Probing excited state ^{1}Hα chemical shifts in intrinsically disordered proteins with a triple resonance-based CEST experiment: Application to a disorder-to-order switch

Ajith Kumar[a], **Kulkarni Madhurima**[a], **Athi N. Naganathan**[b], **Pramodh Vallurupalli**[c], **Ashok Sekhar**[a,*]

[a]Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, Karnataka, India

[b]Department of Biotechnology, Bhupat & Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras; Chennai 600036, India

[c]Tata Institute of Fundamental Research Hyderabad, 36/P, Gopanpally Village, Serilingampally Mandal, Ranga Reddy District, Hyderabad 500046

Abstract

Over 40 % of eukaryotic proteomes and 15 % of bacterial proteomes are predicted to be intrinsically disordered based on their amino acid sequence. Intrinsically disordered proteins (IDPs) exist as heterogeneous ensembles of interconverting conformations and pose a challenge to the structure-function paradigm by apparently functioning without possessing stable structural elements. IDPs play a prominent role in biological processes involving extensive intermolecular interaction networks and their inherently dynamic nature facilitates their promiscuous interaction with multiple structurally diverse partner molecules. NMR spectroscopy has made pivotal contributions to our understanding of IDPs because of its unique ability to characterize heterogeneity at atomic resolution. NMR methods such as Chemical Exchange Saturation Transfer (CEST) and relaxation dispersion have enabled the detection of 'invisible' excited states in biomolecules which are transiently and sparsely populated, yet central for function. Here, we develop a ^{1}Hα CEST pulse sequence which overcomes the resonance overlap problem in the ^{1}Hα-^{13}Cα plane of IDPs by taking advantage of the superior resolution in the ^{1}H-^{15}N correlation spectrum. In this sequence, magnetization is transferred after ^{1}H CEST using a triple resonance coherence transfer pathway from ^{1}Hα (i) to ^{1}HN(i+1) during which the ^{15}N(t_1) and ^{1}HN(t_2) are frequency labelled. This approach is integrated with spin state-selective CEST for eliminating spurious dips in CEST profiles resulting from dipolar cross-relaxation. We apply this sequence to determine the excited state ^{1}Hα chemical shifts of the intrinsically disordered DNA binding domain (CytR^N) of the bacterial cytidine repressor (CytR), which transiently acquires a functional globally folded conformation because of thermal fluctuations. The structure of the excited state, calculated using ^{1}Hα chemical shifts in conjunction with other excited state NMR restraints, is a three-helix bundle incorporating a helix-turn-helix motif that is vital for binding DNA.

---

*Corresponding author ashoksekhar@iisc.ac.in.

Keywords

Intrinsically disordered proteins; nuclear magnetic resonance spectroscopy; chemical exchange saturation transfer; conformational change; alpha proton; protein dynamics

1   Introduction

Intrinsically disordered proteins and protein regions (IDPs) are common in all living organisms and constitute over 40 % of the proteomes of eukaryotes (1). Despite lacking a stable three-dimensional structure, IDPs perform their function in a variety of different ways. Some IDPs fold in the presence of a binding partner (2–5) or upon post-translational modifications such as phosphorylation (6). Other IDPs participate in pathways where a unique structure is not required for function (5) and IDPs can remain disordered in their functional ligand-bound state (7). The amino acid sequences of IDPs are much simpler than those of their folded counterparts, being depleted in large hydrophobic aliphatic and aromatic amino acids like Leu, Ile, Val and Trp that promote order formation. Instead, IDP sequences are enriched in small polar and charged amino acids that include Ser, Gly, Arg and Glu which increase the tendency for disorder in polypeptide chains (8, 9). As a consequence, IDPs exist as structurally heterogeneous ensembles of interconverting conformations. The structural malleability of IDPs and their ability to switch rapidly between conformations enables them to interact with multiple structurally unrelated binding partners (10, 11) and IDPs frequently serve as scaffolds in biological processes such as signal transduction cascades (12, 13) and transcriptional regulation (14, 15). The conformational dynamics of IDPs is central for their function and characterizing the structures populated in the IDP ensemble is a vital step in understanding how IDPs function.

Nuclear magnetic resonance (NMR) spectroscopy is a powerful tool for probing the dynamics of biomolecules such as proteins, nucleic acids and glycans (16–18). The site-specificity afforded by NMR helps in constructing a coherent picture of conformational dynamics across the entire molecule of interest. A number of experiments such as Chemical Exchange and Dark-state Exchange Saturation Transfer (CEST (19–21) and DEST(22)), Carr-Purcell-Meiboom-Gill (CPMG) (23–25), $R_{1\rho}$ Relaxation Dispersion (RD) (26, 27) and Paramagnetic Relaxation Enhancement (PRE) (28, 29) have recently been developed that can detect and structurally characterize transiently ($\tau_E > 25$ μs) and sparsely ($p_E > 0.5$ %) populated biomolecular conformations (referred to as excited states (E) here). The application of these methods has revealed the importance of rare conformational transitions in protein folding (30–32), Hsp70 chaperone binding (33–35), light sensing by the LOV domain (36), enzyme catalysis by tyrosine phosphatases (37) and dihydrofolate reductase (38, 39), post-translational maturation of superoxide dismutase (40), DNA replication fidelity (41, 42) and transcription termination by the fluoride riboswitch (43), as well as in aberrant protein oligomerization (44, 45) and aggregation (46). However, these methods have found limited applicability in the domain of IDPs (47–50), primarily because IDPs are believed to have flat free energy landscapes devoid of kinetically separated excited state conformations (3, 49, 51).

We recently showed that the intrinsically disordered N-terminal DNA binding domain of the cytidine repressor (CytR), $CytR^N$ (CytR(1–66)), is functionally regulated by a conformationally excited state (52). CytR belongs to the LacR family of bacterial transcriptional repressors and regulates the transcription of genes involved in ribonucleotide uptake and degradation (53–56). $CytR^N$ has a high charge-to-hydrophobicity ratio as well as a large number of Ala and Lys residues, which cause it to be intrinsically disordered both in the free form and in full-length CytR (52). In the DNA-bound state, $CytR^N$ folds into a stable three-helix bundle which interacts with DNA through a helix-turn-helix (HTH) motif formed from the first two helices (H1 and H2) (57). While the conformation of $CytR^N$ in the DNA-bound state is very similar to that of its family members such as $FruR^N$ and $LacR^N$ (58, 59), the latter molecules are folded even in the absence of DNA (60–62). We used multinuclear ^{15}N, $^1H^N$, $^{13}C'$ and $^{13}C\alpha$ CEST experiments to show that the disordered ensemble of $CytR^N$ exists in equilibrium with a globally folded conformation that shares the same three-helix bundle topology found in the $CytR^N$-DNA complex (52). We obtained alpha carbon and carbonyl carbon chemical shifts of the excited state using previously reported triple resonance-based CEST experiments that were critical in defining the backbone fold of the $CytR^N$ excited state (63, 64). However, $^1H\alpha$ excited state chemical shifts could not be obtained because the existing aliphatic ^1H-CEST pulse sequence relies on quantifying residue-specific peak intensities in the 2D ^1H-^{13}C correlation spectrum (65), where there is severe overlap of resonances because of the disordered nature of native $CytR^N$ (Figure 1A). The CPMG relaxation dispersion-based method for measuring $^1H\alpha$ excited state chemical shifts also suffers from the same overlap problem (66). This is a significant impediment in characterizing millisecond timescale IDP dynamics because $^1H\alpha$ chemical shifts are sensitive reporters of protein secondary structure (67, 68).

Herein, we report a $^1H\alpha$ CEST pulse sequence (HACACONNH-HA-CEST) that combines spin-state-selective $^1H\alpha$ CEST acquisition with triple resonance coherence transfer that can be used for determining excited state $^1H\alpha$ chemical shifts in folded as well as disordered proteins. We validate this experiment on the L24A variant of the FF domain, in which exchange between the folded state and an on-pathway folding intermediate has been extensively characterized in literature (31). We then demonstrate the applicability of this sequence for IDPs by using it to measure $^1H\alpha$ chemical shifts of 37 residues in natively disordered $CytR^N$. The measurements reinforce the conclusion that disordered $CytR^N$ undergoes a disorder-to-order transition to an excited state structure composed of three helices and containing a helix-turn-helix motif.

2 Results and discussion

2.1 Designing the triple resonance $^1H\alpha$ CEST pulse sequence

The Chemical Exchange Saturation Transfer (CEST) experiment consists of the application of a weak radiofrequency (RF) field of amplitude B_1 at a specific chemical shift offset of the nucleus of interest ($^1H\alpha$ in our case) for an exchange duration T_{ex} (69). Longitudinal magnetization (19) or two-spin-order (34) is present at the beginning of the exchange period and the density operator of interest at the end of T_{ex} is quantified through peak intensities in suitable 2D correlation spectra. The RF field is swept across the offset range for the nucleus

under study and an array of 2D spectra, one for each offset value, is generated in the CEST experiment. The per-residue CEST profile is a plot of the normalized intensity (I/I_0) of the corresponding peak in the 2D spectrum as a function of the offset at which RF irradiation occurs. For a protein in exchange between a major ground state and a minor excited state, the presence of exchange in CEST profiles is visualized as a second minor dip in intensity besides the major saturation dip that occurs at the chemical shift of the nucleus in the excited state.

The chemical exchange saturation transfer (CEST) pulse sequence that is currently used for obtaining excited state aliphatic proton chemical shifts in proteins relies on quantifying peak intensities in a ^1H-^{13}C correlation spectrum (65). Figure 1A shows the $^1H\alpha$-$^{13}C\alpha$ region of the HSQC spectrum of the intrinsically disordered DNA binding domain of the *Escherichia coli* cytidine repressor (CytR^N). There is a lot of resonance overlap in this spectrum, which makes it impossible to reliably extract intensities for many of the $^1H\alpha$-$^{13}C\alpha$ correlations. In addition, it can be challenging to obtain the intensities of $^1H\alpha$-$^{13}C\alpha$ correlations because the $^1H\alpha$ resonances in IDPs cluster near the water resonance. In contrast, resonances are much better resolved in the ^1H-^{15}N HSQC spectrum of CytR^N (Figure 1B), making this 2D plane an attractive candidate for detecting intensity modulations from CEST.

In order to take advantage of the superior resolution in the ^1H-^{15}N 2D plane, our approach for determining excited state $^1H\alpha$ chemical shifts involves carrying out CEST on $^1H\alpha$ at the beginning of the pulse sequence and then transferring the magnetization from $^1H\alpha$ of the ith residue to the ^1HN of the i+1 residue through $^{13}C\alpha$(i), ^{13}CO(i) and ^{15}N(i+1) using one-bond scalar couplings (Figure 1C). The ^1H CEST profile recorded by directly 'saturating' the ^1H resonances as described earlier will often contain additional dips due to ^1H-^1H dipolar interactions (NOEs) with other proximal ^1H spins (70, 71). As these additional dips can mask the presence of a minor state dip and can also be accidentally assigned to a minor state resonance it is imperative to record ^1H CEST profiles that do not contain these spurious dips. Two approaches that exploit the presence of the attached X-nuclei (^{13}C in this case) have been devised to eliminate the NOE induced spurious dips. The first involves carrying X-spin CEST experiments with weak ^1H CW decoupling to indirectly detect the minor state ^1H resonance positions (70), while the second involves carrying out (X-) spin-state-selective ^1H CEST experiments (71). Here we have adopted the spin-state-selective approach to eliminate spurious NOE dips from $^1H\alpha$ CEST profiles.

Consider a $^1H\alpha$ nucleus (H_1) exchanging between two environments corresponding to the ground (G) and excited states of a protein (E) (Figure 1D). In addition, H_1 is dipolar coupled to a proton H_2 which is in close spatial proximity ($< \sim 5$ Å) to H_1 in the ground state. H_1 and H_2 are in-turn scalar coupled to their respective covalently bonded carbons C_1 and C_2. The top panel of Figure 1E (grey) shows the conventional $^1H\alpha$ CEST profile for this system quantifying the intensity of H_{1Z} in the ground state at the end of T_{ex}. The major dip from the ground state and the minor dip from the excited state are centred at the chemical shifts of $(H_1)_G$ (4.5 ppm) and $(H_1)_E$ (5.6 ppm) respectively. Each dip is split by the one-bond scalar coupling of H_1 to C_1 ($^1J_{C1H1}$) because there is no ^{13}C decoupling during T_{ex}; i.e., the major dip splits into dips 1 and 2, while the minor dip splits into dips 3 and 4. The dip arising due to H_2 centred at 3.5 ppm is also split into dips 5 and 6 due to the scalar coupling between

H_2 and C_2. Separating the CEST profiles based on the spin-state (α or β) of $(C_1)_G$ (blue (α) and red (β) in Figure 1E) results in profiles that contain a single dip arising due to H_1 rather than a doublet (dips 2 and 4 for the α spin-state and dips 1 and 3 for the β spin-state). Dips arising due to H_2 remain doublets as they arise due to the $^1J_{C2H2}$ coupling (i.e. the spin-state of C_2) rather than the $^1J_{C1H1}$ coupling. As the dips arising due to H_1 are displaced by $^1J_{C1H1}$ between the two spin-state-selective CEST profiles, subtracting the two profiles will result in antiphase features at the major and minor state H_1 shifts while dips 5 and 6 arising due to H_2 are subtracted out (Figure 1E, green profile). Following the approach detailed in Yuwen et al. (65), this subtraction is done 'in-line' in the HACACONNH-HA-CEST pulse sequence by selecting the $2H_ZC_Z$ operator at the end of T_{ex} since $2H_ZC_Z = H_zC^\alpha - H_ZC^\beta$. This improves the sensitivity of the overall experiment by a factor of $\sqrt{2}$ compared to acquiring separate CEST profiles for $(H_{1z}C_1^\alpha)_G$ and $(H_{1z}C_1^\beta)_G$ and subtracting them post-acquisition.

Figure 2 shows the HACACONNH-HA-CEST pulse sequence, in which CEST on $^1H\alpha$ is carried out at the beginning of the pulse train and $^1H\alpha$ magnetization is aligned along the z-axis at the beginning of the CEST element. The $2H_ZC_Z$ operator is then selected at the end of T_{ex} and magnetization is transferred to the $(i+1)^{th}$ amide proton using a standard coherence pathway (63, 72–74) taking advantage of scalar couplings $^1J_{C\alpha Ha}$, $^1J_{C\alpha CO}$, $^1J_{CON}$ and $^1J_{NH}$ in succession. The amide nitrogen and amide proton of the $(i+1)^{th}$ residue are frequency labelled in the indirect and direct dimensions respectively to extract CEST profiles of the ith $^1H\alpha$ nucleus using the high resolution ^1H-^{15}N 2D plane. The weak B_1 field (usually 5-50 Hz) is swept through the range of $^1H\alpha$ chemical shift offsets (typically 2.5 – 6 ppm) and one 2D plane is collected for each offset. The spacing between offsets is approximately the same as the amplitude of the weak RF field (in Hz), though it has been shown recently that significantly larger spacings can be employed without impairing the robustness of the exchange parameters extracted from the CEST data (75). A reference plane is collected for each CEST dataset using a modified version of the pulse sequence lacking the exchange duration ($T_{ex} = 0$), in which an INEPT pulse train replaces the CEST module for transferring H_z to $2H_zC_z$. Peak intensities in this reference plane (I_0) are used to generate the normalized CEST profile.

Protein CEST profiles on low-γ ^{13}C and ^{15}N nuclei are generally acquired by co-adding scans where the starting magnetization is along the +z and -z directions. This ensures that magnetization decays exponentially to 0 rather than to its equilibrium value of M_0, enabling an accurate 'in situ' measurement of R_1 and eliminating the dependence of the CEST profile on equilibrium magnetization, thereby reducing one fitting variable. It has been previously shown that such an approach is not optimal for ^1H CEST and results in smaller dips than if magnetization decays to M_0 (76). Therefore, $^1H\alpha$ magnetization is always aligned along the +z axis at the start of T_{ex} in the current pulse sequence (Figure 2); residue-specific $^1H\alpha$ longitudinal relaxation rate constants (R_{1H}) are separately estimated by varying T_{ex} using an altered version of the pulse sequence in which an INEPT module is inserted after T_{ex} to transfer magnetization from H_z to $2H_zC_z$, and no B_1 field is applied during T_{ex} in this experiment.

2.2 Testing the sequence on a system without measurable millisecond timescale exchange: ubiquitin:

We first tested the HACACONNH-HA-CEST pulse sequence on the folded protein (Figure 3A) ubiquitin that lacks measurable conformational exchange in the millisecond timescale. The signal-to-noise (S/N) of a reference 2D ^{1}H-^{15}N plane from the HACACONNH-HA-CEST experiment for ubiquitin (averaged over 50 isolated peaks) is 1.6 ± 0.5 – fold higher than a constant-time 2D ^{13}C-^{1}H correlation spectrum (averaged over 31 isolated peaks separated by at least 0.4 ppm from water). This is probably because magnetization evolves on the faster relaxing ^{13}Cα nucleus during the 29.4 ms constant time period in the ^{13}C-^{1}H HSQC, compared to evolution for 24.8 ms on the slower relaxing ^{15}N nucleus in the semi-constant period of the HACACONNH-HA-CEST sequence. In addition, the HACACONNH-HA-CEST sequence incorporates a gradient sensitivity enhancement module that improves sensitivity by a factor of $\sqrt{2}$. The sensitivity of the ^{1}Hα-^{13}Cα 2D correlation map can be further increased by decoupling ^{13}Cβ nuclei with adiabatic decoupling schemes and using a non-constant-time evolution element, as well as by including the sensitivity enhancement scheme for the ^{1}H-^{13}C AX spin pair (77). Nevertheless, it is apparent that the sensitivity disadvantage in using the HACACONNH-HA-CEST experiment is not significant especially for intrinsically disordered proteins that have smaller transverse relaxation rates compared to ordered proteins of the same mass.

Figure 3B shows the ^{1}H-^{15}N HSQC spectrum of ubiquitin and Figure 3C shows the ^{1}Hα CEST profiles obtained using the HACACONNH-HA-CEST pulse sequence by quantifying the intensity ^{1}H-^{15}N correlations of selected residues as a function of chemical shift offset. The ^{1}Hα CEST profiles of K27, Q31 and T66 appear like the difference in two Lorentzians separated by ^{1}J$_{C\alpha H\alpha}$, analogous to the spectrum of the antiphase operator $2H_zC_z$. We refer to this as the 'major dip' in correspondence with the in-phase major dip in ^{15}N or ^{13}C CEST profiles. While the maximum of the major dip occurs at the chemical shift offset of the TROSY component of the ^{1}Hα-^{13}Cα doublet, the minimum occurs at the chemical shift offset of the anti-TROSY component. The downfield component has negative intensity and the upfield component has positive intensity, unlike in Figure 1E, because the reference and CEST planes are phased opposite to each other in this sequence, with $2H_zN_z$ detected in the reference plane and $-2H_zN_z$ in the CEST planes. Notably, there are no NOEs observable in these three ^{1}Hα CEST profiles, confirming that the effect of dipolar cross-relaxation has been suppressed in this pulse sequence.

The ^{1}Hα CEST profile of G10 (Figure 3C, bottom right) shows the presence of two major dips in intensity at the chemical shifts of the two diastereotopic ^{1}Hα1 and ^{1}Hα2 of G10. Saturation transfer carried out on both ^{1}Hα atoms transfer to the same ^{1}HN of K11 and both major dips are observed in the same CEST profile. Therefore, this pulse sequence cannot be used to obtain excited state ^{1}Hα chemical shifts of Gly residues (78).

2.3 Validating the sequence on the L24A FF domain to characterize N-I exchange

Next, we tested the HACACONNH-HA-CEST pulse sequence on the L24A variant of the FF domain, which is known to exchange between its folded native state (N) and a partially folded on-pathway folding intermediate (I) that has a lifetime of 1.8 ms and a population of

5.4 % (303 K) (Figure 4A, (31, 79)). Since the exchange rate constant between these two states is ~ 580 s^{-1}, N↔I exchange can be conveniently studied with CEST. In addition, there is substantial conformational rearrangement in going from N, which is a four-helix bundle, to I, in which the third helix is longer than in N and the fourth helix is disordered (Figure 4A). A number of residues in H3 and H4 are therefore expected to show large ^{1}Hα chemical shift changes, furnishing a convenient testing system for the HACACONNH-HA-CEST pulse sequence.

Figure 4B shows the 2D ^{1}H-^{15}N correlation spectrum of L24A FF and Figure 4C shows ^{1}Hα CEST profiles of four residues L25, R29, E37 and K66 with large (> 0.4 ppm) chemical shift changes between N and I. In addition to the major dip in each profile, there is a minor dip consisting of two Lorentzians of opposite signs separated by ^{1}J$_{CαHα,E}$ (the J-coupling in the excited state) that arises due to conformational exchange between N and I. Since there is partial overlap between the TROSY and anti-TROSY components of the major and minor dips, all four extrema cannot be visualized separately in these CEST profiles.

In order to extract exchange parameters from the ^{1}Hα CEST profiles, we fit profiles of 22 residues with |ϖ_{IN}| > 0.35 ppm acquired at 3 B$_1$ field strengths each (15 Hz, 33 Hz and 54 Hz) globally to a two-state model of conformational exchange using the Bloch-McConnell equations (80). The fitting routine results in parameters of k$_{ex,IN}$ = 582 ± 6 s^{-1} and p$_I$ = 4.24 ± 0.04 %. While k$_{ex,IN}$ matches very well with the previously reported value (580 s^{-1}), p$_I$ is significantly lower (5.4 %). This is likely because of the impact of cross-relaxation on the CEST profiles (71). It is therefore preferable to measure k$_{ex}$ and p$_E$ from ^{15}N CEST experiments and use ^{1}Hα CEST data solely for determining excited state ^{1}Hα chemical shifts.

As the next step, we carried out residue-specific fits of ^{1}Hα CEST profiles of 50 residues to extract chemical shift differences between the native and intermediate states. During this step, k$_{ex,IN}$ and p$_I$ were fixed to the value obtained from the global analysis of ^{1}Hα CEST data (see Materials and Methods). Figure 4D shows a comparison of these ϖ_{IN} values with those determined using CPMG relaxation dispersion methods (31). There is very good agreement between the two measurements, with an RMSD over 50 residues of 0.07 ppm, confirming the robustness of the HACACONNH-HA-CEST pulse sequence.

2.4 Measuring ^{1}Hα chemical shifts of the excited state of intrinsically disordered CytR^N

Having established the reliability of the HACACONNH-HA-CEST pulse sequence for measuring excited state ^{1}Hα chemical shifts, we applied the methodology to the disorder-to-order transition in CytR^N (Figure 5A). The native state of CytR^N is intrinsically disordered, as seen from both sequence-based metrics such as the amino acid composition, charge and hydropathy, as well as from circular dichroism and NMR spectroscopy. We recently showed (52, 81) that CytR^N transiently populates a higher free energy state in which it adopts a globally folded three-helix bundle structure that resembles the conformation of the DNA-bound protein. We also showed that the folded conformation is responsible for binding DNA and that there is no measurable reactive flux directly from the native disordered state to DNA-bound CytR^N (52).

The structure of the globally folded excited state of CytR[N] was calculated using backbone [1]HN, ^{15}N, [13]Cα and [13]C' chemical shifts derived from CEST experiments. [13]Cα and [13]C' chemical shifts were determined using previously published triple resonance schemes (63, 64). In order to obtain [1]Hα chemical shifts in the folded state of CytR[N], we acquired HACACONNH-HA-CEST data at two B_1 fields (13 Hz and 36 Hz, see Materials and Methods) on U-^{13}C,[15]N labelled CytR[N] at 288 K. The folded conformation of CytR[N] in exchange with the intrinsically disordered form can be clearly discerned in [1]Hα CEST profiles as a minor dip in I/I_0, resulting in four extrema (two from the major dip and two from the minor dip) in CEST profiles of T23, T25, V26 and V51 (Figure 5C). [1]Hα chemical shifts for 37 non-overlapped residues (Table S1) were extracted using residue-wise fits of [1]Hα CEST profiles to the two-state Bloch-McConnell equations. Due to overlap in the [13]C-^{1}H plane, minor state [1]Hα chemical shifts from only 18 sites are independently quantifiable using the [1]H-^{13}C based experiment illustrating the efficacy of the HACACONNH-HA-CEST experiment described here. The [1]Hα chemical shift differences between folded and disordered CytR[N] are substantial and vary between -0.79 and 0.58 ppm. The large changes occurring in these shifts as a result of the helix-coil transition in CytR[N] underscores the importance of the [1]Hα nucleus in determining the structures of locally or globally folded conformations transiently sampled by IDPs.

Finally, we assembled 23 [1]Hα shifts belonging to the structured region of the CytR[N] excited state (CytR[N](11–53)) along with 33 [1]HN, 34 ^{15}N, 25 [13]Cα and 26 [13]C' shifts, as well as 31 [1]H-^{15}N RDCs from alignment in C_8E_5/octanol (PEG) and 34 from a 6 % stretched polyacrylamide gel (PAG) to calculate the structure of the excited state using CS-Rosetta (82, 83), as described earlier (52). The structure calculation converges smoothly with a well-defined funnel (Figure 6A) in the energy vs RMSD (to the lowest energy structure) plot and Figure 6B shows the ensemble of 10 lowest energy structures. The structures closely resemble each other with a pairwise all-atom RMSD of 0.97 Å (Cα RMSD = 0.43 Å) and confirm that the excited state adopts a three-helix bundle conformation. Residue-specific [1]Hα chemical shifts predicted using Sparta+ (84) for the lowest energy structure match very well with the experimental values (Figure 6C) and the overall RMSD between the predicted and observed chemical shifts is 0.22 ppm (over 22 residues for which predictions are available (Figure 6C, 6D)), well within the prediction accuracy of Sparta+ (0.25 ppm for [1]Hα) (84). The structures calculated in the absence and presence of [1]Hα chemical shifts (Figure 6E) agree well with each other (RMSD = 0.6 Å over 41 Cα atoms) and have the same helix boundaries for H1 (M12-A19), H2 (T23-M31) and H3 (Q38-V51). The only difference between the two structures is in the small orientations of the H1-H2 and H2-H3 loops. Indeed, the [1]Hα chemical shifts predicted by Sparta+ from the CS-Rosetta structure calculated without input [1]Hα chemical shifts agree well with the experimental chemical shifts (RMSD = 0.23 ppm over 22 residues, Figure 6F,6G), providing an independent validation of the proposed excited state structure of CytR[N]. The [1]Hα chemical shifts therefore reinforce the conclusion that CytR[N] undergoes a disorder-to-order switch from its native disordered form to a three-helix bundle conformation compatible with DNA binding. The [1]Hα measurements also emphasize the fact that the heterogeneity in the structural ensemble of an IDP can include globally folded conformations with stable secondary and tertiary structure.

3   Conclusions

Intrinsically disordered proteins are structurally malleable polymers in which dynamics is central for function. Here, we report the HACACONNH-HA-CEST NMR pulse sequence tailored towards characterizing millisecond dynamics in IDPs that facilitates the measurement of $^1H\alpha$ chemical shifts in IDP excited states. In this sequence, we combine spin state-selective $^1H\alpha$ CEST with triple resonance coherence transfer to the neighbouring ^1HN in order to take advantage the superior resolution in the ^1H-^{15}N 2D available for IDPs. This sequence adds to the rapidly growing toolbox of NMR methods for elucidating the structure of transiently and sparsely populated protein conformations (excited states) and pinpointing their importance for function and malfunction.

In contrast to the CPMG-based approach for determining excited state $^1H\alpha$ chemical shifts (66), a specific isotope labelling scheme is not required for measuring HACACONNH-HA-CEST data and the sequence can applied on the same U-^{13}C,^{15}N labelled samples used for acquiring $^{13}C\alpha$, $^{13}C\beta$ and ^{13}CO CEST. This makes the CEST-based strategy very attractive as it allows us to obtain excited state chemical shifts for all protein backbone nuclei of folded and disordered proteins from a single sample. While CPMG experiments generally require lesser time than CEST, CEST data acquisition can be speeded up using multifrequency irradiation schemes such as DANTE-CEST (85). In addition, the CEST approach automatically provides the sign of the chemical shift difference between the ground and excited states unlike CPMG measurements, where additional experiments must be acquired in order to extract sign information.

4   Materials and methods

4.1   Overexpression and purification of proteins

All proteins were expressed by transforming plasmids containing the gene of interest in BL21(DE3) cells. Ubiquitin was cloned into multiple cloning site of a pET-15b vector while CytR^N and L24A FF were cloned into the pET-29b+ vector. Protein expression was carried out in 500 mL 2xM9 minimal media (86) containing the appropriate selection antibiotic. The media was supplemented with 1g/L ^{15}N labelled ammonium chloride and 3g/L U-^{13}C-labelled glucose as the sole nitrogen and carbon sources respectively. Protein expression was induced by adding 1 mM IPTG to the media when the absorbance at 600 nm (OD) reached 0.8-1 for all proteins. Induction was done at 18-20 °C for all proteins typically for 18 hours. Cells were harvested by centrifugation at 7500 rpm for 20 minutes and lysed by sonication in lysis buffer. The purification protocol followed for each protein is described briefly below.

4.1.1   Ubiquitin—Cells were lysed in 50 mM Tris HCl buffer pH 7.6 containing 1 mM EDTA and 0.05% Triton-X100 by sonication followed by incubation at room temperature with 1 mM phenyl methyl sulfonyl fluoride (PMSF) and 7 mg lysozyme per gram of the cell pellet. The lysate was clarified by centrifugation at 15000 rpm for 1 hr and glacial acetic acid was added dropwise to the clear supernatant while stirring until the solution turned milky white. The lysate was stirred for 30 minutes in room temperature and precipitated bacterial proteins were pelleted down by centrifugation for 30 minutes at 15000 rpm. The supernatant was dialyzed twice against 50 mM sodium acetate pH 4.5 at 4 °C. Dialyzed

fractions were applied onto a SP-Sepharose cation exchange column with the dialysis buffer for column equilibration and wash, and the same buffer with 0.5 M NaCl (Buffer B) for linear gradient elution. Ubiquitin eluted at ~50 % Buffer B concentration. Elution fractions containing ubiquitin were purified further using gel filtration chromatography on a 120 mL Superdex75 column from which ubiquitin eluted between 82-90 mL. NMR data was acquired in Tris -HCl pH 7.6, 1 mM EDTA, 25 mM NaCl and 10 % D_2O.

4.1.2   CytR^N—Cells were lysed as described for ubiquitin above and the supernatant after cell lysis was made up to 60 % of the saturation value of ammonium sulphate to precipitate bacterial proteins. The precipitated proteins were removed by centrifugation and the supernatant was dialyzed into 50 mM sodium phosphate buffer containing 30 mM NaCl and 1 mM EDTA (lysis and equilibration buffer) before being applied to a SP-Sepharose cation exchange column. The equilibration buffer containing 1 M NaCl was used for elution (Buffer B). CytR^N eluted at ~60 % Buffer B concentration. Fractions containing CytR^N were pooled and further purified by gel filtration chromatography on a 120 mL Superdex 75 size exclusion column equilibrated in 150 mM ammonium acetate, from which CytR^N eluted at ~70 mL. Pure fractions were checked on SDS-PAGE, pooled and lyophilized. NMR data were acquired in buffer containing 20 mM sodium phosphate buffer (pH 7.0) and 10 % D_2O.

4.1.3   L24A FF domain—Cells expressing L24A FF were lysed as described for ubiquitin above. Purification was done using cation exchange chromatography on an SP-Sepharose column as mentioned previously for ubiquitin. The lysis buffer was 50 mM sodium phosphate buffer (pH 5.7) containing 1 mM EDTA, and the same buffer for used for column equilibration. Elution was done with the equilibration buffer containing 500 mM NaCl (Buffer B). L24A FF eluted at ~60 % Buffer B concentration; fractions containing protein were further purified on a 120 mL Superdex 75 column, from which L24A FF eluted at a volume of ~83 mL. NMR samples were prepared in 50 mM sodium acetate buffer pH 5.7, 50 mM NaCl, 1 mM EDTA and 10 % D_2O.

4.2   NMR spectroscopy: data acquisition and processing

All NMR experiments were carried out on an Agilent DD2 14.1 T (600 MHz ^1H Larmor frequency) spectrometer equipped with a room-temperature (L24A FF) or cryogenically cooled (ubiquitin and CytR^N) triple resonance single-axis gradient probe. Measurements on ubiquitin, L24A FF and CytR^N were done at 298 K, 303 K and 288 K respectively. All NMR datasets were processed with nmrPipe (87) and visualized with nmrDraw (87) and NMRFAM-Sparky (88).

4.3 $^1H\alpha$ CEST

All $^1H\alpha$ CEST data were acquired in an interleaved pseudo-3D manner with the pulse sequence shown in Figure 2. Each dataset contained a reference plane measured as described in Section 2.1 for obtaining normalizing CEST profiles. Peak intensities (I) in each plane of the CEST dataset were extracted by global lineshape analysis using FuDA and plotted as a ratio with the intensity in the reference plane (I_0) against the chemical shift offset for that particular plane to generate the CEST profile. Uncertainties in the I/I_0 values for each CEST profile were calculated using the scatter in the CEST baselines.

In order to extract ϖ_{GE} values, ^{1}Hα CEST profiles were fit globally or residue-wise to the two-state Bloch-McConnell equations as described earlier using the software package ChemEx (https://github.com/gbouvignies/ChemEx), which numerically integrates the Bloch-McConnell equations (80). The radiofrequency field used in each ^{1}Hα CEST experiment was calibrated using the method of Guennuegues *et al* (89). ^{1}Hα R_{1H} values were measured using a modified version of the HACACONNH-HA-CEST pulse sequence as described in the Section 2.1. Nine delays (including two additional repeat points for error analysis) ranging from 1 to 350 ms were used and intensities were fit to a single exponential decay equation as a function of the delay to extract R_{1H}. ^{1}Hα R_{1H} values were then input into the fitting routine for determining ϖ_{GE} and were not allowed to vary during the fit. It is not essential to measure R_{1H} in order to extract ϖ_{GE}; fixing R_{1H} to a reasonable value (2 s^{-1}) results in comparable fit quality and very similar ϖ_{GE} for CytR^N (maximum deviation of 18 ppb and average deviation of 2.4 ppb over 37 residues).

4.3.1 Ubiquitin—^{1}Hα CEST profiles were acquired at a B_1 field of 28 Hz over a chemical shift range of 3.1 to 6.1 ppm (50 planes) for an exchange duration (T_{ex}) of 250 ms.

4.3.2 L24A FF—^{1}Hα CEST profiles were acquired at B1 fields of 15 Hz (15 Hz spacing, 98 planes), 33 Hz (30 Hz spacing, 50 planes) and 54 Hz (50 Hz spacing, 34 planes) over a chemical shift range of 2.85 to 5.25 ppm (15 Hz and 33 Hz) or 2.71 to 5.38 ppm (54 Hz) for an exchange duration (T_{ex}) of 400 ms. Three B_1 field ^{1}Hα CEST profiles for 22 residues (T13, A17, K18, F21, L25, R29, S35, W36, E37, A39, M40, K41, I43, I44, Y49, A51, L55, S56, K58, K59, N63 and K66) with $|\varpi_{IN}| > 0.35$ ppm were globally fit to the two-state Bloch-McConnell equations to extract $k_{ex,IN}$ (582 ± 6 s^{-1}) and p_I (4.24 ± 0.04 %). The ground state in-phase ^{1}Hα transverse relaxation rate constant for the operator $H_\pm$ ($R_{2,N}$) was allowed to vary, while $R_2 = R_{2,I} - R_{2,N}$ was fixed to 0 to enforce the condition that the excited and ground states have the same transverse relaxation rate constants. Here, $R_{2,I}$ is the in-phase transverse relaxation rate constant for the excited state I. Antiphase transverse relaxation rate constants (*2Ha$_\pm$Ca$_Z$*) of ground and excited state ^{1}Hα were calculated as $R_f(2H_\pm C_Z) = R_{2,i} - R_{1,i}$, while the relaxation rate constant of longitudinal two-spin order $2H_Z C_Z$ was calculated as $R_f(2H_Z C_Z) = R_{1C,i} - R_{1H,i}$. The longitudinal and transverse cross-correlated relaxation rate constants η_z and η_{xy} are also varied during the fit within the constraints $\eta_{z,G} = \eta_{z,E}$ and $\eta_{xy,G} = \eta_{xy,G}$. η_z accounts for any small deviation of the ^{1}Hα CEST profile baseline from 0. $^1J_{C\alpha H\alpha}$ was fixed for both the ground and excited states to 143.5 Hz. For residues with small ϖ_{IN}, the ^{1}Hα CEST profile acquired with a B_1 field of 33 Hz was fit residue-wise by fixing the p_I and $k_{ex,IN}$ to the values obtained from the global fit in order to extract ϖ_{IN}.

4.3.3 CytR^N—^{1}Hα CEST profiles were acquired at B_1 fields of 13 Hz (12 Hz spacing, 127 planes) and 36 Hz (30 Hz spacing, 52 planes) over a chemical shift range of 2.83 to 5.33 ppm for an exchange duration (T_{ex}) of 500 ms. Global fitting to the two-state Bloch-McConnell equations was done as described above for L24A FF using 2 B_1 field ^{1}Hα CEST profiles of A10, V21, T23, A24, T25, V26, T40, E50, V51, Y53 and P55, all of which have $|\varpi_{DE}| > 0.2$ ppm. In this case, R_2 was allowed to float as it resulted in a significant improvement in the quality of the fit compared to the case of $R_2 = 0$. Exchange parameters

of $k_{ex,DE} = 38 \pm 2$ s^{-1} and $p_E = 4.5 \pm 0.1\%$ were obtained from these global fits. For residues with small $|\varpi_{DE}|$, p_E and $k_{ex,DE}$ were fixed to the values determined from the global fits and the 13 and 36 Hz ^{1}Hα CEST profiles were fit together at the residue level to extract ϖ_{DE}.

4.4 CS-Rosetta structure calculations

The structure of the CytR^N excited state was calculated using 119 chemical shifts (including 23 ^{1}Hα shifts) and 65 ^{1}H-^{15}N residual dipolar couplings (determined in (52)) as described in detail in (52).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

5 Funding statement

References

1. Uversky VN. Intrinsically disordered proteins and their "mysterious"(meta) physics. Front Phys. 2019; 7: 10.

2. Wright PE, Dyson HJ. Linking folding and binding. Curr Opin Struct Biol. 2009; 19: 31. [PubMed: 19157855]

3. Gianni S, Dogan J, Jemth P. Coupled binding and folding of intrinsically disordered proteins: what can we learn from kinetics? Curr Opin Struct Biol. 2016; 36: 18. [PubMed: 26720267]

4. Uversky VN. Unusual biophysics of intrinsically disordered proteins. Biochim Biophys Acta: Proteins Proteomics. 2013; 1834: 932.

5. Babu MM. The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. Biochem Soc Trans. 2016; 44: 1185. [PubMed: 27911701]

6. Bah A, et al. Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. Nature. 2015; 519: 106. [PubMed: 25533957]

7. Borgia A, et al. Extreme disorder in an ultrahigh-affinity protein complex. Nature. 2018; 555: 61. [PubMed: 29466338]

8. Forman-Kay JD, Mittag T. From sequence and forces to structure, function, and evolution of intrinsically disordered proteins. Structure. 2013; 21: 1492. [PubMed: 24010708]

9. Piovesan D, et al. DisProt 7.0: a major update of the database of disordered proteins. Nucleic Acids Res. 2017; 45 D219 [PubMed: 27899601]

10. Uversky VN. Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes. Chem Soc Rev. 2011; 40: 1623. [PubMed: 21049125]

11. Fung HYJ, Birol M, Rhoades E. IDPs in macromolecular complexes: the roles of multivalent interactions in diverse assemblies. Curr Opin Struct Biol. 2018; 49: 36. [PubMed: 29306779]

12. Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol. 2015; 16: 18. [PubMed: 25531225]

13. Bondos SE, Dunker AK, Uversky VN. Intrinsically disordered proteins play diverse roles in cell signaling. Cell Communication and Signaling. 2022; 20: 20. [PubMed: 35177069]

14. Smith NC, Kuravsky M, Shammas SL, Matthews JM. Binding and folding in transcriptional complexes. Curr Opin Struct Biol. 2021; 66: 156. [PubMed: 33248428]

15. Brodsky S, Jana T, Barkai N. Order through disorder: The role of intrinsically disordered regions in transcription factor binding specificity. Curr Opin Struct Biol. 2021; 71: 110. [PubMed: 34303077]

16. Palmer AG III. Chemical exchange in biomacromolecules: past, present, and future. J Magn Reson. 2014; 241: 3. [PubMed: 24656076]

17. Anthis NJ, Clore GM. Visualizing transient dark states by NMR spectroscopy. Q Rev Biophys. 2015; 48: 35. [PubMed: 25710841]

18. Sekhar A, Kay LE. An NMR view of protein dynamics in health and disease. Annu Rev Biophys. 2019; 48: 297. [PubMed: 30901260]

19. Vallurupalli P, Bouvignies G, Kay LE. Studying "invisible" excited protein states in slow exchange with a major state conformation. J Am Chem Soc. 2012; 134: 8148. [PubMed: 22554188]

20. Forsén S, Hoffman RA. Study of moderately rapid chemical exchange reactions by means of nuclear magnetic double resonance. J Chem Phys. 1963; 39: 2892.

21. Ward K, Aletras A, Balaban RS. A new class of contrast agents for MRI based on proton chemical exchange dependent saturation transfer (CEST). J Magn Reson. 2000; 143: 79. [PubMed: 10698648]

22. Fawzi NL, Ying J, Ghirlando R, Torchia DA, Clore GM. Atomic-resolution dynamics on the surface of amyloid-[bgr] protofibrils probed by solution NMR. Nature. 2011; 480: 268. [PubMed: 22037310]

23. Mittermaier A, Kay LE. New tools provide new insights in NMR studies of protein dynamics. Science. 2006; 312: 224. [PubMed: 16614210]

24. Loria JP, Rance M, Palmer AG. A relaxation-compensated Carr-Purcell-Meiboom-Gill sequence for characterizing chemical exchange by NMR spectroscopy. J Am Chem Soc. 1999; 121: 2331.

25. Reddy JG, et al. Simultaneous determination of fast and slow dynamics in molecules using extreme CPMG relaxation dispersion experiments. J Biomol NMR. 2018; 70: 1. [PubMed: 29188417]

26. Rangadurai A, Szymaski ES, Kimsey IJ, Shi H, Al-Hashimi HM. Characterizing micro-to-millisecond chemical exchange in nucleic acids using off-resonance R1ρ relaxation dispersion. Prog Nucl Magn Reson Spectrosc. 2019; 112: 55. [PubMed: 31481159]

27. Ban D, Sabo TM, Griesinger C, Lee D. Measuring dynamic and kinetic information in the previously inaccessible supra-τc window of nanoseconds to microseconds by solution NMR spectroscopy. Molecules. 2013; 18 11904 [PubMed: 24077173]

28. Iwahara J, Clore GM. Detecting transient intermediates in macromolecular binding by paramagnetic NMR. Nature. 2006; 440: 1227. [PubMed: 16642002]

29. Clore GM. Exploring sparsely populated states of macromolecules by diamagnetic and paramagnetic NMR relaxation. Protein Sci. 2011; 20: 229. [PubMed: 21280116]

30. Korzhnev DM, Religa TL, Banachewicz W, Fersht AR, Kay LE. A transient and low-populated protein-folding intermediate at atomic resolution. Science. 2010; 329: 1312. [PubMed: 20829478]

31. Korzhnev DM, et al. Nonnative interactions in the FF domain folding pathway from an atomic resolution structure of a sparsely populated intermediate: An NMR relaxation dispersion study. J Am Chem Soc. 2011; 133 10974 [PubMed: 21639149]

32. Tiwari VP, Toyama Y, De D, Kay LE, Vallurupalli P. The A39G FF domain folds on a volcano-shaped free energy surface via separate pathways. Proceedings of the National Academy of Sciences. 2021; 118 e2115113118

33. Sekhar A, Rosenzweig R, Bouvignies G, Kay LE. Mapping the conformation of a client protein through the Hsp70 functional cycle. Proc Natl Acad Sci U S A. 2015; 112 10395 [PubMed: 26240333]

34. Sekhar A, Rosenzweig R, Bouvignies G, Kay LE. Hsp70 biases the folding pathways of client proteins. Proc Natl Acad Sci U S A. 2016. E2794 [PubMed: 27140645]

35. Sekhar A, et al. Conserved conformational selection mechanism of Hsp70 chaperone-substrate interactions. eLife. 2018; 7 e32764 [PubMed: 29460778]

36. Yao X, Rosen MK, Gardner KH. Estimation of the available free energy in a LOV2-Jα photoswitch. Nat Chem Biol. 2008; 4: 491. [PubMed: 18604202]

37. Whittier SK, Hengge AC, Loria JP. Conformational motions regulate phosphoryl transfer in related protein tyrosine phosphatases. Science. 2013; 341: 899. [PubMed: 23970698]

38. Boehr DD, McElheny D, Dyson HJ, Wright PE. The dynamic energy landscape of dihydrofolate reductase catalysis. Science. 2006; 313: 1638. [PubMed: 16973882]

39. Boehr DD, McElheny D, Dyson HJ, Wright PE. Millisecond timescale fluctuations in dihydrofolate reductase are exquisitely sensitive to the bound ligands. Proceedings of the National Academy of Sciences. 2010; 107: 1373.

40. Culik RM, et al. Effects of maturation on the conformational free-energy landscape of SOD1. Proc Natl Acad Sci U S A. 2018; 115 E2546 [PubMed: 29483249]

41. Kimsey IJ, Petzold K, Sathyamoorthy B, Stein ZW, Al-Hashimi HM. Visualizing transient Watson–Crick-like mispairs in DNA and RNA duplexes. Nature. 2015; 519: 315. [PubMed: 25762137]

42. Kimsey IJ, et al. Dynamic basis for dG. dT misincorporation via tautomerization and ionization. Nature. 2018; 554: 195. [PubMed: 29420478]

43. Zhao B, Guffy SL, Williams B, Zhang Q. An excited state underlies gene regulation of a transcriptional riboswitch. Nat Chem Biol. 2017; 13: 968. [PubMed: 28719589]

44. Korzhnev DM, Religa TL, Kay LE. Transiently populated intermediate functions as a branching point of the FF domain folding pathway. Proceedings of the National Academy of Sciences. 2012; 109 17777

45. Sekhar A, et al. Thermal fluctuations of immature SOD1 lead to separate folding and misfolding pathways. eLife. 2015; 4 e07296 [PubMed: 26099300]

46. Neudecker P, et al. Structure of an intermediate state in protein folding and aggregation. Science. 2012; 336: 362. [PubMed: 22517863]

47. Schneider R, et al. Visualizing the molecular recognition trajectory of an intrinsically disordered protein using multinuclear relaxation dispersion NMR. J Am Chem Soc. 2015; 137: 1220. [PubMed: 25551399]

48. Charlier C, et al. Structure and dynamics of an intrinsically disordered protein region that partially folds upon binding by chemical-exchange NMR. J Am Chem Soc. 2017; 139 12219 [PubMed: 28780862]

49. Bolik-Coulon, N, Bouvignies, G, Carlier, L, Ferrage, F. Intrinsically Disordered Proteins. Elsevier; 2019. 65–92.

50. Schneider R, Blackledge M, Jensen MR. Elucidating binding mechanisms and dynamics of intrinsically disordered protein complexes using NMR spectroscopy. Curr Opin Struct Biol. 2019; 54: 10. [PubMed: 30316104]

51. Fisher CK, Stultz CM. Constructing ensembles for intrinsically disordered proteins. Curr Opin Struct Biol. 2011; 21: 426. [PubMed: 21530234]

52. Madhurima K, Nandi B, Munshi S, Naganathan AN, Sekhar A. Functional regulation of an intrinsically disordered protein via a conformationally excited state. Science Advances. 2023.

53. Weickert MJ, Adhya S. A family of bacterial regulators homologous to Gal and Lac repressors. J Biol Chem. 1992; 267 15869 [PubMed: 1639817]

54. Barbier CS, Short SA, Senear DF. Allosteric Mechanism of Induction of CytR-regulated Gene Expression: CytR REPRESSOR-CYTIDINE INTERACTION. J Biol Chem. 1997; 272 16962 [PubMed: 9202008]

55. Lewis M. The lac repressor. C R Biol. 2005; 328: 521. [PubMed: 15950160]

56. Tretyachenko-Ladokhina V, Cocco MJ, Senear DF. Flexibility and adaptability in binding of E. coli cytidine repressor to different operators suggests a role in differential gene regulation. J Mol Biol. 2006; 362: 271. [PubMed: 16919681]

57. Moody CL, Tretyachenko-Ladokhina V, Laue TM, Senear DF, Cocco MJ. Multiple conformations of the cytidine repressor DNA-Binding domain coalesce to one upon recognition of a specific DNA surface. Biochemistry. 2011; 50: 6622. [PubMed: 21688840]

58. Bell CE, Lewis M. A closer view of the conformation of the Lac repressor bound to operator. Nat Struct Biol. 2000; 7: 209. [PubMed: 10700279]

59. Kalodimos CG, Boelens R, Kaptein R. Toward an Integrated Model of Protein-DNA Recognition as Inferred from NMR Studies on the L ac Repressor System. Chem Rev. 2004; 104 3567 [PubMed: 15303828]

60. Slijper M, Bonvin A, Boelens R, Kaptein R. Refined Structure oflacRepressor Headpiece (1-56) Determined by Relaxation Matrix Calculations from 2D and 3D NOE Data: Change of Tertiary Structure upon Binding to thelacOperator. J Mol Biol. 1996; 259: 761. [PubMed: 8683581]

61. Nagadoi A, et al. Structural comparison of the free and DNA-bound forms of the purine repressor DNA-binding domain. Structure. 1995; 3: 1217. [PubMed: 8591032]

62. Penin F, et al. Three-dimensional structure of the DNA-binding domain of the fructose repressor from Escherichia coli by 1H and 15N NMR. J Mol Biol. 1997; 270: 496. [PubMed: 9237914]

63. Long D, Sekhar A, Kay LE. Triple resonance-based 13Cα and 13Cβ CEST experiments for studies of ms timescale dynamics in proteins. J Biomol NMR. 2014; 1

64. Vallurupalli P, Kay LE. Probing slow chemical exchange at carbonyl sites in proteins by chemical exchange saturation transfer NMR spectroscopy. Angew Chem Int Ed. 2013; 52: 4156.

65. Yuwen T, Kay LE. A new class of CEST experiment based on selecting different magnetization components at the start and end of the CEST relaxation element: an application to 1H CEST. J Biomol NMR. 2018; 70: 93. [PubMed: 29352366]

66. Lundström P, Hansen DF, Vallurupalli P, Kay LE. Accurate measurement of alpha proton chemical shifts of excited protein states by relaxation dispersion NMR spectroscopy. J Am Chem Soc. 2009; 131: 1915. [PubMed: 19152327]

67. Wang Y, Jardetzky O. Probability-based protein secondary structure identification using combined NMR chemical-shift data. Protein Sci. 2002; 11: 852. [PubMed: 11910028]

68. Wishart DS. Interpreting protein chemical shift data. Prog Nucl Magn Reson Spectrosc. 2010; 58: 62. [PubMed: 21241884]

69. Vallurupalli P, Sekhar A, Yuwen T, Kay LE. Probing conformational dynamics in biomolecules via chemical exchange saturation transfer: a primer. J Biomol NMR. 2017; 67: 243. [PubMed: 28317074]

70. Bouvignies G, Kay LE. Measurement of proton chemical shifts in invisible states of slowly exchanging protein systems by chemical exchange saturation transfer. J Phys Chem B. 2012; 116 14311 [PubMed: 23194058]

71. Yuwen T, Sekhar A, Kay LE. Separating Dipolar and Chemical Exchange Magnetization Transfer Processes in 1H-CEST. Angew Chem Int Ed. 2017; 56: 6122.

72. Grzesiek S, Bax A. Amino acid type determination in the sequential assignment procedure of uniformly 13 C/15 N-enriched proteins. J Biomol NMR. 1993; 3: 185. [PubMed: 8477186]

73. Löhr F, Rüterjans H. A new triple-resonance experiment for the sequential assignment of backbone resonances in proteins. J Biomol NMR. 1995; 6: 189. [PubMed: 22910800]

74. Bazzo R, Cicero DO, Barbato G. A new three-dimensional pulse sequence for correlating intraresidue NH, N, and CO chemical shifts in 13C, l5N-labeled proteins. J Magn Reson, Ser B. 1996; 110: 65.

75. Bolik-Coulon N, Hansen DF, Kay LE. Optimizing frequency sampling in CEST experiments. J Biomol NMR. 2022; 76: 167. [PubMed: 36192571]

76. Yuwen T, Kay LE. Longitudinal relaxation optimized amide 1 H-CEST experiments for studying slow chemical exchange processes in fully protonated proteins. J Biomol NMR. 2017; 67: 295. [PubMed: 28357518]

77. Bouvignies G, Vallurupalli P, Kay LE. Visualizing side chains of invisible protein conformers by solution NMR. J Mol Biol. 2014; 426: 763. [PubMed: 24211467]

78. Tiwari VP, Pandit S, Vallurupalli P. Exchangeable deuterons introduce artifacts in amide 15 N CEST experiments used to study protein conformational exchange. J Biomol NMR. 2019; 73: 43. [PubMed: 30661150]

79. Sekhar A, Vallurupalli P, Kay LE. Folding of the four-helix bundle FF domain from a compact on-pathway intermediate state is governed predominantly by water motion. Proc Natl Acad Sci U S A. 2012; 109 19268 [PubMed: 23129654]

80. McConnell HM. Reaction rates by nuclear magnetic resonance. J Chem Phys. 1958; 28: 430.

81. Munshi S, Rajendran D, Naganathan AN. Entropic control of an excited folded-like conformation in a disordered protein ensemble. J Mol Biol. 2018; 430: 2688. [PubMed: 29885328]

82. Shen Y, et al. Consistent blind protein structure generation from NMR chemical shift data. Proc Natl Acad Sci U S A. 2008; 105: 4685. [PubMed: 18326625]

83. Nerli, S, Sgourakis, NG. Methods Enzymol. Vol. 614. Elsevier; 2019. 321–362.

84. Shen Y, Bax A. SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. J Biomol NMR. 2010; 48: 13. [PubMed: 20628786]

85. Yuwen T, Kay LE, Bouvignies G. Dramatic Decrease in CEST Measurement Times Using Multi-Site Excitation. ChemPhysChem. 2018; 19 1707 [PubMed: 29663694]

86. Azatian SB, Kaur N, Latham MP. Increasing the buffering capacity of minimal media leads to higher protein yield. J Biomol NMR. 2019; 73: 11. [PubMed: 30613903]

87. Delaglio F, et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR. 1995; 6: 277. [PubMed: 8520220]

88. Lee W, Tonelli M, Markley JL. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. Bioinformatics. 2015; 31: 1325. [PubMed: 25505092]

89. Guenneugues M, Berthault P, Desvaux H. A Method for DeterminingB1Field Inhomogeneity. Are the Biases Assumed in Heteronuclear Relaxation Experiments Usually Underestimated? J Magn Reson. 1999; 136: 118. [PubMed: 9887297]

90. Kay LE, Ikura M, Tschudin R, Bax A. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. J Magn Reson. 1990; 89: 496.

91. Shaka A, Keeler J, Frenkiel T, Freeman R. An improved sequence for broadband decoupling: WALTZ-16. J Magn Reson. 1983; 52: 335.

92. McCoy M, Mueller L. Selective decoupling. Journal of Magnetic Resonance, Series A. 1993; 101: 122.

93. McCoy M, Mueller L. Selective shaped pulse decoupling in NMR: homonuclear [carbon-13] carbonyl decoupling. J Am Chem Soc. 1992; 114: 2108.

94. Kupce , Freeman R. Optimized adiabatic pulses for wideband spin inversion. Journal of Magnetic Resonance, Series A. 1996; 118: 299.

95. Logan TM, Olejniczak ET, Xu RX, Fesik SW. A general method for assigning NMR spectra of denatured proteins using 3D HC (CO) NH-TOCSY triple resonance experiments. J Biomol NMR. 1993; 3: 225. [PubMed: 8477187]

96. Kay L, Keifer P, Saarinen T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. J Am Chem Soc. 1992; 114 10663

97. Schleucher J, Sattler M, Griesinger C. Coherence selection by gradients without signal attenuation: application to the three-dimensional HNCO experiment. Angewandte Chemie International Edition in English. 1993; 32: 1489.

98. Marion D, Ikura M, Tschudin R, Bax A. Rapid recording of 2D NMR spectra without phase cycling. Application to the study of hydrogen exchange in proteins. J Magn Reson. 1989; 85: 393.

99. Vijay-Kumar S, Bugg CE, Cook WJ. Structure of ubiquitin refined at 1.8 Åresolution. J Mol Biol. 1987; 194: 531. [PubMed: 3041007]
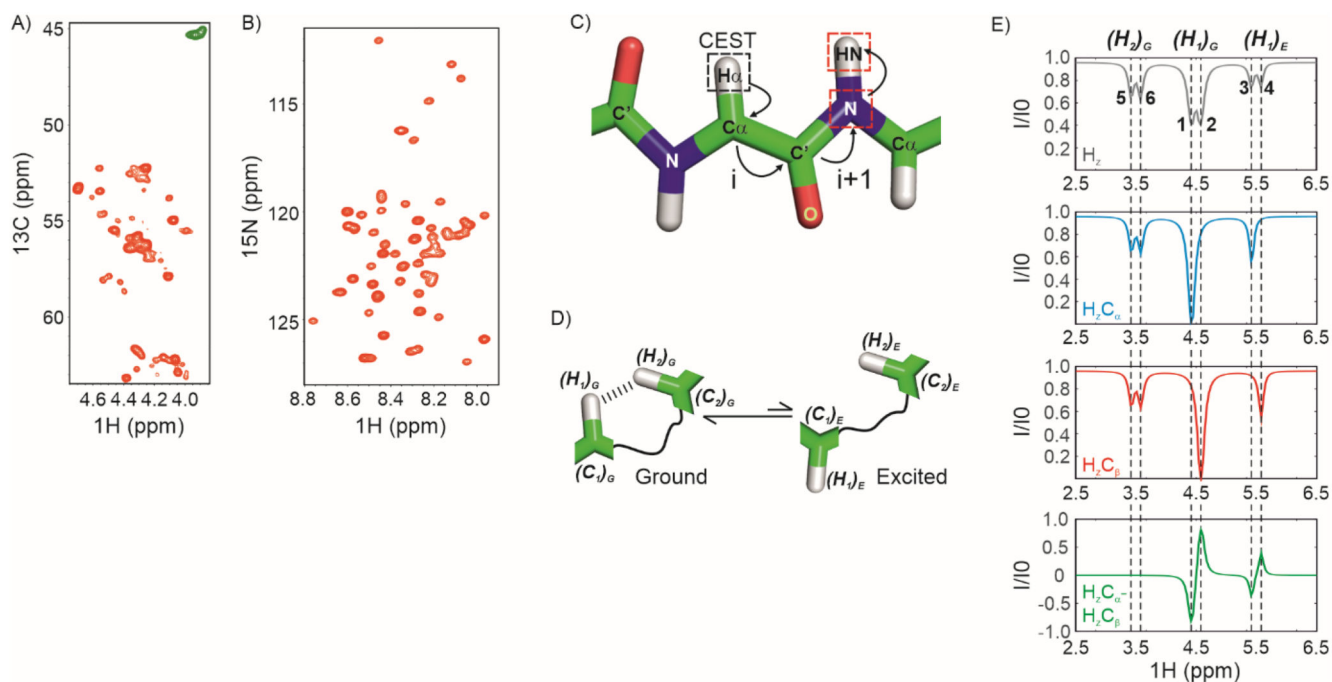
Figure 1. Strategy for acquiring $^1H\alpha$ CEST data on IDPs.

A) The $^1H\alpha$ region of a constant-time ^1H-^{13}C HSQC spectrum of U-^{13}C,^{15}N labelled CytR^N. Green resonances arise from Gly which do not have a scalar coupled sidechain $^{13}C\beta$. B) ^1H-^{15}N HSQC spectrum of CytR^N. C) The coherence transfer pathway in the $^1H\alpha$ CEST experiment reported here. CEST is carried out on $^1H\alpha$ (black dashed square) and the magnetization is then transferred from $^1H\alpha$(i) through $^{13}C\alpha$(i) and ^{13}CO(i) to ^{15}N (i+1) and then to ^1HN(i+1) for detection. Red dashed squares indicate nuclei that are frequency labelled in the experiment. D) Cartoon representation of conformational exchange between ground and excited states of a protein where the alpha proton H_1 comes close (< 5 Å) to H_2 in the ground state. E) The spin-state-selective approach for acquiring proton CEST without interference from dipolar cross-relaxation. (Top) ^1H CEST profile of an alpha proton resonating at 4.5 ppm (major dip), exchanging with a second conformation where the chemical shift of the $^1H\alpha$ is 5.5 ppm (minor dip). The $^1H\alpha$ is proximal to a second $^1H\alpha$ which results in an NOE dip at 3.5 ppm. There is no ^{13}C-decoupling during the CEST period; therefore doublets separated by $^1J_{H\alpha C\alpha}$ are seen for each dip. (Middle) Spin-state-selective CEST data selecting for H_zC_α (blue) and H_zC_β (red). (Bottom) CEST profile of H_zC_α-H_zC_β =$2H_zC_z$, where the NOE dip cancels out (green).
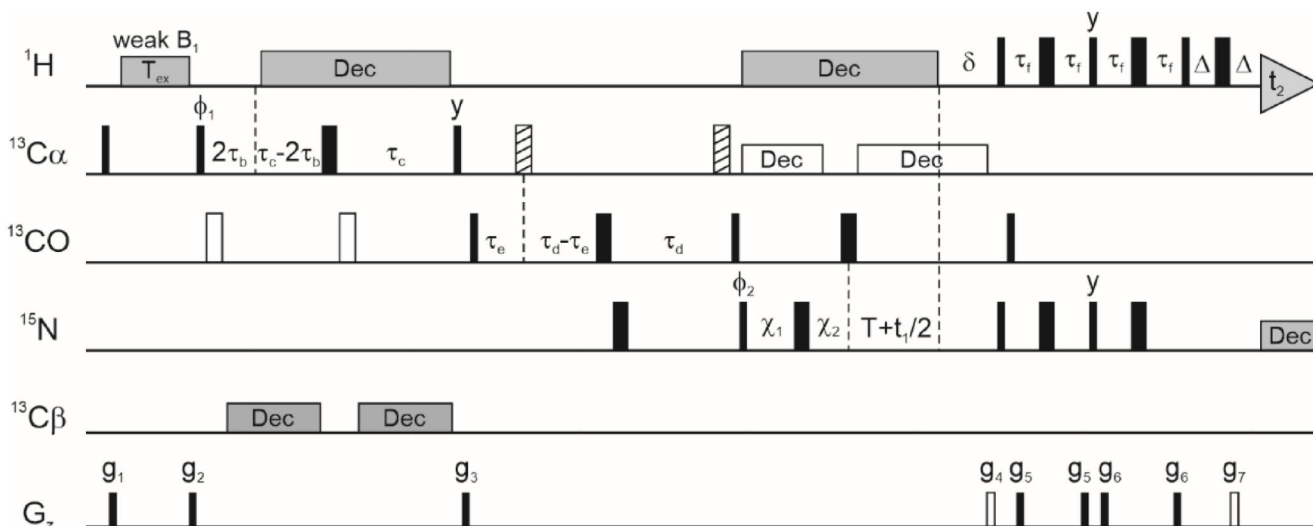
Figure 2. The pulse sequence for acquiring HACACONNH-based ^{1}Hα CEST data.
The ^{1}H carrier is centred on the water resonance except during the CEST period and the ^{15}N carrier at ~ 118 ppm, while the ^{13}C carrier is placed at 58 ppm till gradient g3 and then jumped to 176 ppm for the rest of the pulse sequence. All pulses are applied along the x-axis unless specified otherwise. Narrow and wide rectangles represent 90° and 180° pulses respectively. Open rectangles correspond to rectangular pulses applied 118 ppm off-resonance while the hashed rectangles are rectangular pulses applied -118 ppm off-resonance. ^{1}H and ^{15}N pulses are applied at the highest power while ^{13}C 90° and 180° pulses are applied at RF field amplitudes of $\xi/\sqrt{15}$ and $\xi/\sqrt{3}$ Hz respectively, where ξ is the frequency difference in Hz between the ^{13}Cα and ^{13}CO spectral bands (90). ^{1}H decoupling is implemented via a ~6 kHz WALTZ-16 (91) composite pulse decoupling pattern. Off-resonance ^{13}Cα decoupling is carried out with a cosine modulated SEDUCE-1 pattern (92, 93) with a pulse width of 315 μs (at 600 MHz B_0 field strength), and ^{15}N decoupling during acquisition is carried out with a ~ 1 kHz WALTZ-16 pattern. WURST-2 adiabatic decoupling (94) is used for ^{13}Cβ decoupling in the 16.5-41.5 ppm and 68-72 ppm bands. The phase cycling used in the sequence is: ϕ_1 = (-y,y), ϕ_2 = (x,x,-x,-x) and receiver = (x,-x,-x,x). The delays used in the experiment are: τ_b = 1.8 ms, τ_c = 4.4 ms, τ_d = 12.4 ms, τ_e = 5 ms, τ_f = 2.3 ms, T = 12.4 ms, δ = 5.5 ms and　 = 0.5 ms. The amplitudes and durations of the various gradients in (G/cm, ms) are: g_1: (12000, 0.5), g_2: (-8343, 0.5), g_3: (10000, 0.8), g_4: (30000, 1.25), g_5: (4000, 0.3), g_6: (2000, 0.4), g_7: (30400, 0.125). The two gradients g_4 and g_7 represented as open rectangles are required for coherence selection. Delays χ_1 and χ_2 are set to the larger of the two values (T-t$_1$/2, 0) and (0, t$_1$/2-T) respectively, and this allows ^{15}N chemical shift evolution in t$_1$ to occur in a semi-constant manner (72, 95). Quadrature detection in the F1 dimension is implemented using enhanced sensitivity gradient coherence selection (96–98).
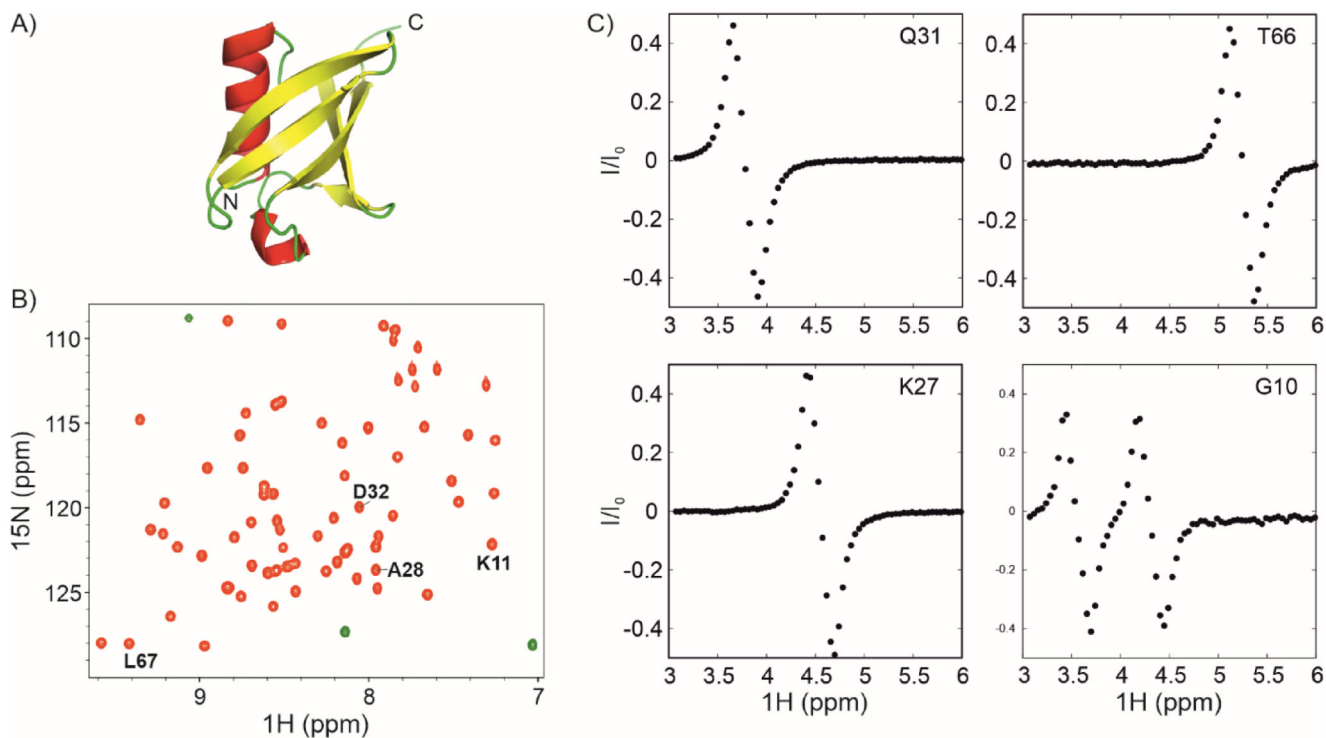
Figure 3. ^{1}Hα CEST on U-^{13}C,^{15}N labelled ubiquitin.
A) Three-dimensional structure of ubiquitin showing helices, sheets and loops in red, yellow and green respectively (PDB ID: 1UBQ (99)). B) ^{1}H-^{15}N HSQC spectrum of ubiquitin. C) ^{1}Hα CEST profiles of selected residues in ubiquitin that were obtained by analyzing the i+1 correlations indicated in the HSQC spectrum in panel B. ^{1}Hα CEST data were acquired using the pulse sequence in Figure 2. The ^{1}Hα CEST profile for Gly10 shows two dips arising from each of the two diastereotopic ^{1}Hα protons of Gly10.
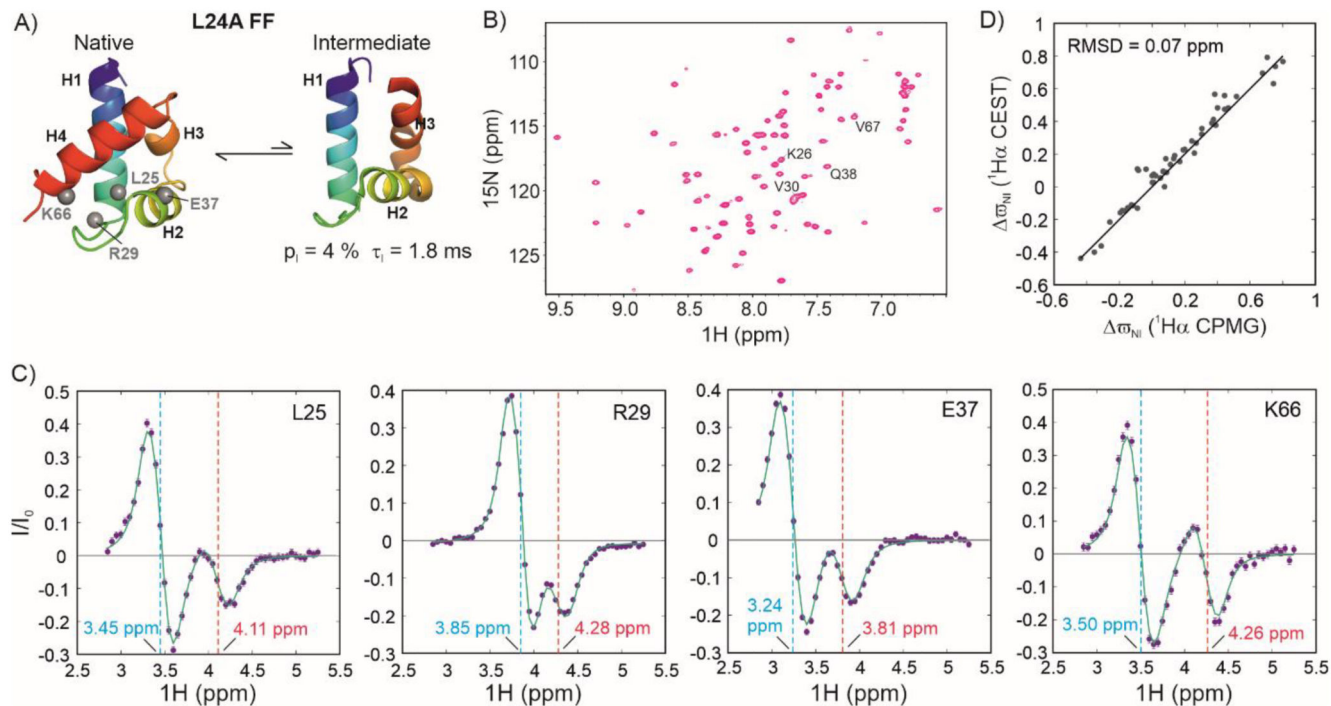
Figure 4. Characterizing exchange in the folded L24A FF domain using ^{1}Hα CEST.
A) Cartoon representation of the native (31) and intermediate states of L24A FF (PDB ID:2L9V (31)) in exchange on the millisecond timescale. B) ^{1}H-^{15}N HSQC spectrum of L24A FF. Resonances whose ^{1}Hα CEST profiles are quantified in panel C are labelled in the spectrum. C) ^{1}Hα CEST profiles of selected resonances of L24A FF. Cyan and red dashed lines indicate the chemical shifts of the native and intermediate states respectively. D) Correlation between the ^{1}Hα chemical shifts of L24A FF intermediate states acquired using ^{1}Hα CPMG (y-axis, (31)) and ^{1}Hα CEST (x-axis). The solid line depicts the y=x function.
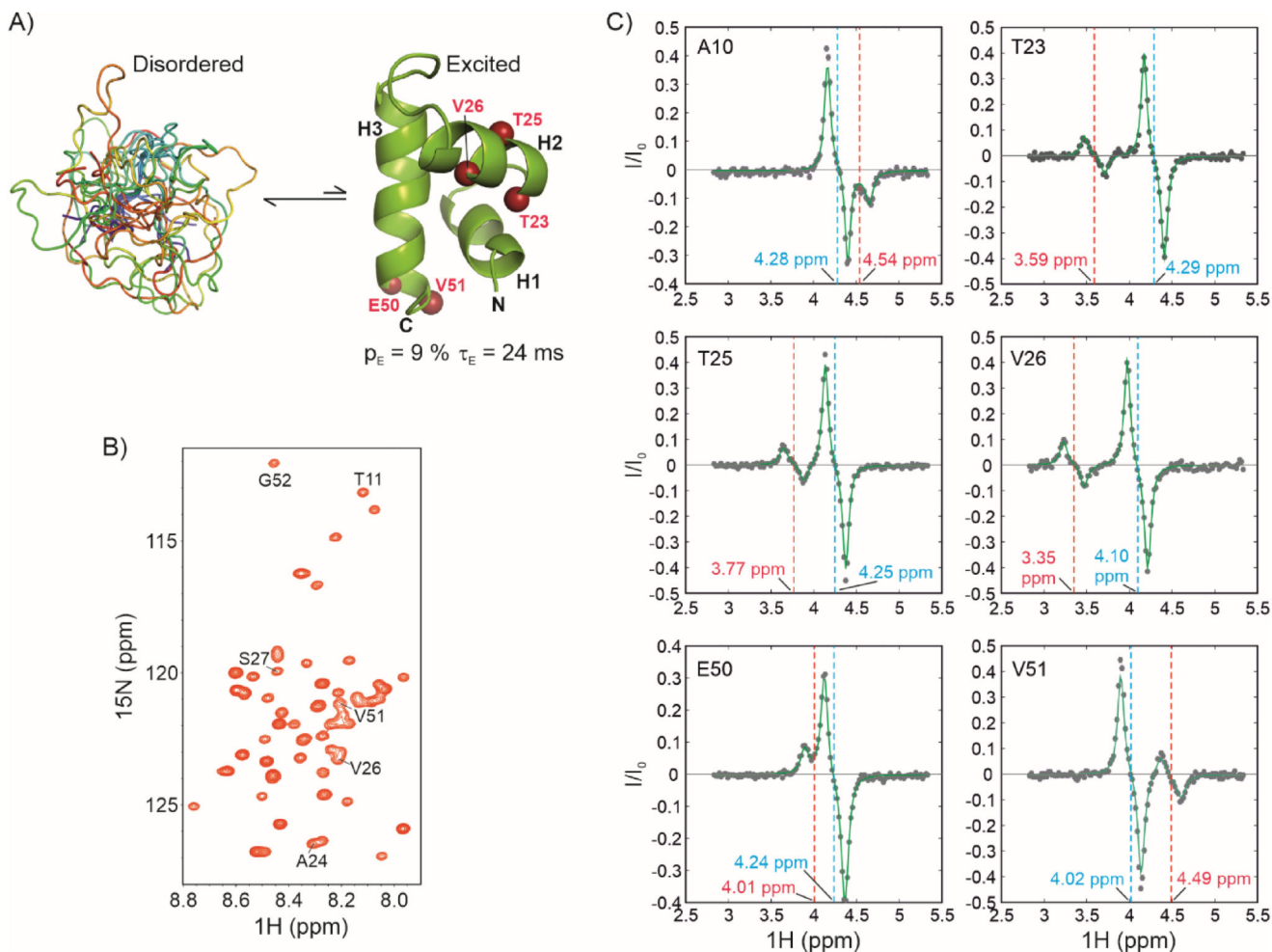
Figure 5. Characterizing exchange in the IDP CytR^N using ^{1}Hα CEST.
A) CytR^N exchanges between a disordered ensemble and a three-helix bundle folded state
(52). B) ^{1}H-^{15}N HSQC spectrum of disordered CytR^N showing limited spectral dispersion
characteristic of an IDP. Resonances for which ^{1}Hα CEST profiles are quantified in panel C
are labelled on the spectrum (except A10, which is not a part of the fragment included in the
structure calculation). C) ^{1}Hα CEST profiles of selected residues of CytR^N. The chemical
shift positions of the ground (disordered) and excited (folded) states are indicated by cyan
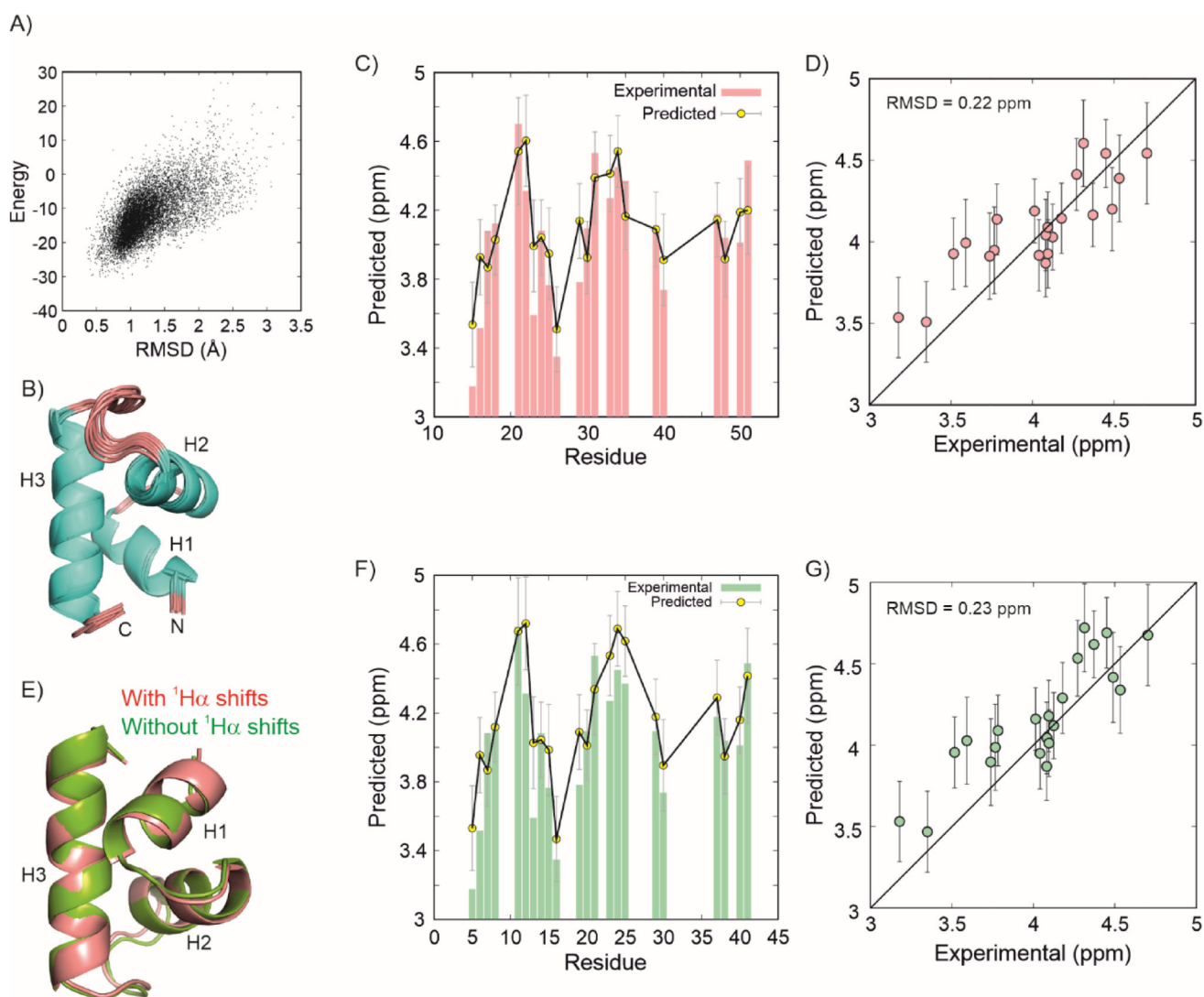and red dashed lines respectively.

Figure 6. Structure of the CytR^N excited state calculated chemical shifts and RDCs.
A) Composite chemical shift and Rosetta energy score of structures plotted against the RMSD of each structure to the lowest energy structure. B) Overlay of the cartoon representations of the 10 structures with the lowest energy. Comparison between the experimental (red bars) and predicted ^{1}Hα chemical shifts (yellow circles) for the CytR^N excited state shown as a function of residues (C) and as a correlation plot (D). Predictions were done using Sparta+ on the lowest energy structure. Error bars are uncertainties in residue-specific chemical shift predictions from Sparta+. E) Overlay of the CytR^N excited state calculated without (green) and with (red) ^{1}Hα chemical shifts. F,G) A comparison of the same experimental data as in panels (C) and (D) with predictions made using Sparta+ (84) on the lowest energy structure calculated without using input ^{1}Hα chemical shifts (52).