

Towards autonomous analysis of Chemical Exchange Saturation Transfer experiments using Deep Neural Networks

Gogulan Karunanithy¹, Tairan Yuwen², Lewis E Kay^{3,4,5,6}, and D Flemming Hansen^{1,*}

- 1) Department of Structural and Molecular Biology, Division of Biosciences, University College London, London, United Kingdom, WC1E 6BT
- 2) Department of Pharmaceutical Analysis & State Key Laboratory of Natural and Biomimetic Drugs, School of Pharmaceutical Sciences, Peking University, Beijing 100191, China
- 3) Department of Molecular Genetics, University of Toronto, Toronto, Canada, M5S 1A8
- 4) Department of Chemistry, University of Toronto, Toronto, ON M5S 3H6, Canada.
- 5) Department of Biochemistry, University of Toronto, Toronto, ON M5S 1A8, Canada.
- 6) Program in Molecular Medicine, Hospital for Sick Children Research Institute, Toronto, ON M5G 0A4, Canada.

Keywords: Artificial Intelligence, NMR Spectroscopy, Deep Learning, Chemical Exchange, CEST

To whom correspondence should be addressed:

D. Flemming Hansen; E-mail: d.hansen@ucl.ac.uk

Abstract

Macromolecules often exchange between functional states on timescales that can be accessed with NMR spectroscopy and many NMR tools have been developed to characterise the kinetics and thermodynamics of the exchange processes, as well as the structure of the conformers that are involved. However, analysis of the NMR data that report on exchanging macromolecules often hinges on complex least-squares fitting procedures as well as human experience and intuition, which, in some cases, limits the widespread use of the methods. The applications of deep neural networks (DNNs) and artificial intelligence have increased significantly in the sciences, and recently, specifically, within the field of biomolecular NMR, where DNNs are now available for tasks such as the reconstruction of sparsely sampled spectra, peak picking, and virtual decoupling. Here we present a DNN for the analysis of chemical exchange saturation transfer (CEST) data reporting on two- or three-site chemical exchange involving sparse state lifetimes of between approximately 3 - 60 ms, the range most frequently observed via experiment. The work presented here focuses on the ^1H CEST class of methods that are further complicated, in relation to applications to other nuclei, by anti-phase features. The developed DNNs accurately predict the chemical shifts of nuclei in the exchanging species directly from anti-phase $^1\text{H}^{\text{N}}$ CEST profiles, along with an uncertainty associated with the predictions. The performance of the DNN was quantitatively assessed using both synthetic and experimental anti-phase CEST profiles. The assessments show that the DNN accurately determines chemical shifts and their associated uncertainties. The DNNs developed here do not contain any parameters for the end-user to adjust and the method therefore allows for autonomous analysis of complex NMR data that report on conformational exchange.

Introduction

Many functional aspects of a macromolecule can be understood from its time-averaged three-dimensional structure. However, often the functionality of these molecules depends on their ability to exchange between different conformational states. Thus, quantifying the interconversion between these states is an important first step towards understanding how these biomolecules work¹⁻⁷. When conformational exchange is present, there is often one major populated state, the ground state, and a set of transiently low-populated states that, despite their low populations and short lifetimes, often play crucial roles for function. Several NMR techniques are now available to characterise reaction dynamics and transiently populated states at atomic resolution, including, chemical exchange saturation transfer (CEST)⁸⁻¹⁰, dark-state exchange saturation transfer (DEST)¹¹⁻¹³, Carr-Purcell-Meiboom-Gill (CPMG)¹⁴⁻¹⁶ relaxation dispersion, and relaxation in the rotating frame ($R_{1\rho}$, $R_{2\rho}$)¹⁷⁻¹⁹. CEST-based methods, which report on conformational exchange involving sparse states with lifetimes ranging from approximately 3 - 60 ms, have expanded tremendously over the last decade and have provided invaluable insights into the function of macromolecules²⁰. However, although several tools are available for the analysis of NMR data reporting on conformational exchange, challenges do exist, particularly when the exchange deviates from a simple two-state model²¹. For ¹H CEST methods reporting on the exchange of amide-protons²² and methyl-protons²³ analyses are further complicated by anti-phase features caused by the requirement to eliminate ¹H-¹H cross-relaxation effects, leading to broad lineshapes, with resolution significantly more limited than for ‘typical’ CEST profiles comprised of absorptive-like dips.

Deep learning and deep neural networks (DNNs) have led to huge advances in many fields of science, including computer vision and natural language processing, and the methodology is now a crucial component of many everyday technologies²⁴. In supervised deep learning, DNNs are trained to map an input to a desired output, and once trained, these networks can perform analyses autonomously. Deep learning is particularly successful at extracting features in complex data²⁵. It has been used for several years within the field of clinical magnetic resonance imaging (MRI) and some of the tools have already been approved by the FDA²⁶ for image enhancement and classification. Within biomolecular NMR there has been a surge in applications of DNNs over the last couple of years, and networks are now available

for the reconstruction of sparsely sampled spectra^{27–30}, peak picking³¹, estimating initial fitting parameters³², and virtual decoupling³³.

A key hurdle with many machine learning applications is that training robust models requires a large amount of curated training data. The in-depth understanding of the theory behind biomolecular NMR and the ability to simulate even complex NMR experiments means that the required amount of realistic training data can be generated synthetically. Importantly, it has now become clear that DNNs trained on fully synthetic data show robust performance on experimental data^{27,28,33}, which allows for sophisticated DNNs to be developed for the transformation and analysis of NMR spectra.

Overall, there is enormous potential for the development of deep learning approaches for the general analysis of NMR data and in particular for experiments reporting on conformational exchange. Below we have designed and trained DNNs to extract chemical shifts from the notably complex amide-proton anti-phase CEST experiment. The DNNs were trained solely on synthetically generated CEST profiles and are able to extract accurate chemical shifts of exchanging species as well as their uncertainties, thereby demonstrating that NMR data reporting on conformational exchange can be analysed autonomously using deep neural networks.

Methods

Deep Neural Network Architectures

Figure S2 shows the architecture for the DNN used to transform time-domain anti-phase CEST profiles into time-domain in-phase CEST profiles, DNN_{TR}. This architecture is built from two modules, a module akin to a block in the FID-Net architecture²⁸ and a modified LSTM module²⁷. The reason for this choice was that the main objective for the DNN is to ‘decouple’ anti-phase CEST profiles, which we have recently shown can be accomplished by the FID-Net architecture²⁸. The PYTHON code for generating the model architecture in Tensorflow/Keras is provided in Supporting Material. The input to the DNN consists of two vectors of size $2 \times 65 = 130$. The first vector, $\mathbf{cest}_{AP}(t) = \mathbf{c}_0$ holds the zero-filled real Fourier transform (real and imaginary components) of the antiphase CEST profile and the second vector holds the time-points associated with the first vector, \mathbf{t}_0 . The output of the network is the in-phase CEST profile, sampled at 128 offsets. The network contained 3,782,423 trainable parameters.

The second DNN, DNN_{CS}, used to determine chemical shifts and their confidences was built using a densely connected convolutional neural network architecture³⁴, Fig S4. The input

for the network is the output from the first transformation described above, that is, frequency domain data describing the in-phase CEST profile, $\text{cest}_{\text{IP}}(\omega)$, a vector of 128 real points. In its current form, the network detects a maximum of three chemical shifts as well as their confidences and the output of the network is therefore a 3×2 tensor, whose elements comprise three chemical shifts and their confidence values. The PYTHON code for generating the model in Tensorflow/Keras is provided in Supporting Material. Overall, the network has 1,591,526 trainable parameters.

Training the deep neural networks

The first DNN, DNN_{TR} , was trained on 15×10^6 anti-phase CEST profiles over 1500 *epochs*, where the range of training data is detailed in Table 1. An *epoch* refers to a single cycle of training of the neural network with training data. The training data was generated on-the-fly using code written in PYTHON and using functions from the Tensorflow and numpy libraries. To obtain smooth simulated CEST profiles, similar to those generated by experiment, previous simulations have used a distribution of B_1 fields or other dephasing methods⁸. Here the dephasing was achieved by only retaining the eigenvectors of the Liouvillian corresponding to real eigenvalues in the propagator. Thus, if \mathbf{L} is the matrix describing the Liouvillian, under which the spin-system evolves during the CEST period, then the eigenvalues and eigenvectors of \mathbf{L} are initially found: $\mathbf{L} \mathbf{\Lambda} = \mathbf{\Lambda} \mathbf{D}$, where $\mathbf{\Lambda}$ is a matrix of eigenvectors and \mathbf{D} is a diagonal matrix of eigenvalues. The submatrix of \mathbf{D} that holds the real eigenvalues is denoted \mathbf{D}_{re} and the matrix holding the eigenvectors corresponding to the real eigenvalues is denoted $\mathbf{\Lambda}_{\text{re}}$. Propagation of the spin-system is carried out with the propagator, $\mathbf{\Lambda}_{\text{re}} \exp(-T_{\text{ex}} \mathbf{D}_{\text{re}}) \mathbf{\Lambda}_{\text{re}}^{-1}$. As an example, for a simple Liouvillian, \mathbf{L} , represented by a 3×3 matrix in the basis set of the three product operators, I_x , I_y , and I_z there is typically only one real eigenvalue. After an eigendecomposition of \mathbf{L} , the matrix holding the eigenvectors, $\mathbf{\Lambda}$, and the diagonal matrix holding the eigenvalues, \mathbf{D} , are 3×3 matrices. The submatrix $\mathbf{\Lambda}_{\text{re}}$ has dimensions 3×1 , \mathbf{D}_{re} , is a 1×1 matrix, and $\mathbf{\Lambda}_{\text{re}}^{-1}$ is a 1×3 matrix. Thus, $\mathbf{\Lambda}_{\text{re}} \mathbf{D}_{\text{re}} \mathbf{\Lambda}_{\text{re}}^{-1}$ produces a 3×3 matrix that is the projection of the original Liouvillian onto the space spanned by the real eigensystem and $\mathbf{\Lambda}_{\text{re}} \exp(-T_{\text{ex}} \mathbf{D}_{\text{re}}) \mathbf{\Lambda}_{\text{re}}^{-1}$ is the propagator corresponding only to the real eigensystem. For the code written with the Tensorflow library functions, where sizes of matrices should remain constant, the dephasing is achieved by multiplying any eigenvalue that has an imaginary part larger than 10^{-3} by 10^9 , which means that evolutions caused by non-real eigenvalues are eliminated within nanoseconds.

The anti-phase CEST profiles were then obtained by propagating the Liouvillian over the first INEPT and the CEST element in the anti-phase $^1\text{H}^\text{N}$ pulse sequence. For each anti-phase CEST profile an in-phase CEST profile was also generated by setting $^1J_{\text{HN}} = 0$ Hz and integrating the Liouvillian over the CEST element⁸. The stochastic ADAM³⁵ optimiser was employed with standard parameters and an adaptive learning rate calculated as $0.0004 \times (L_{\text{freq}} + L_{\text{uncer}})^{3/4}$ (final learning rate of 10^{-6}). A batch size of 256 was used throughout the training and random gaussian noise was added with a standard deviation of 0.01 of the maximum value of each anti-phase CEST profile.

Table 1. Parameters used to generate training data.

<i>Experimental parameters</i>	
B_0	{14.1 T, 16.4 T, 18.8 T, 21.1 T, 23.5 T}
B_1	15 Hz to 50 Hz
Range of offset points	3.4 ppm
Sampled points	50 to 128
Inter-scan delay	0.5 s
CEST delay, T_{ex}	0.4 s
<i>Parameters reporting on the spin system</i>	
Rotation correlation, τ_{M} , used to calculate all relaxation rates of the ground state.	3 ns to 20 ns
^1H - ^{15}N scalar coupling, $^1J_{\text{HN}}$	-91 Hz to -95 Hz
Micro-second exchange contribution added to all states, R_{ex}	Absolute value of a normal distribution: $\mu=1.0 \text{ s}^{-1}$, $\sigma = 2.0 \text{ s}^{-1}$
$R_{2,\text{H}}(\text{E}_1) - R_{2,\text{H}}(\text{G})$ and $R_{2,\text{H}}(\text{E}_2) - R_{2,\text{H}}(\text{G})$	Normal distribution: $\mu=0.0 \text{ s}^{-1}$, $\sigma = 2.0 \text{ s}^{-1}$
<i>Chemical exchange</i>	
Probability of three-site exchange	25%
$k_{\text{ex,GE}}$, $k_{\text{ex,GE1}}$, $k_{\text{ex,GE2}}$	10 s^{-1} to 300 s^{-1}
p_{E} , p_{E1} , p_{E2}	0.01 to 0.15

After training the DNN_{TR} network the DNN_{CS} network was trained. The input data for training the DNN_{CS} network was obtained from output of the trained DNN_{TR} network. Random

gaussian noise with a standard deviation between 0.001 and 0.04 of the maximum value of each anti-phase CEST profile was added to anti-phase CEST profiles before these were transformed with the DNN_{TR} network. A total of 1.5×10^7 CEST profiles were used for training, which was done over 110 *epochs*, with a batch size of 128. The stochastic ADAM³⁵ optimiser with standard parameters and a learning rate of 3.3×10^{-4} was used.

Initial training was carried out using a desktop computer (Intel Core I7-6900K, 3.2 GHz, 64 GB RAM), equipped with an NVIDIA GeForce GTX 1080 TI GPU graphics card and subsequent training carried out using the CAMP cluster (NVIDIA Tesla V100 GPU).

Experimental amide-proton CEST data

A 1.5 mM U-[¹⁵N, ²H] L99A T4L sample produced as described previously³⁶ and dissolved in 50 mM sodium phosphate, 25 mM NaCl, 2 mM EDTA, 2 mM NaN₃, pH 5.5, 90% H₂O/10% D₂O was used to record the anti-phase ¹H^N CEST experiments. L99A T4L anti-phase ¹H^N CEST experiments were performed as described previously²². Briefly, the experiments were measured on a 800 MHz Bruker spectrometer equipped with an x,y,z-gradient cryogenically cooled probe. ¹H^N-CEST measurements were performed with a B_1 field of 30.5 Hz at 282 K using a CEST delay of $T_{\text{ex}} = 400$ ms. A range of ¹H offsets on a regular grid from 6.5–9.5 ppm was used, with step sizes of 30 Hz. An additional reference 2D dataset was obtained by setting the B_1 offset to -12 kHz.

A 1.35 mM sample of [U-¹⁵N,²H; Ile δ_1 -¹³CHD₂; Leu, Val-¹³CHD₂/¹³CHD₂; Met-¹³CHD₂] G48A Fyn SH3 domain was prepared as described previously²². The sample was dissolved in 50 mM sodium phosphate, 0.2 mM EDTA, 0.05% NaN₃, pH 7.0, 90% H₂O/10% D₂O. ¹H^N CEST experiments were measured for the G48A Fyn SH3 domain using a 600 MHz Bruker spectrometer at 285 K (x,y,z-gradient cryogenically cooled probe). The ¹H^N CEST datasets were recorded as described previously²²; specifically, a pair of datasets was recorded using B_1 fields of 26.7 Hz and 42.0 Hz. A CEST delay of $T_{\text{ex}} = 400$ ms was used and B_1 offsets between 5.5 and 10.5 ppm with step sizes of 25 Hz ($B_1 = 26.7$ Hz) or 40 Hz ($B_1 \sim 42$ Hz) were recorded. In addition, a 2D reference dataset was obtained with a B_1 offset of -12 kHz that is equivalent to setting $B_1 = 0$ Hz.

Results and Discussion

Chemical exchange saturation transfer profiles are normally visualised and analysed as, $I(\omega_{\text{offset}})/I_0$, where $I(\omega_{\text{offset}})$ is the intensity observed for a given site when a weak radio-frequency pulse (B_1) is applied at a frequency of ω_{offset} , and I_0 is the corresponding intensity with no B_1 pulse applied. A feature of standard CEST profiles is that they resemble inverted one-dimensional NMR spectra, where the ‘dips’ are centered at the chemical shifts of the exchanging species. Thus, the related CEST profile, $\max(I/I_0) - I/I_0$, resembles a simple NMR spectrum and its real Fourier transform therefore resembles an FID. Analysis of the CEST profiles with DNNs shown below first involved transformation of the data into the time domain, through a real Fourier transform, Figs. 1A,B. It should be noted that for a real Fourier transform, or equivalently a discrete Fourier transform of pure real data (N data points), the output is Hermitian-symmetric and approximately half ($N/2 - 1$ for even N and $(N - 1)/2$ for odd N) of the points are therefore redundant, see Supporting Material and Fig S1.

To show the strength of the developed DNNs for the analysis of CEST data, we consider the amide-proton anti-phase CEST²², whose profiles are complicated relative to those generated by other CEST experiments since the ‘dips’ are anti-phase in nature (*i.e.*, multiplet components from the scalar coupling between one-bond $^1\text{H-X}$ spins are of opposite phase). These CEST profiles are challenging to analyse primarily because the chemical shifts may not be easily accessible directly from the profiles. To facilitate the analysis of amide-proton CEST profiles the overall process is divided into two tasks, each with their own optimal DNN. The first DNN, DNN_{TR} , transforms each anti-phase CEST profile into a ‘classical’ profile, where the doublet nature of the dips are eliminated, thereby improving resolution, and also upsamples the profile to a fixed number of points in the CEST dimension. The second DNN, DNN_{CS} , then determines the ^1H chemical shifts for each of the exchanging species and an associated confidence in the shift values.

A deep neural network for the transformation of amide-proton CEST profiles

We have previously developed DNNs using the FID-Net architecture²⁸ to decouple and analyse NMR spectra^{28,33} by using FIDs as input. Since amide-proton anti-phase CEST profiles resemble anti-phase one-dimensional NMR spectra, our rationale was that a DNN similar to FID-Net can be trained to transform anti-phase CEST profiles into ‘decoupled’ standard CEST profiles. Thus, the DNN_{TR} architecture used was built of two modules, a module akin to a block in the FID-Net architecture²⁸ and a modified Long short-term memory (LSTM) module²⁷. The

architecture is described in detail in Supporting Material, Fig S2, where the PYTHON code for generating the model in Tensorflow/Keras^{37,38} is also provided. The theory for spin-evolution during CEST experiments is well-established^{8,39,40}, and synthetic training data can therefore easily be generated by propagating the Liouvillian over the desired element.

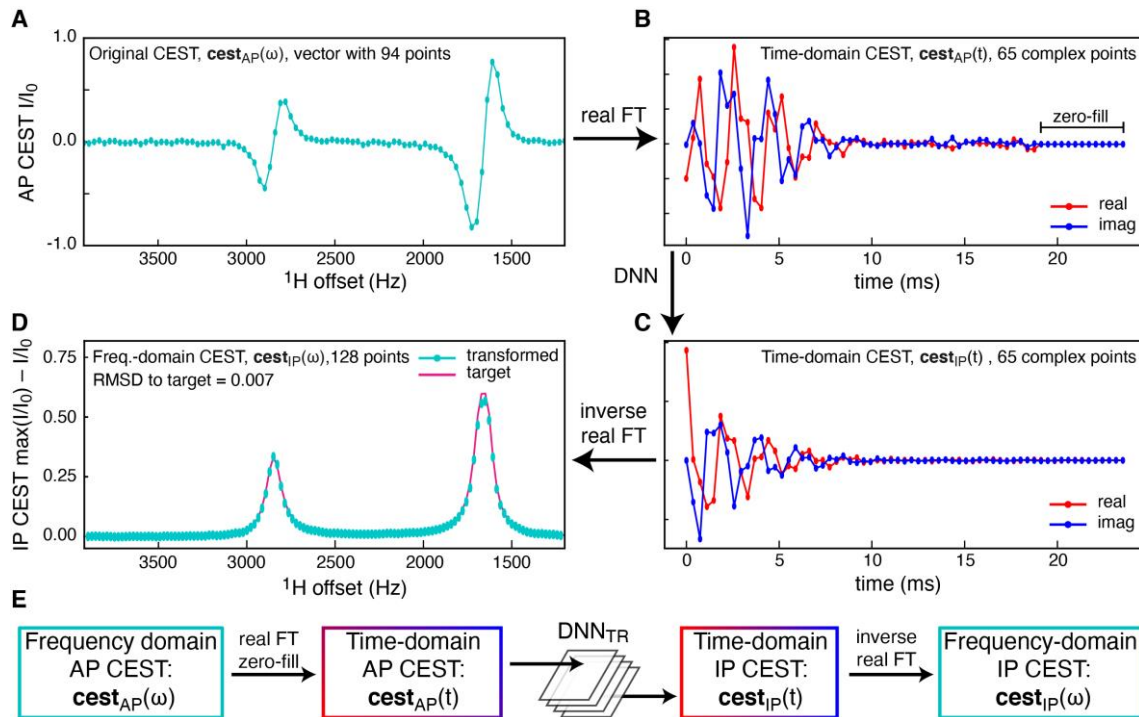


Fig. 1. Transformation of an amide-proton CEST profile from anti-phase to in-phase. Initially the input anti-phase CEST profile (A) is transformed with a real Fourier transform to give the time-domain CEST profile, followed by zero-filling (an additional 17 complex points) to generate a time-domain profile of 65 complex points (B), independent of the size of the original CEST profile. The DNN_{TR} network decouples the time-domain anti-phase CEST profile to give (C), which is transformed with an inverse real Fourier transform to give the final in-phase CEST profile in (D). (E) Schematic representation of the transformation from anti-phase CEST profiles to in-phase CEST profiles with a fixed size.

The first DNN, referred to as DNN_{TR} , was trained to transform an input amide-proton anti-phase CEST profile to the hypothetical CEST profile of an isolated 1H spin, with $^1J_{HN} = 0$ Hz, Fig 1. Thus, DNN_{TR} decouples the anti-phase amide proton CEST profile and upsamples it to 128 points. The upsampling to a constant size, in this case 128 real points, makes the prediction of chemical shifts with a second DNN feasible, since DNNs are typically trained with a constant size of the input and output data (see below). A maximum of three exchanging states was assumed and only the forked three-site exchange model was used to generate the data, that is, $E_1 \rightleftharpoons G \rightleftharpoons E_2$, where E_1 and E_2 are sparsely populated states. For 75% of the training data the population of E_2 was set to zero. Because of the strong correlation between

CEST data reporting on different three-site exchange models, for example, $E_1 \rightleftharpoons G \rightleftharpoons E_2$ *versus* $G \rightleftharpoons E_1 \rightleftharpoons E_2$, it is anticipated that DNN_{TR} will robustly transform anti-phase CEST profiles derived from any three-site exchange process. Briefly, DNN_{TR} was trained on 15×10^6 CEST profiles, where the range of training data is indicated in Table 1. The loss function was calculated from the mean-squared-error between the transformed in-phase CEST profile and the target function, see Fig 1D. The network was trained to a normalised mean-squared-error (MSE) of 4×10^{-4} and a mean-absolute-error (MAE) of 0.01.

The trained DNN_{TR} network was evaluated separately on synthetic data for two- and three-site exchanging systems. Figure 2 shows the evaluation on 100,000 randomly generated CEST profiles for two- (Fig. 2A) and three-site (Fig. 2B) exchanging systems. Figure S3 shows the performance of the DNN transformation as a function of the strength of the weak field, B_1 , the population of the sparse state E, p_E , the overall exchange rate, k_{ex} ($k_{\text{ex}} = k_{GE} + k_{EG}$, for two-site interconversion) and the number of sampled offsets. The transformation of profiles from anti-phase to in-phase by the DNN_{TR} network is robust and there is only limited variation in the performance with different parameters used to generate the CEST profiles. Of particular interest is that the transformation is only minimally affected by the number of points sampled in the input profile, Fig. S3D, suggesting that the upsampling is robust.

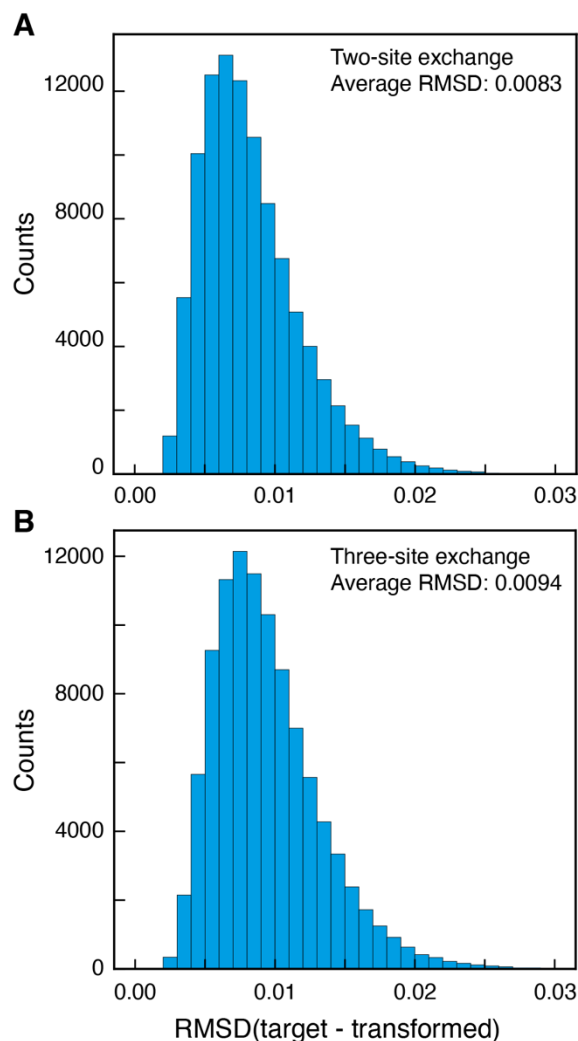


Fig. 2. Evaluation of the transformation of anti-phase ^1H CEST profiles to in-phase CEST profiles using a DNN. (A) RMSD between target and transformed in-phase CEST profile (see Fig 1). Statistics for 100,000 two-site exchange, $\text{G} \rightleftharpoons \text{E}$, CEST profiles, where $10 \text{ s}^{-1} \leq k_{\text{ex}} \leq 300 \text{ s}^{-1}$, $0.01 \leq p_{\text{E}} \leq 0.15$, $15 \text{ Hz} \leq B_1 \leq 50 \text{ Hz}$, $50 \leq \text{Sampled points} \leq 128$, $-95 \text{ Hz} \leq {}^1J_{\text{HN}} \leq -91 \text{ Hz}$, $B_0 \in \{14.1 \text{ T}, 16.4 \text{ T}, 18.8 \text{ T}, 21.1 \text{ T}, 23.5 \text{ T}\}$. (B) Statistics for 100,000 three-site exchange CEST profiles ($\text{E}_1 \rightleftharpoons \text{G} \rightleftharpoons \text{E}_2$), where $10 \text{ s}^{-1} \leq k_{\text{ex,E1}}, k_{\text{ex,E2}} \leq 300 \text{ s}^{-1}$, $0.01 \leq p_{\text{E1}}, p_{\text{E2}} \leq 0.15$, $15 \text{ Hz} \leq B_1 \leq 50 \text{ Hz}$, $50 \leq \text{Sampled points} \leq 128$, $-95 \text{ Hz} \leq {}^1J_{\text{HN}} \leq -91 \text{ Hz}$, $B_0 \in \{14.1 \text{ T}, 16.4 \text{ T}, 18.8 \text{ T}, 21.1 \text{ T}, 23.5 \text{ T}\}$.

Having evaluated the DNN_{TR} network on synthetic data it is important to assess how the DNN performs on experimental anti-phase $^1\text{H}^{\text{N}}$ CEST profiles. Figure 3 shows two examples, where $^1\text{H}^{\text{N}}$ anti-phase CEST profiles for the L99A mutant of T4 lysozyme recorded at 18.8 T have been transformed to in-phase CEST profiles (with the scalar coupling removed). This representation immediately allows estimation of the chemical shifts of ^1H nuclei of the exchanging states, which can be used as initial parameters for a least-squares analysis. However, these experimental CEST profiles are associated with uncertainty and since the

ground truth (exact value) is not known a detailed evaluation of the performance is not directly possible.

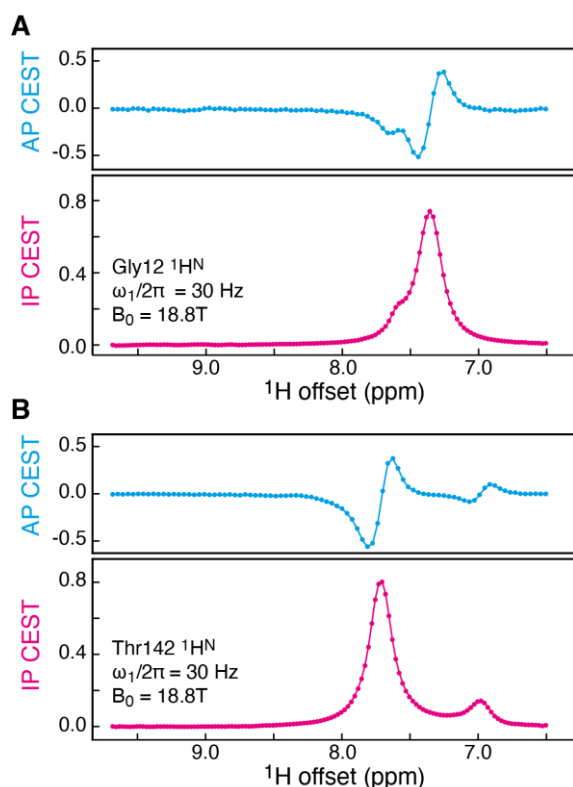


Fig. 3. Transformation of experimental anti-phase $^1\text{H}^{\text{N}}$ CEST profiles (AP CEST) recorded on a sample of the L99A mutant of T4 lysozyme into in-phase CEST profiles (IP CEST). The AP CEST profiles were recorded at a static magnetic field of 18.8 T, a temperature of 284 K, and using a 30 Hz ^1H B_1 field; 86 points were obtained in the CEST dimension. (A) Transformation and upsampling to 128 points of the anti-phase CEST profile for Gly12, (B) transformation and upsampling to 128 points of the anti-phase CEST profile for Thr142.

Determining ^1H chemical shifts in exchanging states using a deep neural network

With the in-phase CEST profiles available it becomes substantially easier to estimate the chemical shifts of the exchanging species. DNNs are particularly adept at locating specific features in data, for example, localising particular elements in an image. Thus, it is expected that a DNN could be trained to determine the position of peaks in one-dimensional NMR spectra and, consequently, trained to determine the chemical shifts of the exchanging species from in-phase CEST profiles or the related profiles, $\max(I/I_0) - I/I_0$. The densely connected convolutional neural network architecture³⁴, which was originally developed for object recognition tasks, was adapted here, Fig S4, to determine the chemical shifts from CEST profiles. Moreover, our goal was not only to determine the chemical shifts of the interconverting conformers, but to also train the DNN to estimate the uncertainties with which

it determined these shifts, thereby providing an output similar to a traditional least-squares fitting procedure.

The output from a DNN is typically a fixed length and a decision about the maximum number of exchanging states therefore has to be made before training the network. Since the time for training the DNN increases rapidly when increasing the maximum number of exchanging states, we chose for this application to only focus on CEST profiles reporting on three or less states, which covers most of the CEST-based studies reported to date. For a maximum of three exchanging states the output from the DNN_{CS} network is a 3×2 matrix whose elements are three chemical shifts, $f_{\omega,\text{pred}}$, and their corresponding confidences, c_{pred} . When the input CEST profile derives from a two-site exchanging system, the DNN should report one confidence approaching zero and when the input CEST profile is only reporting on one state, two of the confidences should tend to zero.

To facilitate an end-to-end analysis, that is chemical shifts and their uncertainties obtained directly from the experimental anti-phase CEST profiles, the network to determine chemical shifts was trained on outputs from DNN_{TR}, *i.e.* in-phase CEST profiles generated from anti-phase profiles. Having the second DNN, referred to as DNN_{CS}, determine both chemical shifts and their confidences requires special attention to the loss function used for training. Naturally, the DNN_{CS} network should be trained to optimise the confidence and thus obtain as accurate peak positions as possible, however, it should also be penalised, when the predicted confidence does not match the accuracy of the predicted chemical shifts.

The last layer of DNN_{CS} has sigmoidal activation, Fig S4, which means that the output values, three values reporting on chemical shifts and three confidences, are between 0 and 1. The predicted chemical shifts in the range (0,1), referred to as $f_{\omega,\text{pred}}$, are easily converted into the range of offsets obtained in the CEST dimension of the original data using a linear mapping. For example, if the CEST profile is recorded with points between 6.6 ppm and 10.0 ppm, then the linear mapping will be $\delta \leftarrow 3.4 \text{ ppm} \times f_{\omega,\text{pred}} + 6.6 \text{ ppm}$. Moreover, a predicted uncertainty, σ_{pred} , was calculated from the predicted confidence as $\sigma_{\text{pred}} = k(1/c_{\text{pred}} - 1)$, where k is a constant and σ_{pred} structured such that it can take values between 0 and infinity. In order to make the predicted uncertainties match actual uncertainties of the prediction, the first part of the loss function was defined in a manner similar to a standard χ^2 , that is:

$$L_{\text{freq}} = \sum_{i=0,1,2} \frac{(f_{\omega,\text{pred},i} - f_{\omega,\text{true},i})^2}{\sigma_{\text{pred},i}^2} \quad (1)$$

where the sum is over the three states. The constant k was initially set to 1 during training, and subsequently set to $(\max(^1\text{H offsets}) - \min(^1\text{H offsets}))\sqrt{L_{\text{freq}}}$ to rescale L_{freq} to have an expectation value of 1 and so that σ_{pred} reports on the expected uncertainty. The purpose of the loss function in Eq (1) is to make the predicted chemical shifts approach their true values. However, if L_{freq} was the only loss function used during training, then training of DNN_{CS} would simply lead to very low confidences (high uncertainties), which would minimise the function in Eq. (1). A second loss function was therefore added during training:

$$L_{\text{uncer}} = 10^{-4} \sum_{i=0,1,2} \mathbf{1}_i \sqrt{\sigma_{\text{pred},i}} \quad (2)$$

where, $\mathbf{1} = \{1,1,1\}$ for three-state exchange input and $\mathbf{1} = \{1,1,0\}$ in the case of two-state exchange, thereby allowing large uncertainties, σ_{pred} , when a state is not present in the input. The loss function in Eq. (2) serves to force DNN_{CS} to predict high confidences (low uncertainties) where, and only where, the input profiles report on a real state. Briefly, DNN_{CS} was trained on 1.5×10^7 randomly generated CEST profiles, with a final value of $L_{\text{freq}} = 7.3 \times 10^{-5}$, and $L_{\text{uncer}} = 2.8 \times 10^{-4}$. For the synthetic CEST data analysed below, the range of ^1H offsets was 3.4 ppm and therefore $k = 0.029$ ppm. Full details of the network architecture and the training are provided in the Methods and Supporting Information sections.

End-to-end one-shot analysis of amide proton CEST

The two DNNs, DNN_{TR} and DNN_{CS} , described above can be applied sequentially to provide an end-to-end one-shot analysis of anti-phase CEST profiles:

$$\begin{aligned} \text{AP-CEST, } \mathbf{cest}_{\text{AP}}(\omega) &\xrightarrow{\text{real FT, } \text{DNN}_{\text{TR}}, \text{ inverse FT}} \text{IP-CEST, } \mathbf{cest}_{\text{IP}}(\omega) \\ &\xrightarrow{\text{DNN}_{\text{CS}}} \{f_{\omega,\text{pred},i}, \sigma_{\text{pred},i}\}_{i=0,1,2} \end{aligned}$$

The overall performance of this sequential DNN was first evaluated using synthetically generated data. Specifically, (i) 100,000 anti-phase CEST profiles were generated for a variety of two-site chemical exchange processes and a further 100,000 profiles for three-site exchange. The range of B_1 offsets used was 3.4 ppm for all profiles, and all other input parameters are given in Table 1. (ii) Random gaussian noise with a standard deviation of 0.01 of the maximum

value of each anti-phase CEST profile was added to the input anti-phase CEST spectrum. (iii) The DNN_{TR} network was first used to transform all the CEST profiles from anti-phase to in-phase. (iv) The second network, DNN_{CS} , was used to determine the chemical shifts of the exchanging states and their associated uncertainties.

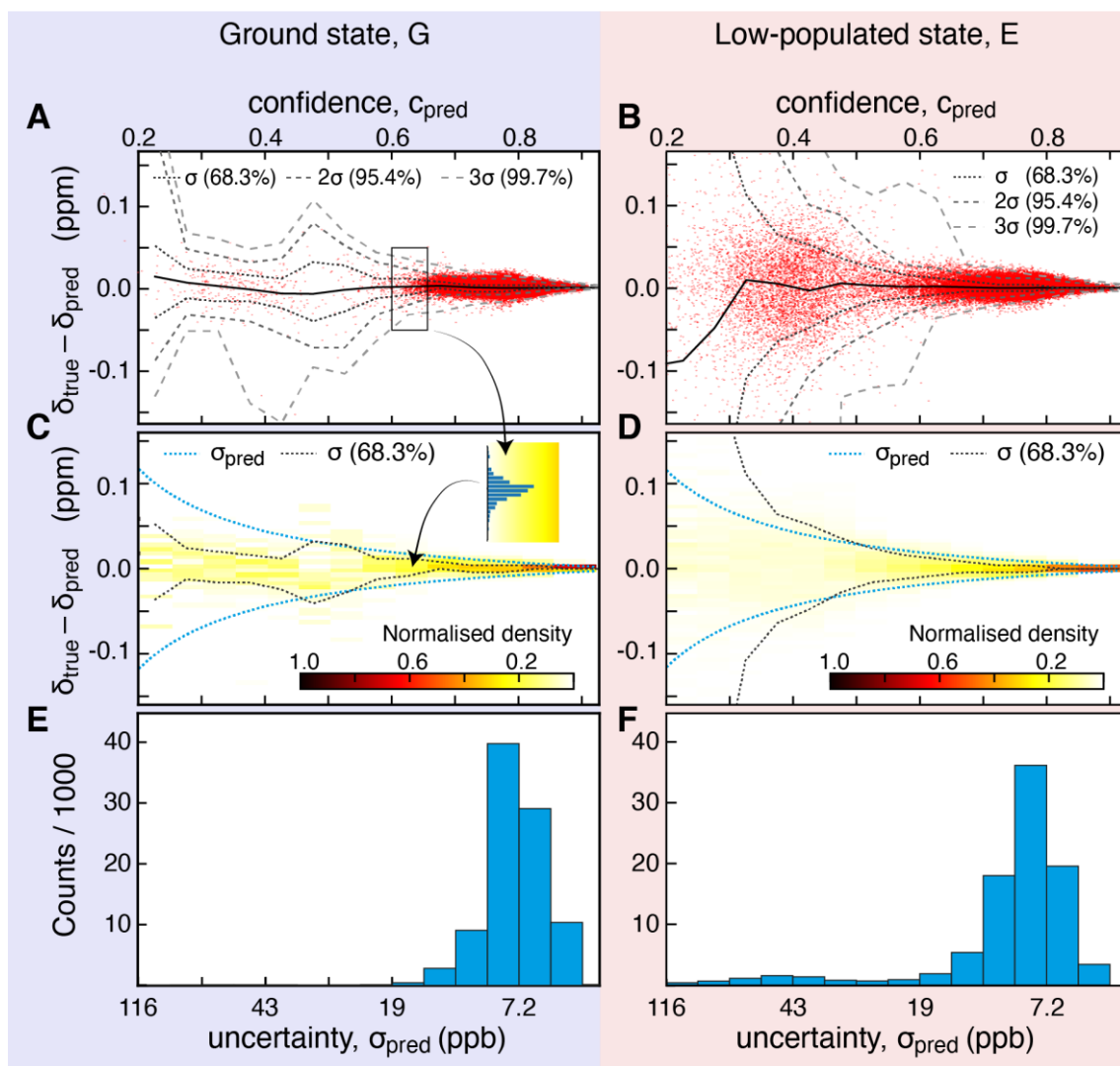


Fig 4. Quantitative assessment using 100,000 synthetic anti-phase $^1\text{H}^{\text{N}}$ CEST profiles reporting on a two-site chemical exchange and analysed using the sequential DNN to determine the chemical shifts of nuclei from the exchanging states. (A,C,E) assessment of the ground-state predictions, (B,D,F) assessment of the predictions of the low-populated state. (A) and (B) show differences between predicted (δ_{pred}) and true (δ_{true}) chemical shift values, *versus* c_{pred} for the 100,000 analysed CEST profiles (red dots). The full-drawn line corresponds to the average and the dashed lines correspond to the standard confidence levels, 68.3%, 95.4%, and 99.7%, respectively. (C) and (D) show 2D histograms of the points in (A) and (B); that is, a 2D histogram of the differences between δ_{pred} and δ_{true} *versus* the predicted confidence, c_{pred} . The histogram was calculated with a resolution of 0.05 along c_{pred} and 0.005 ppm along $\delta_{\text{true}} - \delta_{\text{pred}}$. The blue dashed lines show the predicted uncertainty, $\sigma_{\text{pred}} = 0.029(1/c_{\text{pred}} - 1)$,

which for $c_{\text{pred}} > 0.4$ agrees well with the confidence levels obtained from the analysis of the 100,000 profiles. (E) and (F) shows the distributions of uncertainties obtained from the assessment.

Figure 4 shows a summary of the quantitative assessment of the 100,000 CEST profiles corresponding to a two-state chemical exchange process. From Fig. 4 it is clear that the sequential DNN is able to accurately predict the chemical shifts of exchanging states from anti-phase CEST profiles. From the chemical shift predictions made on the 100,000 random CEST profiles the difference between a predicted chemical shift, δ_{pred} , and a true chemical shift, δ_{true} , was calculated, which gives an estimate of the performance and the confidence levels of the DNN as a function of c_{pred} and σ_{pred} . Importantly, as shown in Fig 4C and 4D, the DNN has also successfully been trained to predict the uncertainty associated with the predicted chemical shifts. Specifically, for $c_{\text{pred}} \geq 0.4$, the predicted uncertainty, σ_{pred} , agrees well with the 68.3% confidence level estimated from the analysis of the 100,000 profiles. For $c_{\text{pred}} < 0.4$, σ_{pred} is no longer an accurate measure of the uncertainty. Not surprisingly, the ground state chemical shifts, Fig 4E, are generally predicted with a higher accuracy than the chemical shifts of the low-populated state, Fig. 4F, where lower confidences are obtained for small chemical shift differences between the two states, see Fig S5. The corresponding assessment carried out on 100,000 synthetic anti-phase $^1\text{H}^{\text{N}}$ CEST profiles reporting on a three-site chemical exchange process, $\text{E}_1 \rightleftharpoons \text{G} \rightleftharpoons \text{E}_2$, is shown in Supporting Material, Fig S6. Figure S7 shows the summary of evaluations where random gaussian noise with a standard deviation of 0.01, 0.02, 0.04 of the maximum value of each anti-phase CEST profile was added to the input anti-phase CEST spectrum. The performance of the stacked DNN shown above strictly only holds for the ranges of data that were used for training and for the quantitative assessments, Table 1. However, as shown below, the performance of the DNN is rather robust and if the parameters of the CEST profile to be analysed deviate only slightly from the training parameters one would still expect the analysis to be valid. The ranges of parameters shown in Table 1 cover those obtained in most of CEST-based studies to date and it is therefore expected that most experimental anti-phase CEST profiles can be accurately analysed using the DNNs.

Assessment of the stacked DNN to analyse experimental CEST profiles

Experimental anti-phase ^1H CEST profiles for the L99A mutant of T4 lysozyme were analysed using the stacked DNN to gain insight into its performance on experimental data. As a validation of the performance of the fully stacked DNN two analyses were performed: in the first all of the 86 B_1 offsets were used to predict chemical shifts, while in the second, half of

the offsets (every second point) were removed. Figure 5A shows the example of Gly12, where the predicted chemical shifts and uncertainties using half of the B_1 offsets agrees well with the values obtained using the full dataset. Generally, this holds for all sites, Fig. 5B and the RMSDs obtained are in line with those expected from the predicted uncertainties, σ_{pred} . The differences in chemical shifts based on analyses of the full and half datasets, for all profiles, as a function of the confidence level are highlighted in Fig. 5C. Finally, it should be noted that the DNN_{TR} network was only trained on profiles with 50-128 input points. The fact that the stacked DNN is able to accurately predict the chemical shifts from profiles with less data (43 points) than those used for training points to the robustness of the DNN.

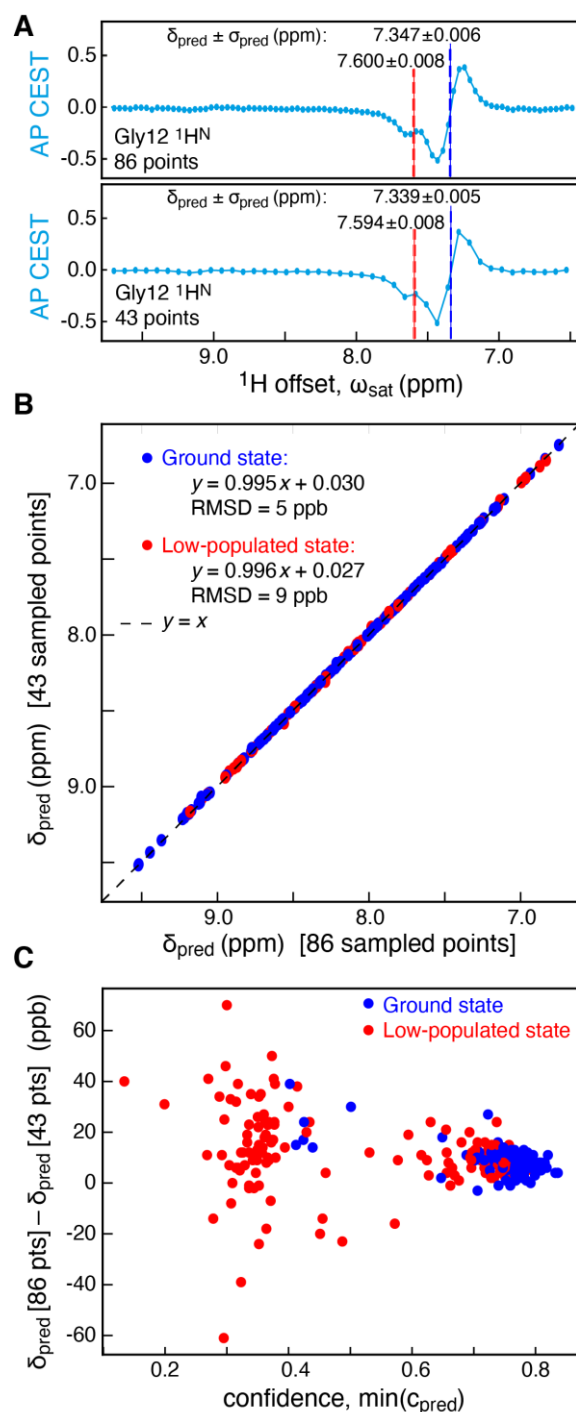


Fig 5. Predicting the chemical shifts of exchanging states of L99A T4 Lysozyme. The DNN for the transformation of anti-phase to in-phase profiles, DNN_{TR} , upsampled the recorded data to 128 points and DNN_{CS} determined the chemical shifts. Two full analyses were performed: One on the original 86 points recorded and another analysis on half of the data. (A) Analysis of the anti-phase profile for Gly12 $^1\text{H}^{\text{N}}$ emphasizes the robustness by which the sequential DNN determines chemical shifts and their predicted uncertainties. (B) Consistency plot showing excellent agreement between the chemical shifts determined from the full dataset (x -axis) and half of the data (y -axis). Only data for which $c_{\text{pred}} > 0.4$ are shown. (C) Differences between the predicted chemical shifts from the full dataset (δ_{pred} [86 pts])

and half of the data (δ_{pred} [43 pts]) versus the minimum of the confidence, $\min(c_{\text{pred}}) = \min(c_{\text{pred}} [86 \text{ pts}], c_{\text{pred}} [43 \text{ pts}])$. All data are included.

To further assess the performance of the stacked DNNs in determining the chemical shifts of the exchanging states, anti-phase CEST profiles were obtained for the G48A mutant of the SH3 domain from Fyn²². At a static magnetic field of 14.1 T, two sets of data were obtained with B_1 fields of 26.7 Hz and 42 Hz. Figure 6A shows that the chemical shifts predicted using the stacked DNNs, independently, on the two different datasets agree well (RMSD of 7 ppb), and Fig 6B highlights the difference in shifts based on the separate analyses of the two full datasets. Subsequently, the two experimental datasets were analysed simultaneously using a standard least-squares analysis²² with the software package ChemEx (<https://github.com/gbouvignies/chemex>) and the results were compared with the predictions made by the DNN, Fig 6C. Again, the agreement between the chemical shifts predicted by the DNN and those obtained by least-squares fitting agree well, with an RMSD of 7 ppb.

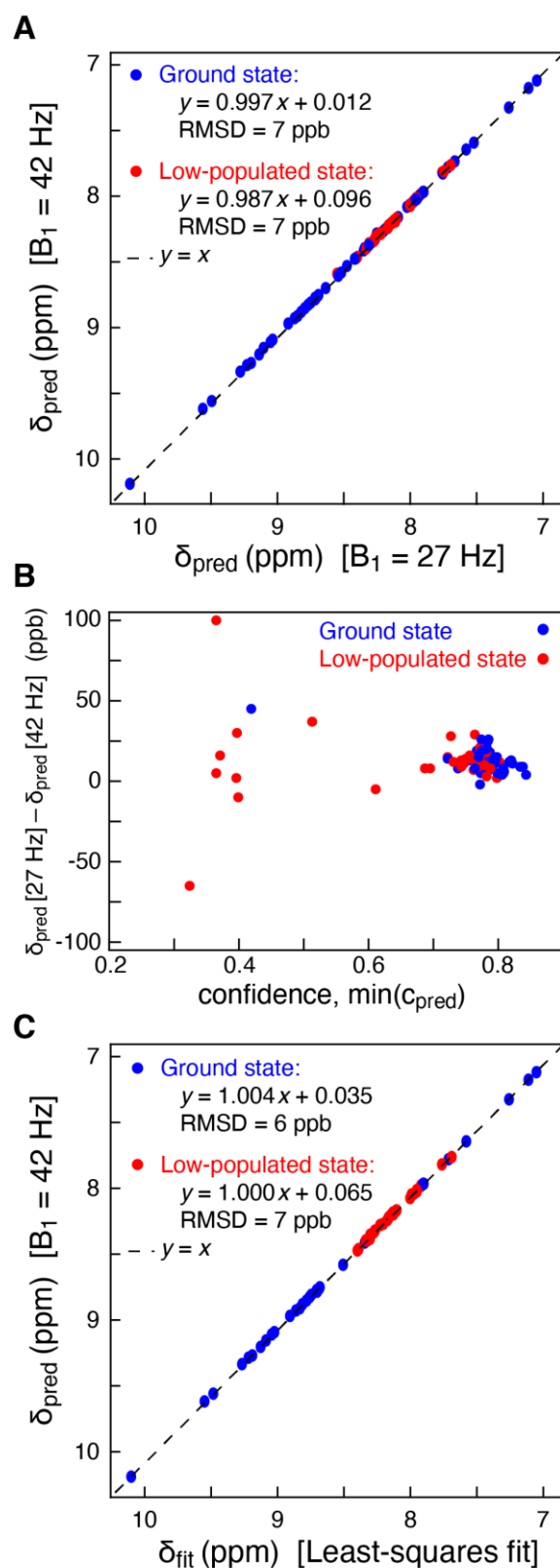


Fig 6. Predicting the chemical shifts of exchanging sites of the G48A mutant of the SH3 domain from Fyn. (A) Two datasets were recorded using different B_1 field strengths, 26.7 Hz and 42 Hz. The consistency plot shows that the chemical shifts determined independently from the two datasets agree. (B) Differences between the predicted chemical shifts from the two datasets ($\delta_{\text{pred}} [27 \text{ Hz}]$ and $\delta_{\text{pred}} [42$

Hz]) versus the minimum of the confidence, $\min(c_{\text{pred}}) = \min(c_{\text{pred}} [27 \text{ Hz}], c_{\text{pred}} [42 \text{ Hz}])$. All data are included. (C) Consistency plot showing that the chemical shifts determined using the stacked DNNs (one dataset) agree with the chemical shifts determined from least-squares fitting (two datasets). Only data for which $c_{\text{pred}} > 0.4$ are shown.

Uncertainties obtained from the covariance matrix in a least-squares analysis of CEST profiles are typically around 1 ppb, which is 6 times smaller than the uncertainties obtained from the DNN, indicating that the stacked DNNs have not fully reached the level of accuracy obtained by least-squares fitting. Still, the predictions obtained from the analysis with the stacked DNNs are of an accuracy where they can be used for downstream analyses and are well beyond the level of accuracy by which these shifts can be predicted from a high-resolution structure⁴¹. Alternatively, the DNN-predicted chemical shifts can serve as excellent starting parameters for a subsequent least-squares analysis. It is also possible that larger or alternative DNN architectures along with longer training periods could improve the performance of the DNN predictions.

Conclusions

A deep neural network was developed and trained to determine amide proton chemical shifts of exchanging states from anti-phase $^1\text{H}^{\text{N}}$ CEST profiles. The approach first leads to the conversion of anti-phase to in-phase $^1\text{H}^{\text{N}}$ CEST profiles, whereafter the chemical shifts are predicted along with their uncertainties. Compared with other analysis tools, the DNN does not require any additional training and there are no user adjustable parameters, which makes the analysis autonomous and suitable for automated processing pipelines. Thus far, the DNN only predicts chemical shifts. If additional parameters are sought, such as exchange rates and populations, the output shift values from the DNN can then serve as excellent starting points for a least-squares fitting procedure. The methodology and DNNs presented here add to the growing applications of deep learning and artificial intelligence for the analysis of NMR data, and provide an example of the autonomous analysis of complex NMR data reporting on macromolecular dynamics and chemical exchange.

Acknowledgements

T.Y. acknowledges post-doctoral support from the Canadian Institutes of Health Research (CIHR). Computational aspects of this work were supported by the Francis Crick Institute

(DFH) through provision of access to the Scientific Computing STP and the Crick data Analysis and Management Platform (CAMP). The Francis Crick Institute receives its core funding from Cancer Research UK (FC010233), the UK Medical Research Council (FC010233), and the Wellcome Trust (FC010233). DFH is supported by the Biotechnology and Biological Sciences Research Council UK (BBSRC) (ref: BB/T011831/1). LEK acknowledges support from the CIHR and the Natural Sciences and Engineering Council of Canada. LEK would like to dedicate this paper to the unbelievable goal scored by Connor McDavid of the Edmonton Oilers against the New York Rangers, November 5, 2021.

References

1. Boehr, D. D., Dyson, H. J. & Wright, P. E. An NMR perspective on enzyme dynamics. *Chem. Rev.* **106**, 3055–79 (2006).
2. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964–972 (2007).
3. Yang, H. *et al.* Protein conformational dynamics probed by single-molecule electron transfer. *Science* (80-.). **302**, 262–266 (2003).
4. Karplus, M. & Kuriyan, J. Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 6679–6685 (2005).
5. Xie, T., Saleh, T., Rossi, P. & Kalodimos, C. G. Conformational states dynamically populated by a kinase determine its function. *Science* (80-.). **370**, eabc2754 (2020).
6. Faust, O. *et al.* HSP40 proteins use class-specific regulation to drive HSP70 functional diversity. *Nature* **587**, 489–494 (2020).
7. Wurm, J. P. *et al.* Molecular basis for the allosteric activation mechanism of the heterodimeric imidazole glycerol phosphate synthase complex. *Nat. Commun.* **12**, 2748 (2021).
8. Vallurupalli, P., Bouvignies, G. & Kay, L. E. Studying ‘invisible’ excited protein states in slow exchange with a major state conformation. *J. Am. Chem. Soc.* **134**, 8148–61 (2012).
9. Zhou, J. & Zijl, P. C. M. van. Chemical exchange saturation transfer imaging and spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.* **48**, 109–136 (2006).
10. Ward, K. ., Aletras, A. . & Balaban, R. . A New Class of Contrast Agents for MRI Based on Proton Chemical Exchange Dependent Saturation Transfer (CEST). *J. Magn. Reson.* **143**, 79–87 (2000).
11. Fawzi, N. L., Ying, J., Ghirlando, R., Torchia, D. A. & Clore, G. M. Atomic-resolution dynamics on the surface of amyloid- β protofibrils probed by solution NMR. *Nature* **480**, 268–72 (2011).
12. Hansen, D. F. & Led, J. J. Determination of the geometric structure of the metal site in a blue copper protein by paramagnetic NMR. *Proc Natl Acad Sci USA* **103**, 1738–1743 (2006).
13. Bertini, I. *et al.* High-Field NMR Studies of Oxidized Blue Copper Proteins: The Case of Spinach Plastocyanin. *J. Am. Chem. Soc.* **121**, 2037–2046 (1999).
14. Meiboom, S. & Gill, D. Modified Spin-Echo Method for Measuring Nuclear

- Relaxation Times. *Rev. Sci. Instrum.* **29**, 688–691 (1958).
15. Tollinger, M., Skrynnikov, N. R., Mulder, F. A. A., Forman-Kay, J. D. & Kay, L. E. Slow dynamics in folded and unfolded states of an SH3 domain. *J. Am. Chem. Soc.* **123**, 11341–11352 (2001).
 16. Loria, J. P., Rance, M. & Palmer, A. G. A relaxation-compensated Carr-Purcell-Meiboom-Gill sequence for characterizing chemical exchange by NMR spectroscopy. *J. Am. Chem. Soc.* **121**, 2331–2332 (1999).
 17. Hansen, A. L., Nikolova, E. N., Casiano-Negroni, A. & Al-Hashimi, H. M. Extending the range of microsecond-to-millisecond chemical exchange detected in labeled and unlabeled nucleic acids by selective carbon R(1rho) NMR spectroscopy. *J. Am. Chem. Soc.* **131**, 3818–9 (2009).
 18. Palmer, A. G. & Massi, F. Characterization of the Dynamics of Biomacromolecules Using Rotating-Frame Spin Relaxation NMR Spectroscopy. *Chem. Rev.* **106**, 1700–1719 (2006).
 19. Chao, F.-A. & Byrd, R. A. Geometric Approximation: A New Computational Approach To Characterize Protein Dynamics from NMR Adiabatic Relaxation Dispersion Experiments. *J. Am. Chem. Soc.* **138**, 7337–7345 (2016).
 20. Vallurupalli, P., Sekhar, A., Yuwen, T. & Kay, L. E. Probing conformational dynamics in biomolecules via chemical exchange saturation transfer: a primer. *J. Biomol. NMR* **67**, 243–271 (2017).
 21. Neudecker, P., Korzhnev, D. M. & Kay, L. E. Assessment of the Effects of Increased Relaxation Dispersion Data on the Extraction of 3-site Exchange Parameters Characterizing the Unfolding of an SH3 Domain. *J. Biomol. NMR* **34**, 129–135 (2006).
 22. Yuwen, T., Sekhar, A. & Kay, L. E. Separating Dipolar and Chemical Exchange Magnetization Transfer Processes in 1 H-CEST. *Angew. Chemie Int. Ed.* **56**, 6122–6125 (2017).
 23. Yuwen, T., Huang, R. & Kay, L. E. Probing slow timescale dynamics in proteins using methyl 1H CEST. *J. Biomol. NMR* **68**, 215–224 (2017).
 24. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
 25. Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning*. (MIT Press, 2016).
 26. Chaudhari, A. S. *et al.* Prospective Deployment of Deep Learning in MRI: A Framework for Important Considerations, Challenges, and Recommendations for Best Practices. *J. Magn. Reson. Imaging* **54**, 357–371 (2021).
 27. Hansen, D. F. Using Deep Neural Networks to Reconstruct Non-uniformly Sampled

- NMR Spectra. *J. Biomol. NMR* **73**, 577–585 (2019).
28. Karunanithy, G. & Hansen, D. F. FID-Net: A versatile deep neural network architecture for NMR spectral reconstruction and virtual decoupling. *J. Biomol. NMR* **75**, 179–191 (2021).
 29. Qu, X. *et al.* Accelerated Nuclear Magnetic Resonance Spectroscopy with Deep Learning. *Angew. Chemie* **132**, 10383–10386 (2020).
 30. Luo, J., Zeng, Q., Wu, K. & Lin, Y. Fast reconstruction of non-uniform sampling multidimensional NMR spectroscopy via a deep neural network. *J. Magn. Reson.* **317**, 106772 (2020).
 31. Klukowski, P. *et al.* NMRNet: a deep learning approach to automated peak picking of protein NMR spectra. *Bioinformatics* **34**, 2590–2597 (2018).
 32. Beckwith, M. A., Erazo-Colon, T. & Johnson, B. A. RING NMR dynamics: software for analysis of multiple NMR relaxation experiments. *J. Biomol. NMR* **75**, 9–23 (2021).
 33. Karunanithy, G., Mackenzie, H. W. & Hansen, D. F. Virtual Homonuclear Decoupling in Direct Detection Nuclear Magnetic Resonance Experiments Using Deep Neural Networks. *J. Am. Chem. Soc.* jacs.1c04010 (2021) doi:10.1021/jacs.1c04010.
 34. Huang, G., Liu, Z., van der Maaten, L. & Weinberger, K. Q. Densely Connected Convolutional Networks. (2016).
 35. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. (2014) doi:1412.6980.
 36. Bouvignies, G. *et al.* Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* **477**, 111–117 (2011).
 37. Abadi, M. *et al.* TensorFlow: Large-scale machine learning on heterogeneous systems. (2015).
 38. Chollet, F. and others. Keras. (2015).
 39. Hansen, D. F., Vallurupalli, P., Lundstrom, P., Neudecker, P. & Kay, L. E. Probing chemical shifts of invisible states of proteins with relaxation dispersion NMR spectroscopy: how well can we do? *J Am Chem Soc* **130**, 2667–2675 (2008).
 40. Helgstrand, M., Hard, T. & Allard, P. Simulations of NMR pulse sequences during equilibrium and non- equilibrium chemical exchange. *J Biomol NMR* **18**, 49-63. (2000).
 41. Han, B., Liu, Y., Ginzing, S. W. & Wishart, D. S. SHIFTX2: significantly improved protein chemical shift prediction. *J. Biomol. NMR* **50**, 43–57 (2011).

