resonance.

To further assess the relative performance of the excipient attenuating experiments, each was recorded on the NISTmAb sample in 10 mM sodium acetate at pH 5.2, using identical values for common acquisition parameters and compared against the benchmark gsHSQC experiment. To avoid incorporating the variability in the acetate signal from the various methods as well as effects specific to the application of weak RF during the SIERRA filter, spectra were analyzed over the area from -0.7 ppm to 1.75 ppm in ^1H and 11 ppm to 28 ppm in ^{13}C with noise estimated from the spectral area over 8 ppm to 10.7 ppm in ^1H and 19 ppm to 28 ppm in ^{13}C . The S/N values for the various experiments are summarized in Table 1. The fully methyl-selective gsHSQC experiment exhibits a slight loss of sensitivity relative to the standard gsHSQC experiment, likely as a result of additional T_2 relaxation during the ^{13}C selective pulses. Similarly, when considering the SIERRA-gsHSQC experiment, we find that even when no RF power is applied to the ^1H and ^{13}C contact pulse (CP off), modest signal losses of ca. 5% are incurred. These additional losses likely result from of the second τ_{cp} delay during the ^1H RF refocusing period in which the ^1H magnetization is inverted leaving ^1H T_1 relaxation active, as well as from the two additional ^1H 180° pulses. On the other hand, when ^1H and ^{13}C CP RF power is applied (CP on) at the frequency of the acetate resonance, additional losses of ca. 4% relative to the case with the CP RF off are observed, perhaps as a result of spin diffusion; still, total protein signal losses from the SIERRA filter remain less than 10%. To better illustrate the advantage of this methods over diffusion based methods, a standard pulse-field gradient stimulated echo (PFG-STE) diffusion filtered ^1H - ^{13}C HSQC (Rajagopalan et al. 2004) was recorded with both short (5.5 ms) and long (30 ms) filter delays. The former is the minimum delay to accommodate application of the gradients while the latter was chosen as the optimal delay to remove the acetate peak from the spectrum. As shown in Table 1, even when employing the shorter delay, significant attenuation of the protein spectrum is observed and would require 4 times more experimental duration (~17 hrs.) just to achieve a benchmark S/N of 10:1, and 20 times more experimental duration (~86 hrs.) to match the gsHSQC experiment. Furthermore, we note that even if the increased experimental duration was deemed acceptable, the PFG-STE experiment fails for excipients that do not exhibit large disparities in correlation time relative to the protein of interest, such as polysorbate-80, which is known to form micelles under the experimental conditions (Fig. S4) (Braun et al. 2015).

Given that the SIERRA filter results in protein signal losses due to relaxation effects as well as from spin diffusion, we examined the performance of the sequence as a function of the number of repetitions of the SIERRA element. Shown in Fig. 4a is the performance for three different scenarios of CP RF application; targeting the acetate signal, on resonance in ^1H (@ 1 ppm) but off resonance in ^{13}C (@ 24.3 ppm) and on resonance in ^{13}C (@19ppm) but off resonance in ^1H (@ 2.0 ppm). Signal integrals were taken for the 1st FID of the SIERRA-gsHSQC experiment over 1.95 ppm to -0.5 ppm in ^1H and at each repetition, intensities were normalized relative to the case where CP RF power was not applied. For up to four successive applications of the SIERRA filter, total losses for all cases remain less than 50%, and even at 10 repetitions, the signal losses are less than those observed for the PFG-STE HSQC experiment. Looking more closely at the different scenarios of applied RF, it can be seen that for acetate on-resonance and ^{13}C on-resonance, losses from RF saturation remain relatively modest (<25% for four repetitions), conversely for the case of ^1H on-resonances losses from RF saturation are much more pronounced.

The question arises as to which case better represents a practical experimental outcome where the ^1H and ^{13}C RF carrier frequencies are moved between repetitions to target different excipient signals. To explore this further, ten total targets were chosen; eight based on signal frequencies for acetate, polysorbate-80 (4), L-arginine, L-alanine, and L-methionine as well as two additional two arbitrarily chosen frequencies. Where carrier frequencies based on the chosen excipient signals overlapped with protein resonances, they were offset slightly to a nearby empty region of the spectrum. The signal intensities were measured as before, with the order of the carrier

frequencies arranged in two different scenarios. The resulting intensities were then normalized to the case when the carrier was on resonance in ^1H at each repetition (Fig S5). For limited applications of the SIERRA element (≤ 4), the effect of moving the carrier position on total signal intensity is on the order of (0.91) to slightly better (1.31) than that of the fixed carrier on ^1H , depending on the exact order of the applied frequencies. In the first case, the ^1H carrier began on the acetate signal and was then moved into the center of the spectrum in ^1H (1.96, 1.04, 1.42 and 1.58 ppm). T_2 losses incurred during RF application at multiple points in the heart of the ^1H dimension are clearly manifest. In the second case, the ^1H carrier was initially placed far upfield and then alternated between signals in the extreme down and upfield positions (-0.30, 2.54, 0.19, 2.25 ppm). Because fewer protein resonances were degenerate with the ^1H carrier position, losses were mild with respect to the fixed carrier at 1 ppm. Ultimately the exact nature of signal loss will be a function of the extent of overlap between targeted excipient resonances and protein resonances in the ^1H dimension.

To further investigate whether use of the SIERRA filter would compromise the spectrum for statistical analysis, a series of 2D spectra with up to ten SIERRA repetitions were recorded with both the RF CP power off as well as RF CP power on, targeting the acetate signal. Following acquisition, spectra were subject to direct principal component analysis (PCA) of their spectral intensity matrix (Fig. 4b) (Arbogast et al. 2017). Differentiation of both RF on and RF-off spectra as a function of SIERRA repetition number is observed in the 2nd principal component, which we attribute to variation in intensity from relaxation losses during the element. For low repetition numbers, and especially for a single SIERRA application, RF-on spectra cluster quite closely to the RF-off reference spectra. For higher repetition numbers the RF-on spectra begin to look more relaxed owing to greater influence of T_2 effects, clustering with RF-off spectra of a higher repetition number. For extreme repetition values (>6), RF-on spectra begin to deviate appreciably from RF-off spectra with large separation in the 3rd principal component, suggesting the effects of spin-diffusion are becoming manifest. However, these results indicate that for limited applications of the SIERRA filter, such as for the cases presented herein, the resulting spectra remain suitable for chemometric analysis.

Post-Acquisition Removal of Residual Excipient Signal

It may be envisioned that in certain instances, the SIERRA filter may not achieve removal of the excipient signal to the noise baseline, such as when that signal is orders of magnitude larger than the protein signal or when it lies in a heavily overlapped spectral region in the ^1H dimension, making optimization of the null condition difficult. Ideally, formulation signals should be removed completely to baseline to allow for robust chemometric analysis. However, once the excipient signal is attenuated to the point that it is on the order of the protein signal, with no accompanying baseline distortions, it is possible to employ post-processing techniques to more completely eliminate the unwanted resonance. For example, the SMILE algorithm, which iteratively decomposes features in the frequency domain by simulating ideal time domain signals and then removes these from the measured time-domain data, can in principle, also be used to model and remove any arbitrary signal at will (Ying et al. 2017). To demonstrate this, we consider the scenario where the acetate peak has been only partially attenuated ($\tau_{cp} = 4.2$ ms) to mitigate associated baseline distortions. To begin, SMILE is used to analyze the spectrum including the partially attenuated excipient signals. The result includes a list of ideal time-domain signal parameters which taken together form a complete model of the spectral features to a desired degree of approximation. A model spectrum can be generated by Fourier transform of the ideal time-domain signals using the same processing protocols applied to the experimental data. In this case, the SMILE result for the entire spectrum consists of ca. 6,000 ideal 2D time-domain signals, which together approximate the peaks and their possible multiplicity, phase distortion, and line shape irregularities, while also approximating solvent artifacts and some of the noise. Then, a simulated spectrum is generated from a subset of the SMILE results, using only model signals in the immediate region of the excipient peak; ^1H , 2.12 to 2.07 ppm, ^{13}C , 27.0 to 25.0 ppm. This yields six signals which together are used to generate a synthetic noise-free spectrum of the acetate peak. Subsequent difference of the original spectrum with the synthetic SMILE data results in an excipient-free spectrum with no residual baseline distortion at the excipient frequency (Fig. 5). Notably, unlike the outcome for subtraction of an actual measured buffer blank spectrum, since the SMILE synthetic excipient spectrum is noise free, the difference spectrum retains the signal to noise of original spectrum (Fig. S6).

To further demonstrate how excipient signals may be removed using these methods, we consider the NISTmAb in two other commercial mAb excipients; 0.3 % w/v Tween-80 and 50 mM L-methionine. For the first case, four signals from the excipient are present in the methyl window (Fig. 6a). The various polysorbate-80 signals overlap with protein signals in the direct dimension, and initial estimates of the SIERRA filter parameters did not achieve the null condition, resulting in negative signals for three of the four excipient peaks. The fourth signal at 1.45, 23.5 ppm, was purposely left only modestly attenuated due to its large overlap with protein signals of interest (Fig. 6b). Using the SMILE method, each residual polysorbate-80 signal can be targeted and removed to the baseline (Fig. 6c). The second case illustrates the performance of the SIERRA element when the excipient signal is orders of magnitude larger than that of the protein signal, as shown in Fig. 7a. Application of the SIERRA filter substantially attenuates the methionine excipient signal and removes the attendant baseline distortions; however a residual excipient signal still remains (Fig. 7b). In the immediate vicinity of the residual signal at 2.2, 14 ppm, a signal which may belong to a protein methionine residue is seen. Under the assumption that the signal is indeed a protein signal, which we wish to keep, the SMILE algorithm can be deployed to surgically remove the residual methionine excipient signal while leaving the remaining suspected protein signal as illustrated in Fig. 7c.

On a final note, applications presented herein have been based on a theoretical ideal case for the perfect and complete conversion of initial I_z magnetization into unobservable $2I_xS_y$ multiple quantum magnetization. However, when performing a cross polarization experiment in liquids, it is well known that violations of the Hartmann-Hahn condition are easily achieved and can result in a loss of magnetization (Pelupessy and Chiarparin 2000). For non-ideal experimental conditions, impure spin state outcomes are expected; however most will not pass through into subsequent the HSQC element leaving the sequence quit robust to experimental imperfections. Indeed, if we set the contact time to greater than $1/J$, and systematically vary the ^1H CP power level, we can achieve a similar null condition (Fig. S7). Thus, it is likely that multiple scenarios exist to achieve severe attenuation of a targeted signal via double resonance irradiation. As such, one can have confidence that using estimated parameters followed by a brief parameter optimization, even if the ideal condition is not achieved, it will still be sufficient to reduce excipient signal amplitudes and, in conjunction with the SMILE algorithm, remove the signal to the baseline.

Conclusions

2D ^1H - ^{13}C methyl correlated NMR at natural isotopic abundance can provide residue level structural characterization of biotherapeutics and has the potential to impact nearly all life-cycle stages of a product from discovery through to lot release. Indeed, with the recent multi-national NMR mAb interlaboratory study, application of 2D NMR in the industry is moving from an investigational technology towards a harmonized method. (R. Brinson, personal communication). A major hindrance to wide spread adoption of the methodology is interference from commonly employed aliphatic excipients. The use of selective pulse techniques including traditional band selective methods as well as the novel SIERRA filter, combined with SMILE-based signal subtraction presented herein show promise at mitigating these interfering effects and removing a major obstacle to the routine application of NMR in the biopharmaceutical industry. The ability to record a clean spectrum of formulated protein therapeutics is especially useful in conjunction with new modes of chemometric analysis of biotherapeutic NMR spectra under development and should extend the utility of NMR in the industry and open up new avenues for structural characterization and comparability.

Appendix I

Basic theoretical treatment of J-driven selective cross polarization

Here we present the theoretical framework for selective cross polarization carried out under the simplifying assumptions of matched RF field strengths and furthermore with the RF field nutation frequency set to the so-called “magic field”, which is defined and described below. This theoretical framework starts from the concept of fictitious spin-1/2 operators as originally described by Chiarparin and Pelupessy, (Pelupessy and Chiarparin 2000; Chiarparin

et al. 1998) but carry the development further as we are interested in the evolution of components of the system beyond the scope of that work.

We start by the introduction of the ZQ and DQ fictitious spins-1/2 which are defined analogously to the work of Chiarparin and Pelupessy, but in an untilted frame and with different Cartesian labeling. The ZQ and DQ fictitious spins are defined in turn with respect to the Cartesian spin operators corresponding to a coupled pair of spin-1/2 nuclei I and S,

$$\begin{aligned} D_x &= \frac{1}{2}(I_x + S_x) & Z_x &= \frac{1}{2}(I_x - S_x) \\ D_y &= I_y S_z + S_y I_z & Z_y &= I_y S_z - S_y I_z \\ D_z &= I_z S_z - S_y I_y & Z_z &= I_z S_z + S_y I_y \end{aligned}$$

Note that the ZQ and DQ spins are governed by the same commutation relationships observed for an ordinary spin pair, ie.

$$[Z_i, Z_j] = i\varepsilon_{ijk}Z_k; [D_i, D_j] = i\varepsilon_{ijk}D_k; [Z_i, D_j] = 0$$

where i, j, and k represent Cartesian components x, y, and z and the ε_{ijk} represent components of the Levi-Civita tensor. In the conventional treatment of cross polarization, both the spin Hamiltonian and the state of the system can be represented in terms of these two fictitious spin operators allowing straightforward calculation of the time evolution of the system. Of interest here are cases in which a complete set of operators, beyond those corresponding to D and Z, spins must be introduced in order to describe the state of the system.

We consider here evolution of the system under dual on resonance RF irradiation to the coupled IS spin pair. Expressed in the doubly rotating frame, the spin Hamiltonian can be written,

$$H_{CP}^{DR} = \omega_1^I I_x + \omega_1^S S_x + 2\pi J I_z S_z$$

Assuming matched RF fields (ie. $\omega_{1s} = \omega_{1i}$) and expressing in terms of fictitious spin operators we obtain,

$$H_{CP}^{DR} = 2\omega_1 D_x + \pi J D_z + \pi J Z_z$$

Looking forward, in order to calculate the evolution of the system starting from an arbitrary initial state described by $\sigma(0)$ it suffices to consider evolution in what we term the ZQ and DQ effective field frames. Due to the commutivity of the Z and D spin operators the evolution of the system with respect to each effective field can be considered sequentially. Additionally, system evolution within the DQ effective field frame will first require transformation into the appropriate tilted frame.

Transformation into DQ effective field frame

This transformation requires a rotation about the D_y axis by an angle which we denote θ_D . The DQ effective field is shown in Fig. 8 with the angle θ_D determined according to

$$\theta_D = \arctan\left(\frac{4v_1}{J}\right)$$

As such, to transform an arbitrary untilted operator O into its representation within the DQ effective field frame, O', we must perform the following transformation,

$$O' = T^{-1}OT$$

where

$$T = \exp[-i\theta(I_y S_z + S_y I_z)]$$

We note that this mirrors the classic transformation into the effective field frame except in this case carried out within the DQ spin manifold,

$$I_z S_z - I_y S_y \xrightarrow{DQ(eff)} (I'_z S'_z - I'_y S'_y) \cos \theta_D - \frac{1}{2}(I'_x + S'_x) \sin \theta_D$$

$$\frac{1}{2}(I_x + S_x) \xrightarrow{DQ(eff)} \frac{1}{2}(I'_x + S'_x) \cos \theta_D - (I'_z S'_z - I'_y S'_y) \sin \theta_D$$

System evolution in the ZQ and DQ effective field frames

We will carry out the evolution of the system within the corresponding DQ and ZQ effective field frames. Evolution of the system for a time interval τ is calculated using the rotation operators shown below respectively,

$$R_D(\tau) = \exp[-i\omega_D \tau (I'_z S'_z - I'_y S'_y)]$$

$$R_Z(\tau) = \exp[-i\omega_Z \tau (I'_z S'_z + I'_y S'_y)]$$

where ω_Z and ω_D refer to ZQ and DQ precessional frequencies and the primed operators denote their expression in the corresponding effective field frame.

Under the influence of the ZQ part of the dual RF Hamiltonian the system will precess about the (untilted) Z_z axis at a frequency $\omega_Z = 0.5J$ (in Hz). With respect to the DQ part, the system will precess about the effective field vector D'_z at a frequency ω_D determined according to,

$$\omega_D = \sqrt{\frac{1}{4}J^2 + 4v_1^2}$$

For the purposes of this work the RF nutation frequency must be set to a specific value referred to as the magic field. The magic field value of smallest possible magnitude is,

$$v_1 = \frac{\sqrt{3}}{4}J$$

which results in a value of $\omega_D = J$ (in Hz). Such a situation in which the DQ frequency is a multiple of the ZQ frequency is precisely the relationship required for different possible values of the magic field.

Commutation relationships

In order to handle any possible state of the system, commutation relationships between the ZQ and DQ operators and the remaining operators are needed. We state them here for convenience.

$$[I_z S_z \pm I_y S_y, 2I_x S_y] = \frac{1}{2}iI_z$$

$$[I_z S_z \pm I_y S_y, I_y] = -iI_x S_z$$

$$[I_z S_z \pm I_y S_y, I_x S_z] = \frac{1}{2}iI_y$$

$$[I_z S_z \pm I_y S_y, I_z] = \pm \frac{1}{2}iI_x S_y$$

$$[I_y S_z + I_z S_y, 2I_x S_y] = -\frac{1}{2}iI_y$$

$$[I_y S_z + I_z S_y, I_y] = iI_x S_y$$

$$[I_y S_z + I_z S_y, 2I_x S_z] = \frac{1}{2} i I_z$$

$$[I_y S_z + I_z S_y, I_x] = -i I_x S_z$$

System evolution under magic field cross polarization Hamiltonian

Now we can derive our central result. Starting from Zeeman magnetization for spin I, what is the state of the system after evolution under the magic field CP Hamiltonian for a duration $\tau_{CP} = 1/J$?

First we consider evolution under the ZQ part of the Hamiltonian,

$$I_z \xrightarrow{\omega_z \tau \hat{Z}_z} 2I_x S_y$$

Corresponding to a π rotation about Z_z . Now we need to transform the system into the DQ effective field frame. We note that a magic field set as described above corresponds to a value of $\theta_D = 60$. This leads to,

$$2I_x S_y \xrightarrow{DQ(eff)} 2I'_x S'_y \cos \frac{1}{2} \theta_D - I'_y \sin \frac{1}{2} \theta_D$$

Evolution under the DQ part of the Hamiltonian (by an angle 2π) then leads to,

$$2I'_x S'_y \cos \frac{1}{2} \theta_D - I'_y \sin \frac{1}{2} \theta_D \xrightarrow{\omega_D \tau \hat{D}_z} -2I'_x S'_y \cos \frac{1}{2} \theta_D + I'_y \sin \frac{1}{2} \theta_D$$

This is then followed by transformation back into the lab frame,

$$-2I'_x S'_y \cos \frac{1}{2} \theta_D + I'_y \sin \frac{1}{2} \theta_D \xrightarrow{DQ(eff)^{-1}} -2I_x S_y$$

for the desired net result of

$$I_z \xrightarrow{H_{CP} \tau_{CP}} -2I_x S_y$$

Associated Content

Pulse sequence and parameter files as well as protocols and example processing scripts can be found at www.ibbr.umd.edu/sierra/home

Supporting Information

Additional information as noted in text.

Acknowledgements

We gratefully acknowledge the help and support of Jinfa Ying for assistance with modifying the SMILE algorithm. We further acknowledge the support of the NIST Biomanufacturing Initiative and W.M. Keck for support of the biomolecular NMR instrumentation.

References

1. Aubin Y, Gingras G, Sauvé S (2008) Assessment of the Three-Dimensional Structure of Recombinant Protein Therapeutics by NMR Fingerprinting: Demonstration on Recombinant Human Granulocyte Macrophage-Colony Stimulation Factor. *Anal Chem* 80: 2623–2627.
2. Aubin Y, Jones C, Freedberg DI (2010) Using NMR spectroscopy to obtain the higher order structure of pharmaceutical products : simple methods can characterize polysaccharide vaccines and recombinant

- cytokines at high resolution. *Biopharm Int* 23: 28–34.
3. Wishart DS (2013) Characterization of biopharmaceuticals by NMR spectroscopy. *TrAC Trends Anal Chem* 48: 96–111.
 4. Arbogast LW, Brinson RG, Marino JP (2015) Mapping Monoclonal Antibody Structure by 2D ^{13}C NMR at Natural Abundance. *Anal Chem* 87: 3556–3561.
 5. Arbogast LW, Brinson RG, Marino JP (2016). Application of Natural Isotopic Abundance ^1H – ^{13}C - and ^1H – ^{15}N -Correlated Two-Dimensional NMR for Evaluation of the Structure of Protein Therapeutics. in *Methods in Enzymology* 566: 3–34.
 6. Singh SM, Bandi S, Jones DNM, Mallela KMG (2017) Effect of Polysorbate 20 and Polysorbate 80 on the Higher-Order Structure of a Monoclonal Antibody and Its Fab and Fc Fragments Probed Using 2D Nuclear Magnetic Resonance Spectroscopy. *J Pharm Sci* 106: 3486–3498.
 7. Ghasriani H, Hodgson DJ, Brinson RG, McEwen I, Buhse LF, Kozlowski S, Marino JP, Aubin Y, Keire DA (2016) Precision and robustness of 2D-NMR for structure assessment of filgrastim biosimilars. *Nat Biotechnol* 34: 139–141.
 8. Brinson RG, Ghasriani H, Hodgson DJ, Adams KM, McEwen I, Freedberg DI, Chen K, Keire DA, Aubin Y, Marino JP (2017) Application of 2D-NMR with room temperature NMR probes for the assessment of the higher order structure of filgrastim. *J Pharm Biomed Anal* 141: 229–233.
 9. Japelj B, Ilc G, Marušič J, Senčar J, Kuzman D, Plavec J (2016) Biosimilar structural comparability assessment by NMR: from small proteins to monoclonal antibodies. *Sci Rep* 6: 32201.
 10. Chen K, Park J, Li F, Patil SM, Keire DA (2018) Chemometric Methods to Quantify 1D and 2D NMR Spectral Differences Among Similar Protein Therapeutics. *AAPS PharmSciTech* 19: 1011–1019.
 11. Arbogast LW, Delaglio F, Schiel JE, Marino JP (2017) Multivariate Analysis of Two-Dimensional ^1H , ^{13}C Methyl NMR Spectra of Monoclonal Antibody Therapeutics To Facilitate Assessment of Higher Order Structure. *Anal Chem* 89: 11839–11845.
 12. Poppe L, Jordan JB, Lawson K, Jerums M, Apostol I, Schnier PD (2013) Profiling Formulated Monoclonal Antibodies by ^1H NMR Spectroscopy. *Anal Chem* 85: 9623–9629.
 13. Franks J, Glushka JN, Jones MT, Live DH, Zou Q, Prestegard JH (2016) Spin Diffusion Editing for Structural Fingerprints of Therapeutic Antibodies. *Anal Chem* 88: 1320–1327.
 14. Ying J, Delaglio F, Torchia DA, Bax A (2017) Sparse multidimensional iterative lineshape-enhanced (SMILE) reconstruction of both non-uniformly sampled and conventional NMR data. *J Biomol NMR* 68: 101–118.
 15. Delaglio F, Grzesiek S, Vuister G, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6: 277–293.
 16. Arbogast LW, Brinson RG, Formolo T, Hoopes JT, Marino JP (2016) 2D ^1H N, ^{15}N Correlated NMR Methods at Natural Abundance for Obtaining Structural Maps and Statistical Comparability of Monoclonal Antibodies. *Pharm Res* 33: 462–475.
 17. Ferrage F, Eykyn TR, Bodenhausen G (2000) Coherence transfer by single-transition cross-polarization: Quantitation of cross-correlation effects in nuclear magnetic resonance. *J Chem Phys* 113: 1081–1087.
 18. Arbogast L, Majumdar A, Tolman JR (2013) Unraveling long range residual dipolar coupling networks in strongly aligned proteins. *J Magn Reson* 235: 26–31.
 19. Ferrage F, Eykyn TR, Bodenhausen G (2004) Frequency-Switched Single-Transition Cross-Polarization: A Tool for Selective Experiments in Biomolecular NMR. *ChemPhysChem* 5: 76–84.
 20. Rajagopalan S, Chow C, Raghunathan V, Fry CG, Cavagnero S (2004) NMR spectroscopic filtration of polypeptides and proteins in complex mixtures. *J Biomol NMR* 29: 505–516.
 21. Braun AC, Ilko D, Merget B, Gieseler H, Germershaus O, Holzgrabe U, Meinel L (2015) Predicting critical micelle concentration and micelle molecular weight of polysorbate 80 using compendial methods. *Eur J Pharm Biopharm* 94: 559–568.
 22. Pelupessy P, Chiarparin E (2000) Hartmann-Hahn polarization transfer in liquids: An ideal tool for selective experiments. *Concepts Magn Reson* 12: 103–124.
 23. Chiarparin E, Pelupessy P, Bodenhausen G (1998) Selective cross-polarization in solution state NMR. *Mol Phys* 95: 759–767.
 24. Schleucher J, Schwendinger M, Sattler M, Schmidt P, Schedletzky O, Glaser SJ, Sørensen OW, Griesinger C (1994) A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients. *J Biomol NMR* 4: 301–6.
 25. Emsley L, Bodenhausen G (1990) Gaussian pulse cascades: New analytical functions for rectangular selective inversion and in-phase excitation in NMR. *Chem Phys Lett* 165: 469–476.

Table 1 Signal to Noise Analysis of Excipient Suppressing Experiments

Experiment*	Signal to Noise	Ratio of Signal to Noise	
		Standard	Selective
gsHSQC	22.938	1	N/A
Selective	21.260	0.927	1
SIERRA-gsHSQC CP RF-off	20.232	0.882	0.952
SIERRA-gsHSQC CP RF-on	19.390	0.845	0.912
PFGSTE-gsHSQC 5 ms	5.079	0.221	N/A
PFGSTE-gsHSQC 30 ms	3.765	0.164	N/A

*All data were collected with 128 scans, a recycle delay of 1.5 s acquisition times of 100 ms and 10 ms over spectral widths of 14 ppm and 26.5 ppm in the F2 and F1 domains with an experimental duration of 257 min.

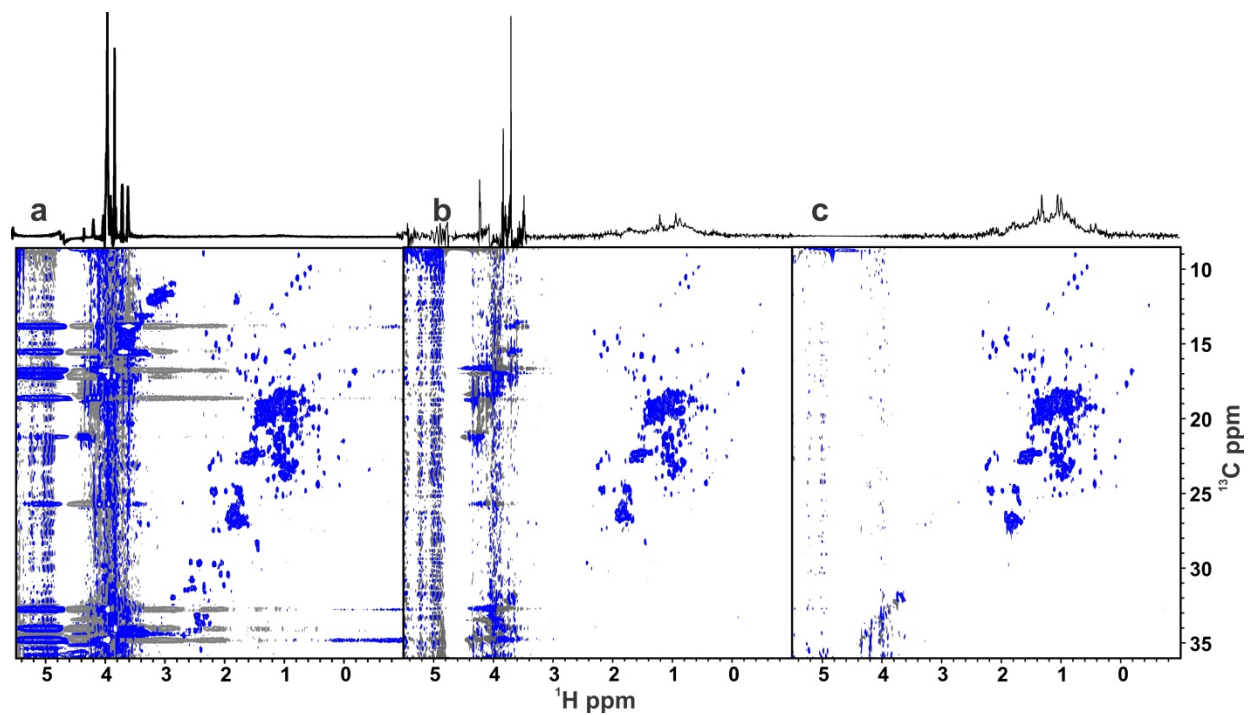


Fig. 1 2D ^1H , ^{13}C methyl correlation spectra of NISTmAb ‘formulated’ in 10% w/v sucrose and 10 mM sodium acetate (d_3) at pH 5.2. 1D ^1H projections are shown above. **a)** A standard gsHSQC spectrum. **b)** The semi-selective gsHSQC with broadband flip-back pulses and **c)** the fully-selective gsHSQC. Intensity of ^1H projections of **a** and **b** are scaled to the maximum sucrose signal to illustrate relative intensity of the carbohydrate and protein signals.

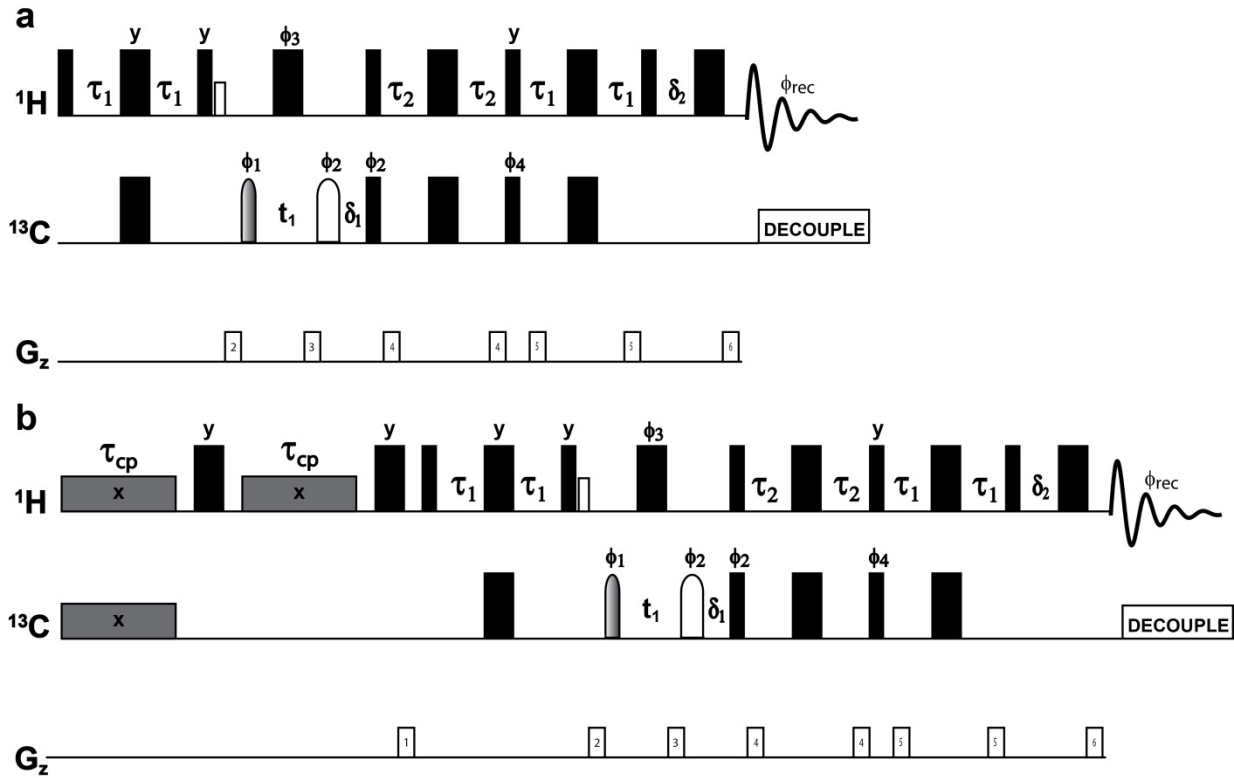


Fig. 2 Schematic representation of the modified 2D ^1H - ^{13}C gradient-selected, sensitivity-enhanced HSQC (Schleucher et al. 1994) experiments for selective removal of excipient signals. **a)** 2D ^1H - ^{13}C methyl-selective gradient-selected, sensitivity-enhanced HSQC experiment. All narrow (wide) rectangular pulses are applied with a flip angle of 90° (180°) and with a phase of x unless otherwise specified. Band selective $760 \mu\text{s}$ Q3 and 1.365 ms Q5 Gaussian cascade pulses (Emsley and Bodenhausen 1990) were employed for the 180° ^{13}C refocusing pulse and optional 90° ^{13}C pulse (shaded) respectively. A $400 \mu\text{s}$ broadband selective water flip back pulse was employed. The delays used were $\tau_1 = 1/4J_{\text{HC}} = 1.72 \text{ ms}$ $\tau_2 = 1/8J_{\text{HC}} = 862 \mu\text{s}$ $d_1 = 1.22 \text{ ms}$ and $d_2 = 1.24 \text{ ms}$. The following phase cycle was employed; $\phi_1 = x, -x$; $\phi_2 = x, x, -x, -x$ $\phi_3 = y, y, -y, -y$ $\phi_4 = x, x, -x, -x$, $\phi_{\text{rec}} = -x, x, x, -x$. Quadrature was achieved in F1 by the echo-anti-echo method. The gradients employed were g_2 , 35 G/cm for 1 ms ; g_3 , 56 G/cm for 1 ms ; g_4 , 3.5 G/cm for 0.5 ms ; g_5 , -1.4 G/cm for 1 ms and g_6 14.1 G/cm for 1 ms . **b)** ^1H - ^{13}C SIERRA-filtered methyl-selective HSQC experiment. All parameters are the same as in **a** except, ^1H and ^{13}C contact pulses are applied with an RF nutation frequency of $\sim 54 \text{ Hz}$, and τ_{cp} was set to ca. $1/J = 8 \text{ ms}$ for all targeted ^1H - ^{13}C methyl correlations and gradient g_1 was applied at 66.5 G/cm for 1.5 ms .

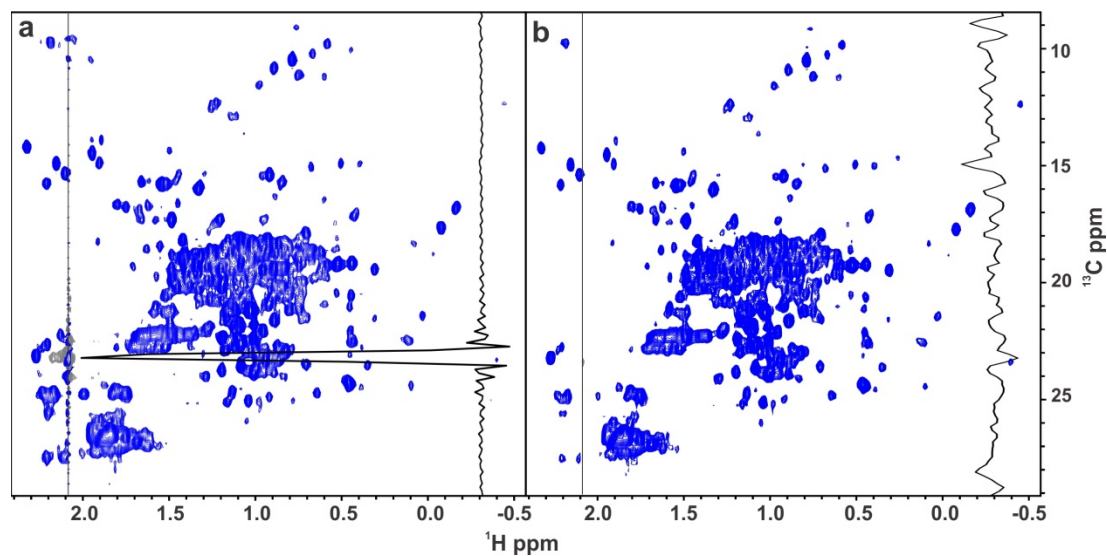


Fig. 3 2D ^1H - ^{13}C methyl SIERRA-filtered gsHSQC spectra of NISTmAb in 10 mM sodium acetate at pH 5.2. **a)** Reference spectrum acquired without RF power applied to the ^1H and ^{13}C contact pulses. **b)** SIERRA-filtered spectrum with ^1H and ^{13}C CP RF power. ^{13}C traces through the acetate resonance at 2.084 ppm in ^1H are shown to illustrate the how the removal of the acetate signal affects the baseline.

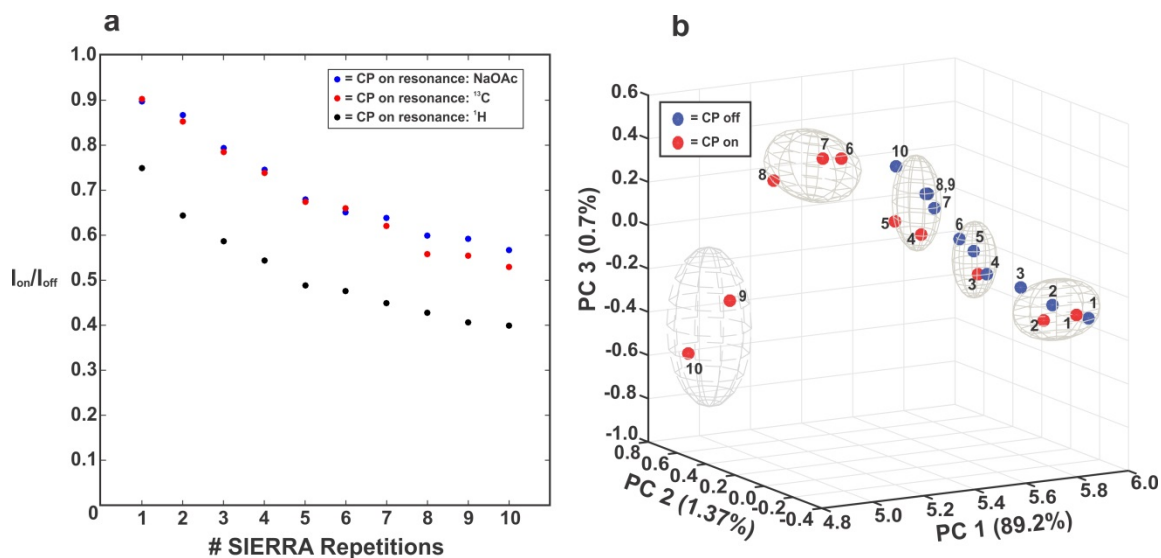


Fig. 4 Spectral analyses as a function of number of SIERRA filter repetitions. **a)** Normalized signal intensity values as a function of repetition number. Values normalized against the value obtained with the CP RF power off at each integer repetition of the SIERRA filter. Signal integrals were taken for the 1st FID of the SIERRA-gsHSQC experiment over 1.95 ppm to -0.5 ppm in ¹H. **b)** Principal component analysis (PCA) of 2D SIERRA-filtered gsHSQC spectra with and without CP RF power applied. PCA was performed directly on the processed spectral data matrices after each was normalized to uniform maximum intensity.

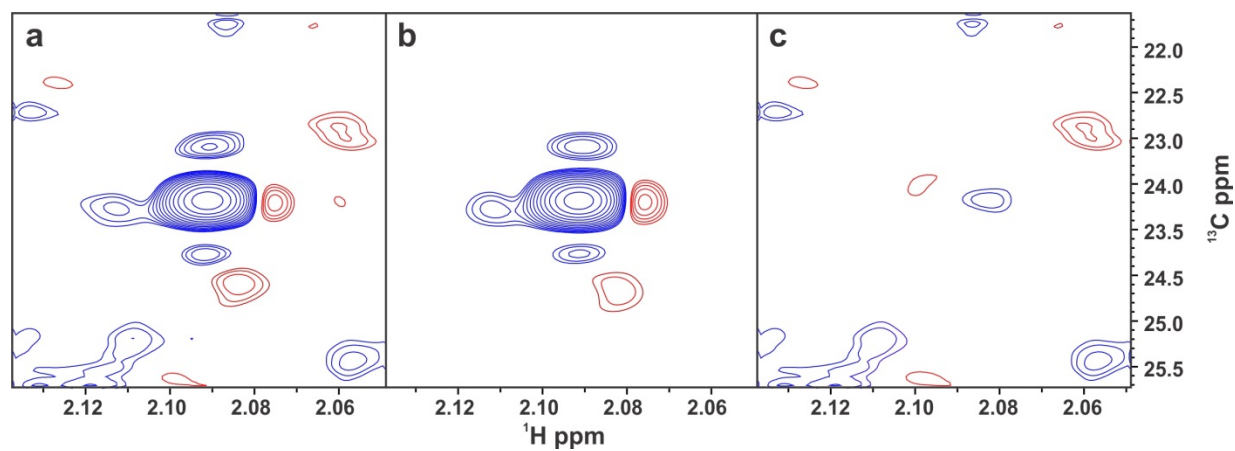


Fig. 5 Expanded region of a ^1H - ^{13}C methyl gsHSQC spectrum of NISTmAb in 10 mM sodium acetate at pH 5.2, illustrating the process of SMILE signal subtraction. **a)** The acetate signal after partial attenuation using the SIERRA filter. **b)** Synthetic SMILE spectrum of the acetate signal from **a**. **c)** Difference spectrum of **a** and **b**, where the residual acetate signal has been completely removed to the baseline.

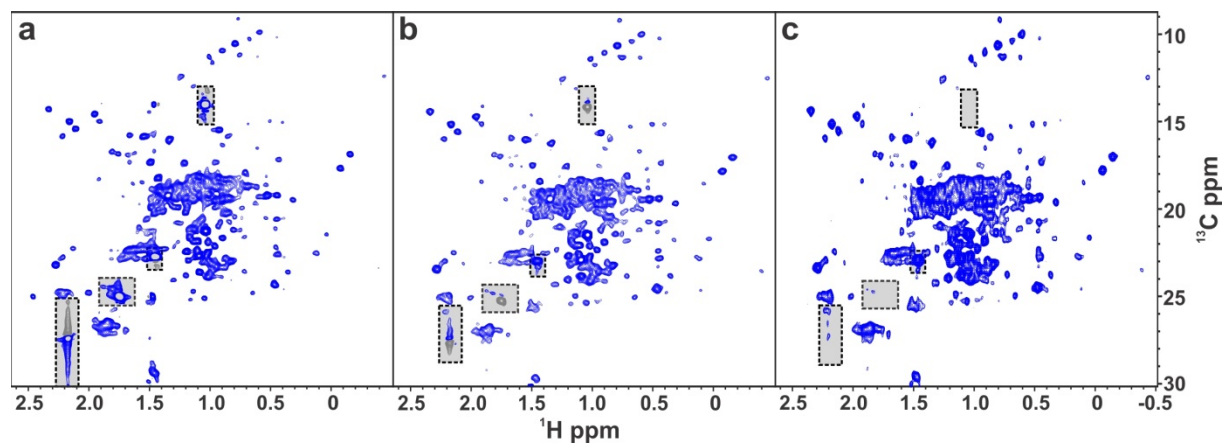


Fig. 6 ^1H - ^{13}C methyl spectra of NISTmAb in 0.3% w/v polysorbate-80, collected with 128 scans. **a)** Standard gsHSQC spectrum **b)** SIERRA-filtered gsHSQC spectrum, with the filter serially applied to the four interfering polysorbate-80 resonances. **c)** SIERRA-filtered gsHSQC spectrum after SMILE subtraction of the residual polysorbate-80 signals. Resonances position of the polysorbate-80 signals are denoted by the shaded boxes.

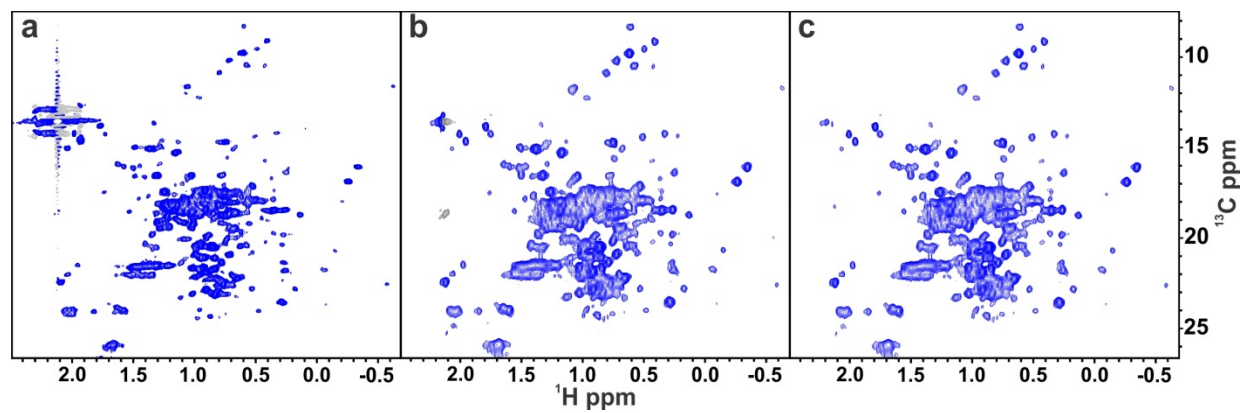


Fig. 7 ^1H - ^{13}C methyl spectra of NISTmAb ‘formulated’ in 75 mmol/L L-methionine and 25 mmol/L L-histidine at pH 6.0, collected with 128 scans. **a)** Standard gsHSQC spectrum **b)** SIERRA-filtered gsHSQC spectrum. **c)** SIERRA-filtered gsHSQC spectrum after SMILE subtraction of the residual methionine signal.

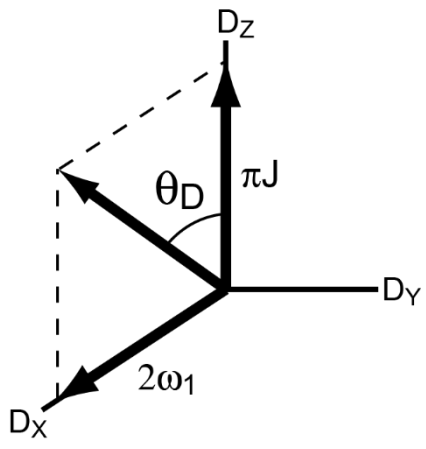


Fig. 8 DQ effective field frame.

Supporting Information

Selective Suppression of Excipient Signals in 2D ^1H - ^{13}C Methyl Spectra of Biopharmaceutical Products

Luke W. Arbogast, Frank Delaglio, Joel R. Tolman and John P. Marino

Table S1 Sample Dynamic Range

Sample	Max [†]	Min [‡]	Ratio
10% w/v sucrose, gsHSQC	4.72×10^{10}	3.83×10^5	1.23×10^5
10% w/v sucrose, gsHSQC partially-selective	6.99×10^7	2.34×10^5	2.99×10^3
10% w/v sucrose, gsHSQC fully-selective	3.52×10^7	4.62×10^5	7.62×10^1
10 mM NaOAc, SIERRA cp-off	2.62×10^8	5.99×10^5	4.37×10^2
10 mM NaOAc, SIERRA cp-on	4.94×10^7	4.81×10^5	1.03×10^2
0.3% w/v Tween, SIERRA cp-off	1.09×10^8	2.31×10^5	4.72×10^2
0.3% w/v Tween, SIERRA cp-on	1.33×10^7	2.58×10^5	5.14×10^1
25 mM L-methionine, SIERRA cp-off	1.06×10^9	3.99×10^5	2.65×10^3
25 mM L-methionine, SIERRA cp-of	2.17×10^7	3.13×10^5	6.95×10^1

[†] For gsHSQC, gsHSQC partially-selective and SIERRA cp-off spectra, the maximum intensity peak was from the excipient. For gsHSQC fully-selective and SIERRA cp-on spectra, the maximum intensity peak was the protein signal at 1.33/19.27 ($^1\text{H}/^{13}\text{C}$) ppm.

[‡] The minimum intensity peak was taken to be the protein signal at 0.286/15.33 ($^1\text{H}/^{13}\text{C}$) ppm, the smallest peak that could reliably identified in all the spectra.

Peak heights were determined using the Sparky 3 software.¹

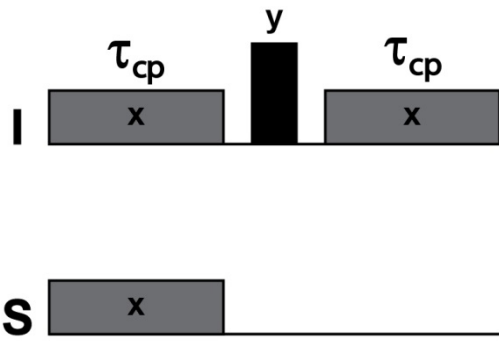


Fig. S1 Schematic representation of the SIERRA filter. τ_{cp} is set to $1/J$ or ca. 8 ms for signal removal.

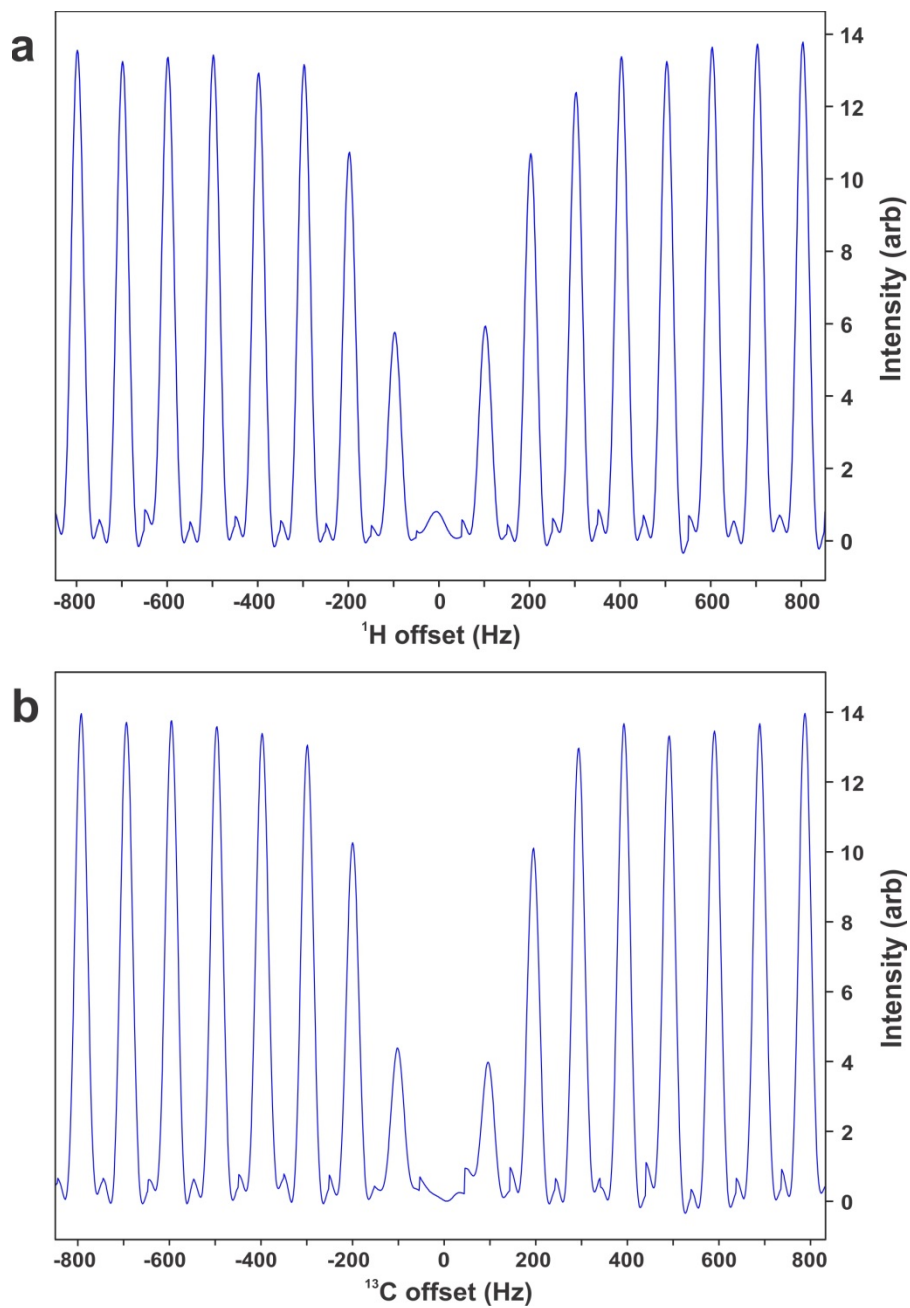


Fig. S2 Signal intensity of the acetate resonance as a function of SIERRA filter RF offsets for the case of **a)** a fixed on-resonance ^{13}C carrier frequency and a varied ^1H carrier frequency and **b)** a fixed on-resonance ^1H carrier frequency and a varied ^{13}C carrier frequency.

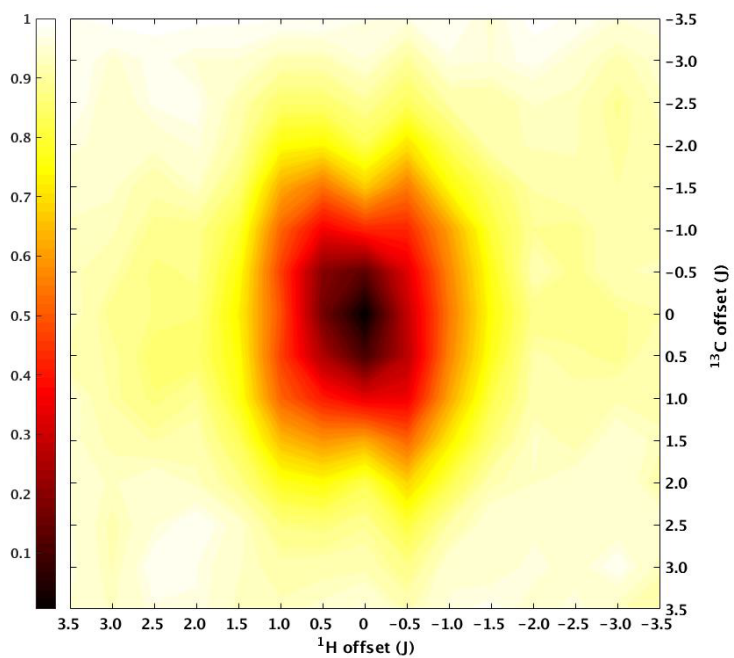


Fig. S3 Signal intensity contour plot of the acetate resonance as a function of SIERRA filter RF offsets. Both the ^1H and ^{13}C carriers were systematically varied along a grid of $0.5 \text{ } ^1J_{\text{HC}}$ (62.5 Hz).

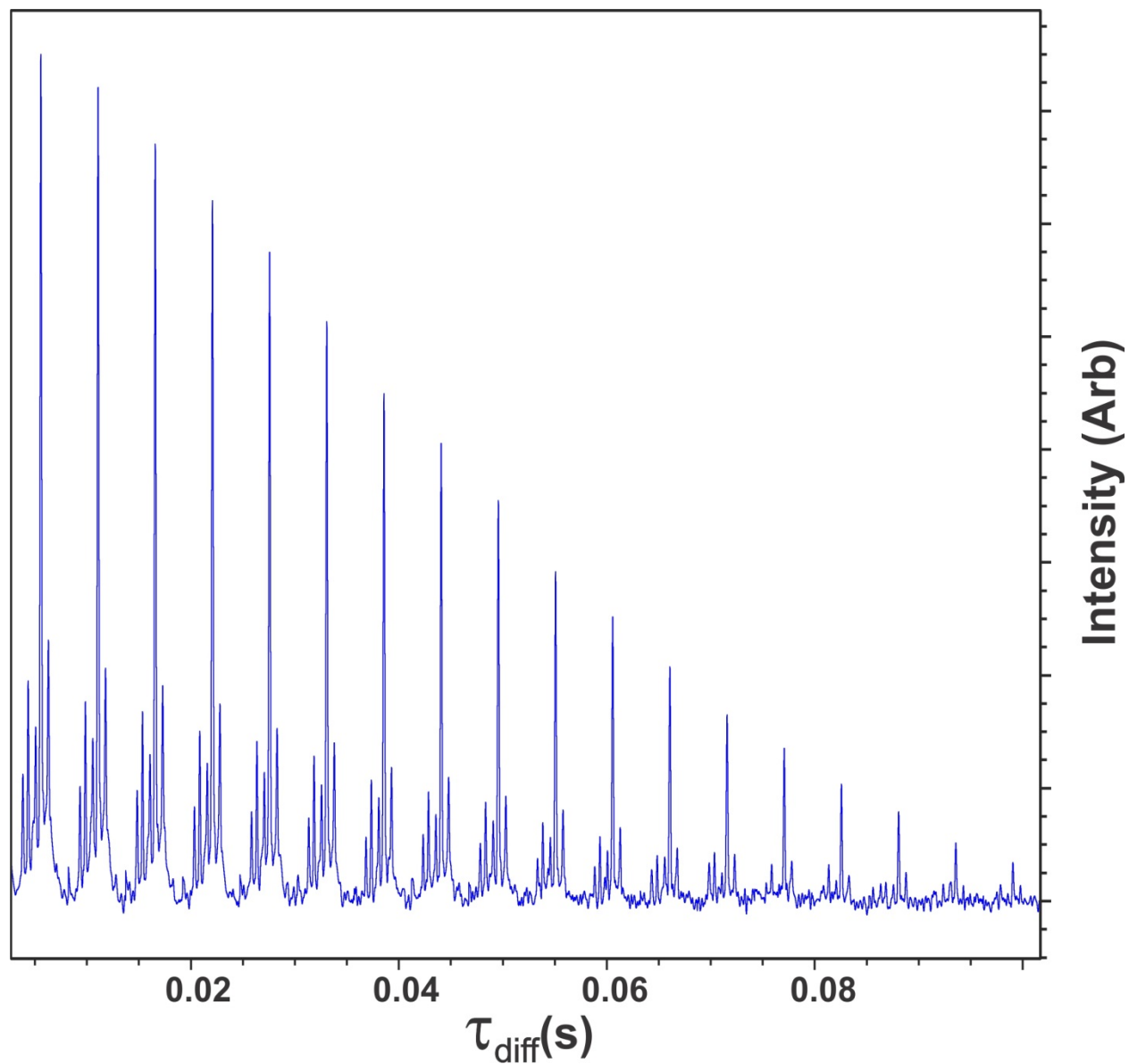


Fig. S4. Intensity profile of the polysorbate-80 signal as a function of the diffusion delay time in a 1D PFG-STE experiment. Even at 100 ms of diffusion time, residual polysorbate-80 signal remains.

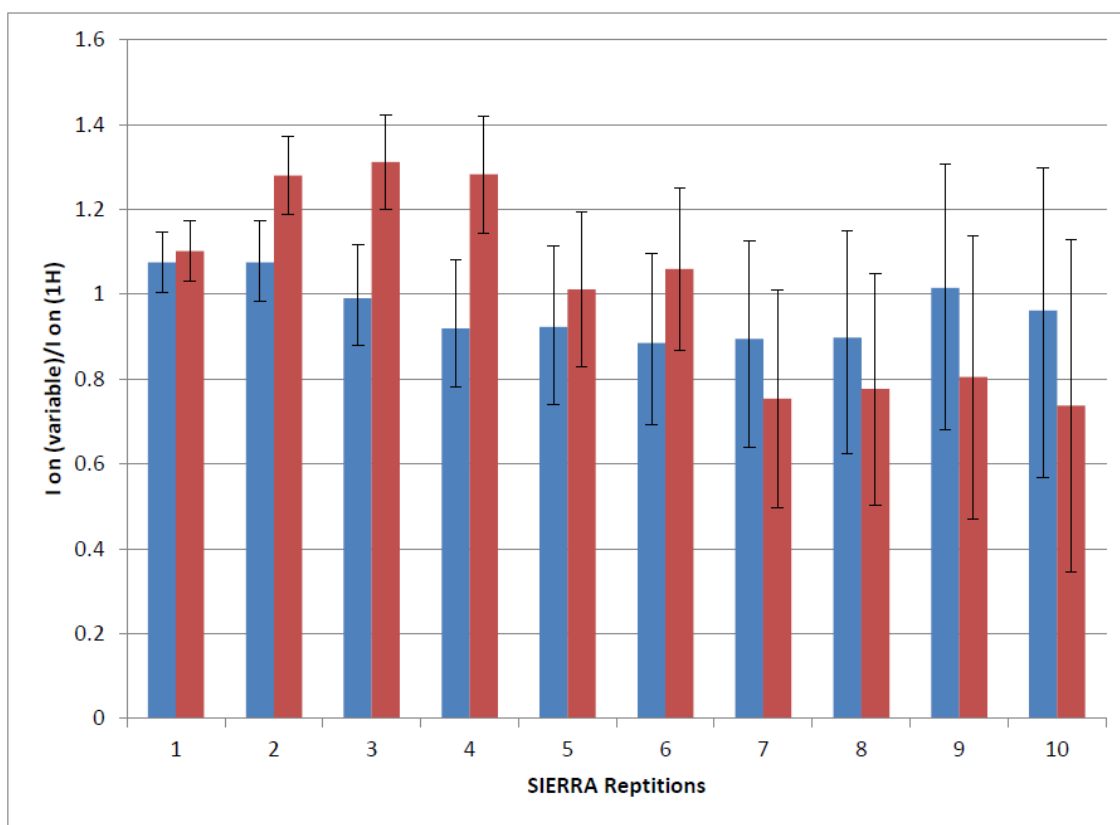


Fig. S5 Normalized signal intensity values as a function of SIERRA filter repetitions for the case when the ^1H and ^{13}C carrier are varied between repetitions. Blue bars correspond to $^1\text{H}/^{13}\text{C}$ carriers applied at: 1.96/23.24 ppm, 1.04/14.00 ppm, 1.42/23.41 ppm, 1.58/25.05 ppm, 2.25/24.61 ppm, 2.06/30.48 ppm, 2.19/20.76 ppm, 2.54/13.27 ppm, -0.30/20.50 ppm, and 0.19/31.50 ppm. Red bars correspond to $^1\text{H}/^{13}\text{C}$ carriers applied at: -0.30/20.50 ppm, 2.54/13.27 ppm, 0.19/31.50 ppm, 2.25/24.61 ppm, 1.04/14.00 ppm, 2.06/30.48 ppm, 1.42/23.41 ppm, 1.96/23.24 ppm, 1.58/25.05 ppm, and 2.19/20.76 ppm. In both cases, signal intensities were normalized to against the values for the case where the carrier was applied on resonance in ^1H for all repetitions.

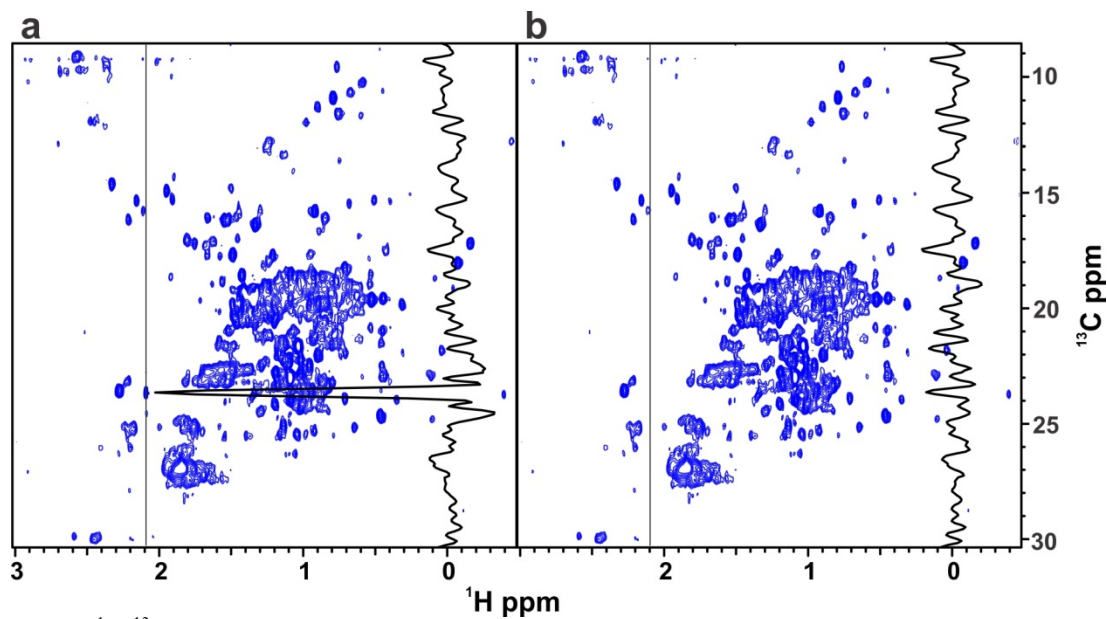


Fig. S6 ^1H - ^{13}C SIERRA-filtered gsHSQC spectrum of NISTmAb in 10 mmol/L sodium acetate at pH 5.2. **A)** Spectrum as recorded, the τ_{cp} delay was set to 4.2 ms resulting in incomplete attenuation of the acetate signal as shown in the 1D trace. **b)** Spectrum after SMILE subtraction of the residual acetate signal, resulting in a clean baseline without increase in the noise level as demonstrated in the 1D trace.

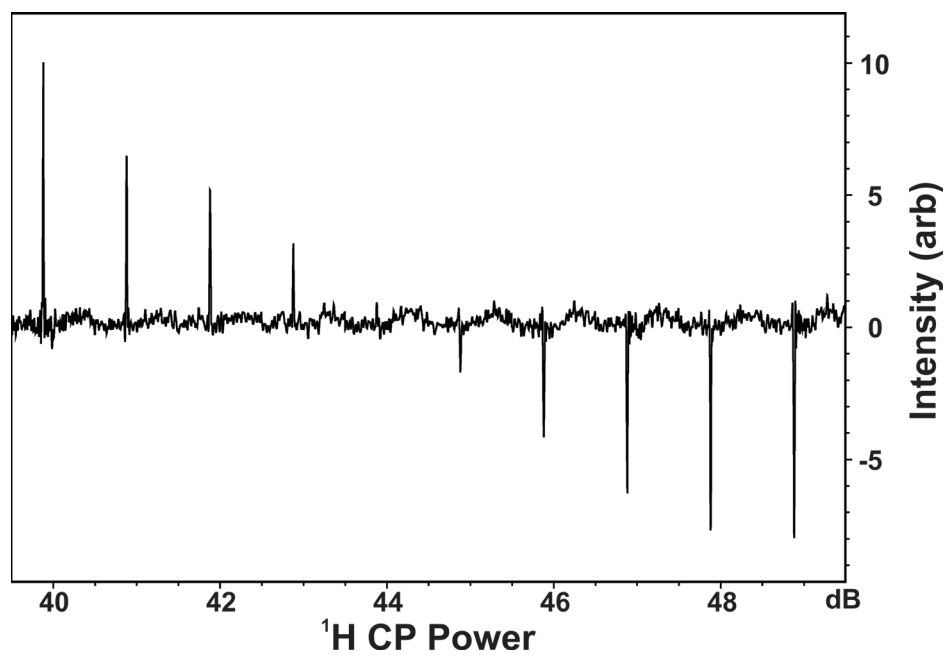


Fig. S7 Signal intensity of the acetate resonance as a function of the SIERRA filter ^1H CP RF power. ^{13}C RF nutation frequency was set according to eq. 1 and τ_{cp} was set to 11 ms.

References

- (1) Goddard, T.; Kneller, D. SPARKY 3. *University of California:San Francisco*, 2007.