Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

# Probing Invisible, Excited Protein States by Non-Uniformly Sampled Pseudo-4D CEST Spectroscopy**

*Dong Long,* Frank Delaglio, Ashok Sekhar, and Lewis E. Kay**

anie_201504070_sm_miscellaneous_information.pdf

*Sample Preparation.* An NMR sample of U-[$^{15}$N, $^{13}$C] G48A Fyn SH3 domain (2 mM protein, 0.2 mM EDTA, 0.05% $NaN_3$, 50 mM sodium phosphate, pH7.0, 90% $H_2O$/10 % $D_2O$) was prepared as described previously[1]. A 0.57 mM U-[$^{15}$N, $^{13}$C, $^{2}$H] L99A T4L sample was generated following the protocol described by Bouvignies *et al*.[2] Buffer conditions were 50 mM sodium phosphate, 25 mM NaCl, 2 mM EDTA, 2 mM $NaN_3$, pH 5.5, 90% $H_2O$/10 % $D_2O$. A 1.5 mM sample of U-[$^{15}$N, $^{13}$C] apoSOD1$^{2SH}$ dissolved in 20 mM HEPES pH 7.4, 1 mM TCEP, 1 mM $NaN_3$ and 90% $H_2O$ /10% $D_2O$ was prepared as described by Getzoff *et al*.[3]

*NMR Spectroscopy*. Conventional pseudo-3D HSQC-based $^{15}$N CEST[4] and pseudo-4D HNCO-based CEST (Figure S1A, CEST flag=$^{15}$N) experiments were recorded on G48A Fyn SH3 using a Varian Inova spectrometer, 11.7 T, 25 ℃ with a weak $B_1$ field of 26.3 Hz and a CEST mixing period, $T_{EX}$, of 200 ms. The positions of the weak $B_1$ field ranged from 102 ppm to 138 ppm with a step size of 25 Hz for both experiments. Sixty uniformly sampled complex points were recorded in the indirect dimension ($^{15}$N) for the pseudo-3D CEST experiment (4 transients, experimental time of 0.8 days). For the pseudo-4D data set an 8-by-8 complex point sampling matrix was used for $^{13}$CO and $^{15}$N dimensions along with 4 transients to give an experimental time of 1.7 days. In this application, as for the others listed below, a Poisson-gap sampling schedule was employed[5]. The first time point for $^{15}$N was always set to 0 and for $^{13}$CO chosen such that the total evolution time is $1/(2SW_{CO})$, where $SW_{CO}$ is the spectral width in the $^{13}$CO dimension.

Pseudo- 3D and 4D (Figure S1B, CEST flag=$^{15}$N) CEST experiments were measured for L99A T4L on a 14.0 T Varian Inova spectrometer (10 $^{o}$C) equipped with a cryogenically cooled probe. A weak ($^{15}$N) $B_1$ field of 28.5 Hz, ranging from 97 ppm to 138 ppm in step sizes of 30 Hz,

was applied for $T_{EX}$ = 700 ms in both experiments. Eighty uniformly sampled complex points were recorded in the indirect dimension ($^{15}$N) for the standard pseudo-3D CEST experiment (2 transients, experimental time 0.73 days), while a NUS matrix comprising 12-by-12 complex points for $^{13}$CO and $^{15}$N dimensions was used in the pseudo-4D experiment (4 transients, experimental time 5.3 days).

$^{15}$N and $^{13}$CO pseudo-4D CEST experiments (Figure S1A) were recorded on a sample of U-[$^{15}$N, $^{13}$C] apoSOD1$^{2SH}$, 25 ℃, 14.0 T. For the $^{15}$N experiment a weak $B_1$ field of 24 Hz was applied for $T_{EX}$ =350 ms over a frequency range extending from 98 ppm to 139 ppm and with a step size of 40 Hz. The $^{13}$CO CEST data set was acquired with a weak $B_1$ field of 25 Hz ranging from 169.5 ppm to 179.6 ppm and with a step size of 40 Hz ($T_{EX}$ of 350 ms). Sampling schedules of 13-by-13 complex points were used for $^{13}$CO ($t_1$) and $^{15}$N ($t_2$) chemical shift evolution, with total measurement times of 2.5 ($^{13}$CO CEST) and 4.0 ($^{15}$N CEST) days.

In all of the above experiments the $B_1$ field strengths were calibrated using the method of Guenneugues and Berthault[6] and a uniformly sampled reference HNCO spectrum (i.e. $T_{EX}$ = 0) was acquired from which $^{1}$H, $^{15}$N and $^{13}$CO chemical shifts of individual resonances were obtained.

*Data Processing*. Processing of pseudo-4D CEST data sets was achieved by initial Fourier transformation of the direct ($^{1}$H) dimension using the NMRPipe software package[7]. Subsequently, a set of 2D $^{13}$CO-$^{15}$N time domain planes was selected, one plane at each $^{1}$H frequency, $\omega_H$, of a peak in the HNCO 3D data set and analyzed by *NUSCEST* (a MATLAB program written in-house for NUS processing; available upon request). As described in the main text the experimental signal of interest can be modeled by Eq. [4] (including an additional 3

terms for the quadrature components of the signal). In practice a phase correction term, *f*, must be added to the $^{13}$CO dimension,

$$S_{i,k} = \sum_n \alpha_n \cdot e^{-R_n^{13CO} t_i^{13CO}} \cos(\omega_n^{13CO} t_i^{13CO} + f) \cos(\omega_n^{15N} t_k^{15N}) \quad \text{(S1)}$$

that can be obtained from a reference HNCO spectrum. In fits of the time-domain data the resonance frequencies and relaxation rates ( $\omega_n^{13CO}$ , $\omega_n^{15N}$ and $R_n^{13CO}$ ) are treated as global parameters that are invariant for all $B_1$ offsets in the optimization while the intensities, $\alpha_n$, depend on each $B_1$ offset. For a given set of parameters ( $\omega_n^{13CO}$, $\omega_n^{15N}$ and $R_n^{13CO}$) $S_{i,k}$ has a linear dependence on $\alpha_n$ and can therefore be expressed in matrix form as a set of coupled linear equations,

$$\mathbf{S} = \mathbf{K}\boldsymbol{\alpha} \quad \text{(S2)}$$

In the case where *p* and *q* $t_i^{13CO}$ and $t_k^{15N}$ points are recorded, respectively, **S** is a column vector with *j* elements (*j*=4*p*×*q*) that includes all quadrature components of the signal, **K** is a *j*-by-*N* matrix whose elements follow directly from Eq. [S1] and related equations where the cosine terms are replaced by sine terms, *N* is the number of cross-peaks in the $^{13}$CO-$^{15}$N plane chosen and **α** is a column vector with *N* elements. In principle, for each $^{13}$CO-$^{15}$N plane centered at a given proton frequency, $\omega_H$, cross-peaks are included in the fit if their proton chemical shifts lie within ±0.05 ppm of $\omega_H$. To speed up the computation, the minimization, $\min\left\|\mathbf{S}^{\text{expt}} - \mathbf{S}\right\|^2$, is executed as a two-step iterated process, whereby the intensities $\alpha_n$ are solved by singular value decomposition[8] for each proposed set of $\omega_n^{13CO}$, $\omega_n^{15N}$ and $R_n^{13CO}$ (Eq. [S2]). This set is then

optimized by the trust region reflective algorithm[9], with $\omega$ values allowed to vary within ±5 Hz of their starting positions and the process repeated.

CEST profiles extracted from pseudo- 3D and 4D experiments were fit to a two-state exchange model by numerical integration of the Bloch-McConnell equations,[10] including evolution due to all one-bond scalar couplings, as described previously[1]. Initially only those profiles with distinct minor dips or major dips with pronounced asymmetry where chosen to extract $p_b$, $k_{ex}$, which were then fixed in subsequent fits of all CEST profiles to obtain $\Delta\tilde{\omega}$ values. In this manner 28 and 14 CEST profiles were initially selected for G48A Fyn SH3 and L99A T4L, respectively. CEST data from SOD1 were analysed on a per-residue basis as described previously for this system[11].

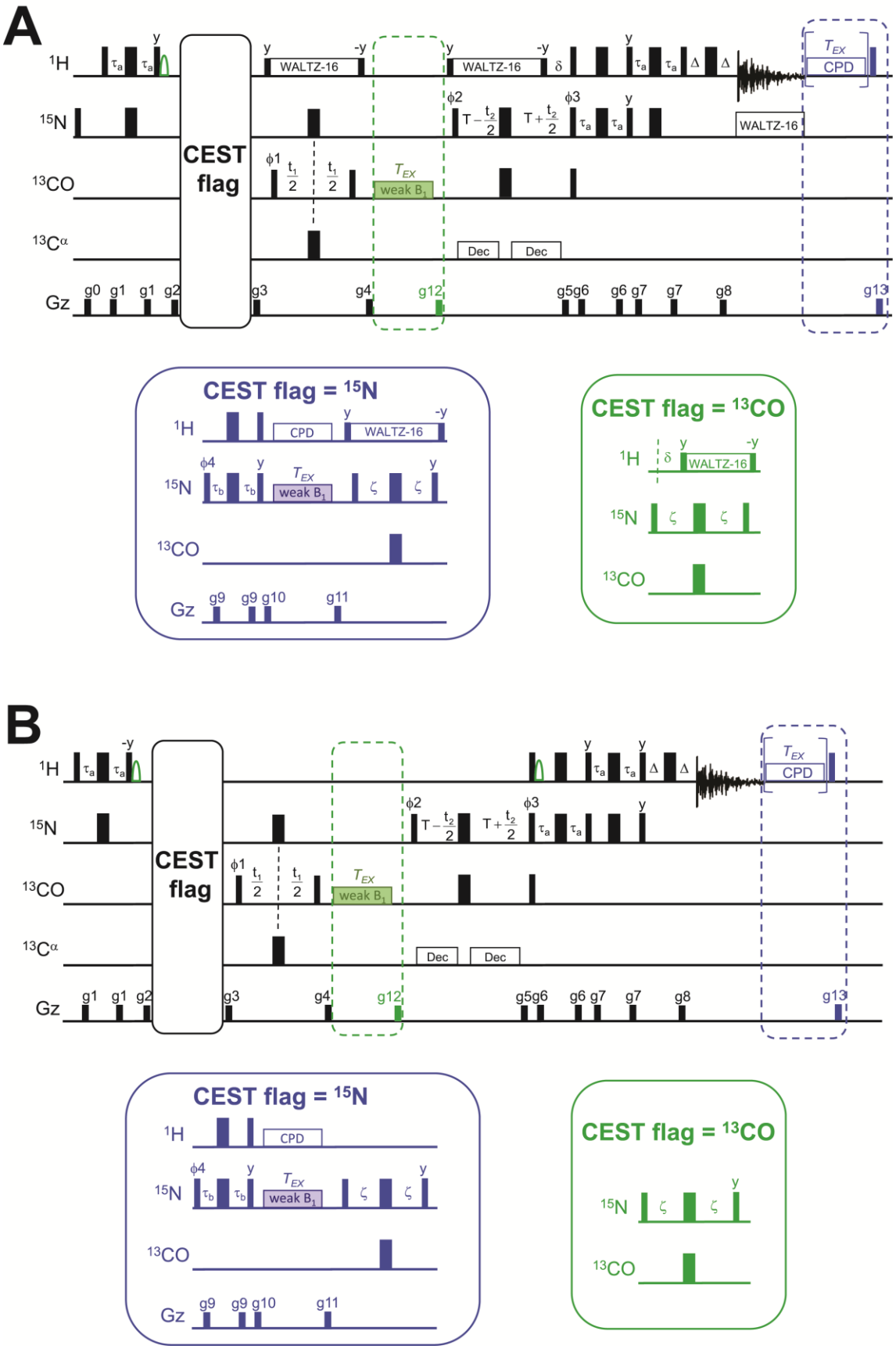
A
1H
15N
13CO
13Cα
Gz
CEST flag
WALTZ-16
CPD
Dec
weak B1
TEX
g0 g1 g1 g2 g3 g4 g12 g5g6 g6 g7 g7 g8 g13
CEST flag = 15N
CEST flag = 13CO
B
g1 g1 g2 g3 g4 g12 g5g6 g6 g7 g7 g8 g13
g9 g9 g10 g11

***Figure S1***. Pulse schemes for non-TROSY (A) and TROSY (B) versions of the HNCO-based pseudo-4D CEST experiment. Either $^{15}$N or $^{13}$CO CEST experiments are possible by selecting CEST flag = $^{15}$N or $^{13}$CO, respectively, and by including the elements in blue (or green) when CEST flag is set to $^{15}$N ($^{13}$CO). Note that the blue (green) dashed element is used for $^{15}$N ($^{13}$CO) CEST, while the green (blue) dashed element is removed. $^{1}$H, $^{15}$N, and $^{13}$CO carrier frequencies are centered at water, 118 ppm ($^{15}$N) and 176 ppm ($^{13}$CO) except during the CEST periods. During the $^{15}$N CEST period (CEST flag = $^{15}$N), the $^{1}$H carrier is placed at the center of the amide proton region with the $^{15}$N carrier positioned at distinct frequencies that are systematically varied. During the $^{13}$CO CEST period (CEST flag = $^{13}$CO), the $^{13}$CO carrier is varied from one HNCO data set to the next. Narrow and wide rectangles correspond to 90 ° and 180 ° pulses, respectively, applied along the *x*-axis unless otherwise indicated. All $^{1}$H pulses are applied at the highest power possible with the exception of the 6 kHz WALTZ-16 scheme[12] that is used for proton decoupling, the corresponding flanking $^{1}$H 90$^{o}$ pulses (A) and the $^{1}$H decoupling element during the $^{15}$N CEST delay period $T_{EX}$ where a $90_x240_y90_x$ composite pulse scheme[13] of 3 kHz is used. $^{1}$H shaped pulses are water selective and are of an approximate duration of 2 ms. $^{15}$N pulses are applied at the highest possible power, with the exception of the CEST element (CEST flag = $^{15}$N) that uses a very weak field of 10 – 50 Hz and for $^{15}$N decoupling during acquisition that is achieved using a 1 kHz WALTZ-16 field (A). $^{13}$C 90 ° and 180 ° rectangular pulses are applied with fields of $\xi/\sqrt{15}$ and $\xi/\sqrt{3}$ Hz, where $\xi$ is the frequency difference in Hz between the centers of the $^{13}$CO (176 ppm) and $^{13}$C$^{\alpha}$ (58 ppm) resonance frequencies[14]. Weak $^{13}$CO fields of between 25- 50 Hz are used for the $^{13}$CO CEST element. $^{13}$C$^{\alpha}$ decoupling is achieved using a WURST-2 adiabatic scheme[15] with a bandwidth from 44 to 66 ppm and a maximum (rms) RF amplitude of 0.59 (0.36) kHz (14.0 T). For $^{15}$N CEST applications a $^{1}$H decoupling

block at the end of each of the sequences (in parenthesis) is included for compensating the differential heating between the reference spectrum ($T_{EX}$ =0) and all other spectra for which $T_{EX}$ ≠0. This element is not present when the CEST element is used (i.e. $T_{EX}$ ≠0). Delays are: $\tau_a$ = 2.3 ms, $\tau_b$ = 2.72 ms, $\zeta$ = 12.4 ms, T = 12.4 ms, $\delta$ = 5.5 ms, $\Delta$ = 0.5 ms. The phase cycle employed is: $\phi_1$ = (*x*, -*x*), $\phi_2$ = *x*, $\phi_3$ = *x*, $\phi_4$ = (*x*, *x*, -*x*, -*x*), receiver = (*x*,-*x*) for CEST flag = $^{13}$CO, receiver = (*x*,-*x*,-*x*, *x*) for CEST flag = $^{15}$N. Gradient levels and durations (Gauss/cm; ms) are: g0 = (16; 0.5), g1 = (10; 0.5), g2 = (30; 2), g3 = (40; 0.75), g4 = (-24; 0.6), g5 = (60; 1.25), g6 = (10; 0.3), g7 = (20; 0.2), g8 = (59.1; 0.125), g9 = (-12; 0.3), g10 = (-24; 1), g11 = (-10; 0.5), g12 = (-10; 0.5), g13 = (36; 1). Quadrature detection in $F_1$ is achieved using States-TPPI[16] of $\phi_1$, while quadrature detection in $F_2$ makes use of the gradient enhanced sensitivity method[17,18] by recording a pair of data sets with ($\phi_3$, g5) and ($\phi_3$+180 °, -g5). For each successive $t_2$ point $\phi_2$ and the receiver phase are incremented by 180 °.[16]

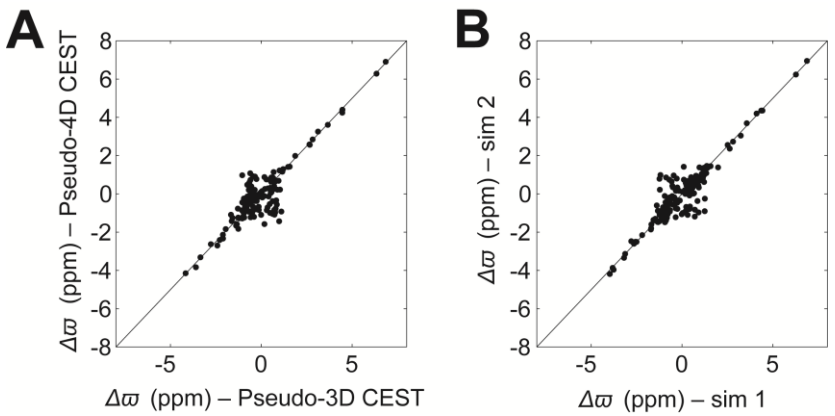


***Figure S2***. (A) Comparison of $\Delta\tilde{\omega}$ values extracted from uniformly sampled, Fourier transformed pseudo-3D CEST and from non-uniformly sampled pseudo-4D HNCO-based CEST for L99A T4L. (B) Comparison of $\Delta\tilde{\omega}$ values extracted from fits of synthetic CEST data sets. CEST profiles for each data set were generated directly via computation using experimentally measured L99A T4L exchange parameters and $\Delta\tilde{\omega}$ values, with different sets of random errors added to each computed profile based on the noise of the experimental profiles obtained from the pseudo-4D experiment. The resulting profiles were then fit following the same approach as for the experimental data, chemical shift differences extracted and plotted as shown. Note the very similar profiles in A and B indicating that the scatter observed for $|\Delta\tilde{\omega}| < 1$ ppm does not derive from artefacts in the NUS scheme but is an inherent limitation based on the width of the CEST dips and the inherent signal-to-noise of the CEST profiles.

In order to evaluate whether shorter measurement times could be used to acquire the pseudo-4D CEST data we have re-analyzed the data recorded on L99A T4L using only 50% of the experimental NUS complex points (corresponding to an experimental time of 2.65 days). Figure S3 compares the extracted $\Delta\tilde{\omega}$ values with those obtained from analysis of the complete NUS data set. Very similar correlations to those of Figure S2 are obtained, along with similar exchange parameters (see legend to Figure S3). This suggests that for many applications involving medium sized proteins 2-3 days acquisition will be sufficient to obtain robust pseudo-4D CEST data and CEST sampling schemes on the order of 10x10 complex points in ($^{13}$CO,$^{15}$N) dimensions will be sufficient.

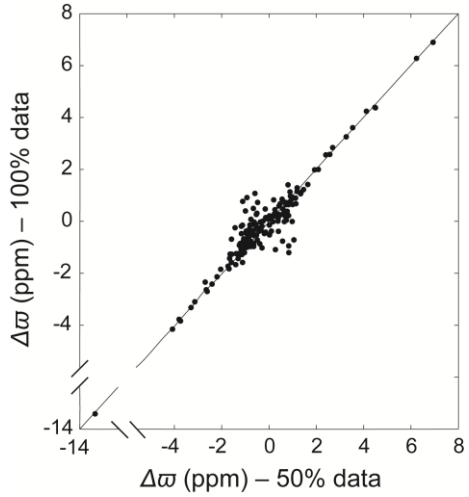


***Figure S3***. Comparison of $\Delta\tilde{\omega}$ values obtained from analyses of the complete NUS experimental pseudo-4D data set recorded on L99A T4L, 10 $^{o}$C, and a data set obtained by removing 50% of the ($t_1$,$t_2$) points. Fitted ($k_{ex}$,$p_b$) values from the complete and reduced data sets are (292±30 s$^{-1}$, 1.6±0.l %) and (330±40 s$^{-1}$, 1.6±0.l %) respectively. Note that the CEST profile for which $\Delta\tilde{\omega}$ = -13.4 ppm could not be obtained from 2D data.

The NUS processing method presented in this work is based on the direct fitting of time domain data and in order to obtain meaningful extracted parameters it is clear that the number of sparsely sampled data points must exceed the number of fitting parameters (for any given $^{15}$N-$^{13}$CO plane). It is of interest to establish the minimum number of sampling points that is required to accurately reproduce the obtained CEST profiles. In practice this remains a difficult question to answer because the minimum sampling number depends critically not only on the number of cross-peaks in the plane of interest, but also on spectral overlap, on the dynamic range of peaks, the exchange parameters themselves, as well as on the inherent signal-to-noise. Rather than address each of these issues in detail separately we have, instead, considered a single synthetic pseudo-4D CEST data set, and focussed on the set of $^{15}$N-$^{13}$CO planes corresponding to a given $\omega_H$, the amide $^1$H frequency, that are generated as a function of different weak $B_1$ offsets. In constructing this plane we have assumed the presence of 20 different resonances (*i.e.*, 20 cross-peaks) including the peak of interest that will be analyzed (denoted as P1), with the resonance position of P1 at ($\omega^{13}$CO,$\omega^{15}$N) = (177.8 ppm,116.0 ppm), and with the $^{13}$CO and $^{15}$N chemical shifts of the 19 other resonances randomly distributed within [171 ppm, 182 ppm] and [103 ppm, 133 ppm], respectively. It has not been assumed that the maxima of all 20 resonances are in the plane considered (*i.e*, at the chosen $\omega_H$); peak intensities were therefore randomly chosen to range between 0.1-1, although the intensity of P1 was set to 1 (in the absence of the weak $B_1$ field). The synthetic time domain signals, either with or without added noise, were subsequently analyzed using the NUS processing method described above, and the resultant CEST profiles for P1 compared with the expected profile in the absence of noise (red solid lines, Fig. S4). As can be seen in Fig. S4A, sampling as sparse as 0.38% (corresponding to 24 real data points) reproduces the correct CEST profile for an ideal noise-free case, while 0.25% of the data

(corresponding to 16 real data points) fails since the linear set of equations, Eq. [S2], becomes underdetermined (20 peaks and 16 points). In practice, significantly more data points must be sampled to achieve sufficient sensitivity than the minimal number based on the number of cross-peaks, but the exact number will be a function of several factors, as described above. Fig. S4B shows profiles extracted from the synthetic data with the addition of random noise, added at a level that is similar to what is observed in the experimental L99A T4L data. It is clear that sampling in the vicinity of 5-10% is required for an accurate reproduction of CEST profiles and subsequent robust extraction of exchange parameters ($p_b$, $k_{ex}$, $\Delta\omega$).

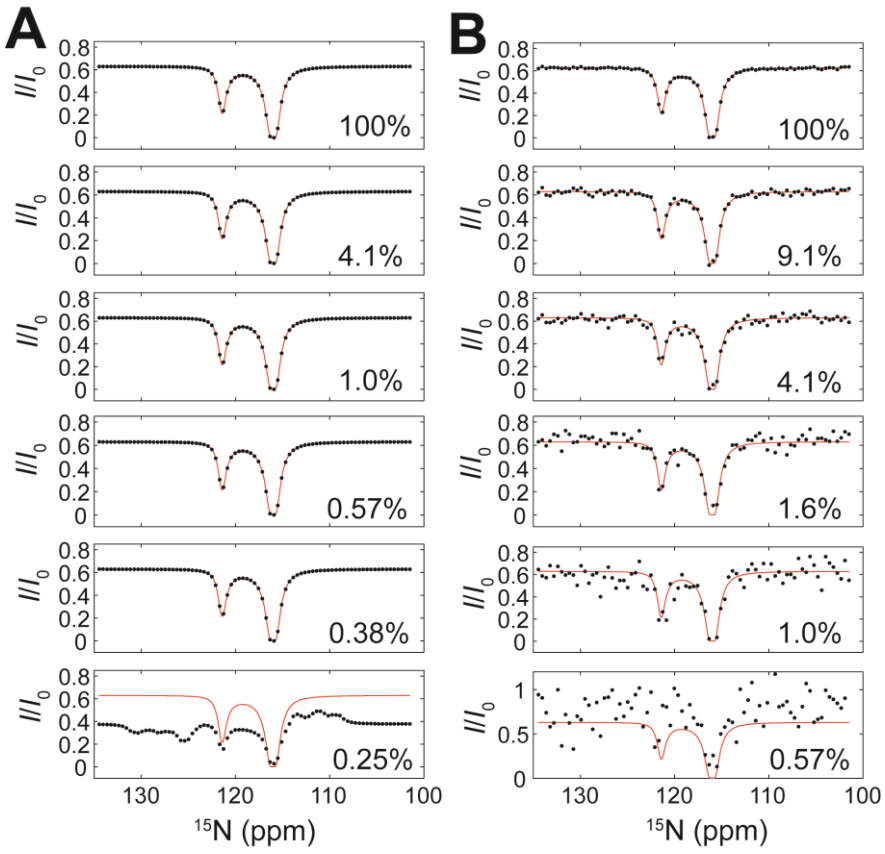


***Figure S4***. CEST profiles extracted from synthetic time domain signals without (A) and with (B) random noise added based on the average signal-to-noise of the experimental L99A T4L data set. All profiles were simulated for a data set constructed with 81 $B_1$ offsets, using a weak $B_1$ field of

15 Hz, $T_{EX}$ = 0.4 s, and assuming a two-site chemical exchange model with $p_b$ = 6% and $k_{ex}$ = 108 $s^{-1}$. The top panels (100% sampling) represent the cases of uniformly sampled synthetic data with 35×45 complex points ($^{13}$CO,$^{15}$N). The lower panels show results from sparse data sets taken from the full (uniform) data using Poisson-gap sampling schedules. The sparsely sampled complex points for individual panels are 12×12 (9.1%), 8×8 (4.1%), 5×5 (1.6%), 4×4 (1.0%), 3×3 (0.57%), 2×3 (0.38%), 2×2 (0.25%). The red solid lines represent the input CEST intensities for peak P1 used to generate the synthetic data, shown here for comparison.

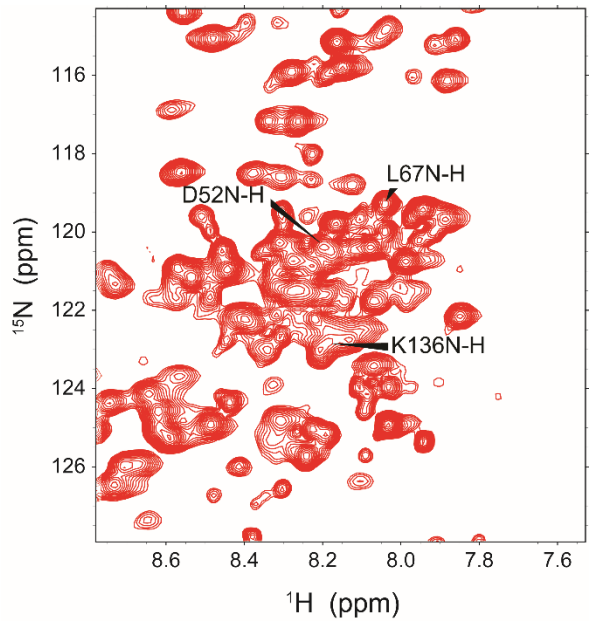


***Figure S5***. Selected region of the $^{1}$H-$^{15}$N HSQC spectrum of apoSOD1$^{2SH}$ showing the poor resolution that leads to difficulties in analysis of CEST data recorded using 2D data sets. Residues D52, L67 and K136, from which CEST profiles reporting on conformational exchange processes are readily obtained in pseudo-4D CEST, are labeled. The spectrum was acquired at 25 ℃, 14.0 T.

## References


[1] G. Bouvignies, P. Vallurupalli, L. E. Kay, *J. Mol. Biol.* **2014**, *426*, 763–774.

[2] G. Bouvignies, P. Vallurupalli, D. F. Hansen, B. E. Correia, O. Lange, A. Bah, R. M. Vernon, F. W. Dahlquist, D. Baker, L. E. Kay, *Nature* **2011**, *477*, 111–114.

[3] E. D. Getzoff, D. E. Cabelli, C. L. Fisher, H. E. Parge, M. S. Viezzoli, L. Banci, R. A. Hallewell, *Nature* **1992**, *358*, 347–351.

[4] P. Vallurupalli, G. Bouvignies, L. E. Kay, *J. Am. Chem. Soc.* **2012**, *134*, 8148–8161.

[5] S. G. Hyberts, K. Takeuchi, G. Wagner, *J. Am. Chem. Soc.* **2010**, *132*, 2145–2147.

[6] M. Guenneugues, P. Berthault, H. Desvaux, *J. Magn. Reson.* **1999**, *136*, 118–126.

[7] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, *J. Biomol. NMR* **1995**, *6*, 277–293.

[8] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical Recipes. The Art of Scientific Computing, 3rd Edition. Cambridge University Press.*, **2007**.

[9] T. F. Coleman, Y. Li, *SIAM J. Optim.* **1996**, *6*, 418–445.

[10] H. M. McConnell, *J. Chem. Phys.* **1958**, *28*, 430–431.

[11] A. Sekhar, J. Rumfeldt, H. R. Broom, C. M. Doyle, G. Bouvignies, E. M. Meiering, L. E. Kay, *Elife* **2015**, *in press*.

[12] A. . Shaka, J. Keeler, T. Frenkiel, R. Freeman, *J. Magn. Reson.* **1983**, *52*, 335–338.

[13] M. H. Levitt, R. Freeman, *J. Magn. Reson.* **1979**, *33*, 473–476.

[14] L. E. Kay, M. Ikura, R. Tschudin, A. Bax, *J. Magn. Reson.* **1990**, *89*, 496–514.

[15] Ē. Kupce, R. Freeman, *J. Magn. Reson. Ser. A* **1996**, *118*, 299–303.

[16] D. Marion, M. Ikura, R. Tschudin, A. Bax, *J. Magn. Reson.* **1989**, *85*, 393–399.

[17] L. E. Kay, P. Keifer, T. Saarinen, *J. Am. Chem. Soc.* **1992**, *114*, 10663–10665.

[18] J. Schleucher, M. Sattler, C. Griesinger, *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 1489–1491.