

Cite this: *Chem. Commun.*, 2019, 55, 7820Received 7th May 2019,  
Accepted 10th June 2019

DOI: 10.1039/c9cc03530c

rsc.li/chemcomm

## Rapid NMR assignments of intrinsically disordered proteins using two-dimensional $^{13}\text{C}$ -detection based experiments†

 Sujeesh Sukumaran,<sup>ab</sup> Shahid A. Malik,<sup>ab</sup> Shankararama Sharma R.,<sup>bc</sup>  
Kousik Chandra<sup>\*d</sup> and Hanudatta S. Atreya<sup>id</sup> <sup>\*b</sup>

**An approach for rapid backbone resonance assignments in proteins using only two 2D NMR experiments is presented. The new method involves a combination of high-resolution  $^{13}\text{C}$ -detected NMR experiments and selective unlabeled amino acid residues. The  $^{13}\text{C}$  detected 2D hNCA and 2D hNcoCA spectra of a uniformly labeled sample of the protein are analysed in concert with the 2D hNCA spectrum obtained for a selectively unlabeled sample. The combinatorial set of amino acid residues for selective unlabeled is chosen optimally to maximize the assignments. The method is useful for rapid assignment of proteins with low stability such as intrinsically disordered proteins and is applicable to deuterated proteins. This approach helped in assignments of 14.5 kDa human  $\alpha$ -synuclein during the course of its aggregation.**

Intrinsically disordered proteins (IDPs) are an important class of proteins comprising ~30% of the total proteins in organisms, and characterized by the lack of a well-defined three-dimensional (3D) structure.<sup>1</sup> IDPs are involved in various biological functions, such as regulation of transcription and translation, storage of small molecules, regulation of self-assembly of large multi-protein complexes, *etc.*<sup>2</sup> Due to the absence of a well-defined 3D structure and high flexibility, structural studies of IDPs are difficult using most of the structural biology techniques.<sup>3</sup> NMR spectroscopy plays a significant role in the study of IDPs, because it is the only method that allows a high resolution description of structural and dynamic features of IDPs in solution.<sup>4</sup> The potential for NMR spectroscopy to study the dynamic properties of proteins under physiological conditions makes it a powerful technique to understand the functional properties of IDPs.<sup>5,6</sup>

The first step towards the structure determination of proteins using NMR spectroscopy is the assignment of resonances observed

in the NMR spectra to individual nuclei of the protein.<sup>7</sup> Although conventional approaches for connecting sequential spin systems using 3D and higher dimensional NMR experiments have been successful for structured proteins, they are time-consuming and become ineffective in the presence of spectral overlap or if the protein aggregates or degrades rapidly such as in the case of IDPs.<sup>8</sup> The process of resonance assignments in IDPs is thus challenging due to their unique properties.<sup>9</sup> Chemical shift dispersion is much less in the conventional  $^1\text{H}$ -detected experiments of IDPs. The signal overlap in the spectra of IDPs strongly limits the possibility of standard amide proton-detection based NMR experiments to provide sufficient resolution for their characterization.<sup>10</sup> Heteronuclear NMR experiments based on  $^{13}\text{C}$  direct detection are useful in such cases, as they display a better chemical shift dispersion compared to  $^1\text{H}$ -detected experiments.<sup>11</sup>

In general, the conventional approach involves analysis of the NMR spectra obtained from multidimensional triple resonance experiments in concert, and linking the backbone  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{15}\text{N}$  chemical shifts of a continuous stretch of amino acids and mapping them onto the primary sequence of the protein for obtaining sequence-specific resonance assignments.<sup>7</sup> When there is considerable spectral overlap like in the case of larger molecular weight proteins or IDPs, it is necessary to selectively identify spin systems corresponding to the different amino acid types to aid resonance assignments. Amino acid selective “unlabeling” is an efficient method for the selective identification of the type of spin systems.<sup>12</sup> In this method, selective unlabeled of a set of amino acid type(s) in a polypeptide is performed by feeding the host microorganism with nutrients containing NMR active atoms.  $^{15}\text{NH}_4\text{Cl}$  is used as a source of nitrogen and  $^{13}\text{C}$  ( $^1\text{H}/^2\text{H}$ )-D-glucose is used as a source of carbon. The desired set of amino acid(s) in the unlabeled form (*i.e.*, containing  $^{12}\text{C}$  and  $^{14}\text{N}$ ) are also added to the growth medium for protein expression.<sup>13</sup> As a result, peaks due to these unlabeled residues will not be observed in the double and triple resonance spectra. The peaks corresponding to the unlabeled and labeled residues can be distinguished by comparing the spectrum recorded on the unlabeled sample with a control spectrum involving

<sup>a</sup> Solid State and Structural Chemistry Unit (SSCU), Indian Institute Of Science, Bangalore – 560 012, India

<sup>b</sup> NMR Research Centre, Indian Institute of Science, Bangalore – 560 012, India.  
E-mail: hsatreya@iisc.ac.in

<sup>c</sup> IISc Mathematics Initiative, Indian Institute of Science, Bangalore – 560 012, India

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c9cc03530c



**Fig. 1** Magnetization transfer pathway in (a) 2D hNCA providing  $(C_i^\alpha, N_i)$  and  $(C_i^\alpha, N_{i+1})$  chemical shift correlations and (b) 2D hNcoCA providing  $(C_i^\alpha, N_{i+1})$  shift correlations.

uniformly  $^{13}\text{C}$  or/and  $^{15}\text{N}$  labeled protein.<sup>12</sup> Selective unlabeled is used in combination with other NMR methods to help in rapid assignment of resonances observed in the NMR spectra.<sup>12–15</sup>

By exploiting the advantages of  $^{13}\text{C}$  detection and optimal combinatorial selective unlabeled<sup>14</sup> of amino acid residues, rapid resonance assignments can be obtained as shown here. In this method, we use two 2D NMR experiments:  $^{13}\text{C}^\alpha$  detected 2D hNcoCA and  $^{13}\text{C}^\alpha$  detected 2D hNCA. These experiments are recorded on a ( $^{13}\text{C}$ ,  $^{15}\text{N}$ ) uniformly labeled and a selectively unlabeled protein sample. The 2D hNCA provides chemical shift correlation of  $C^\alpha$  of residue  $i$  with that of  $^{15}\text{N}$  of residue  $i$  ( $C_i^\alpha, N_i$ ) and  $^{15}\text{N}$  of residue  $i+1$  ( $C_i^\alpha, N_{i+1}$ ) (Fig. 1a). Of these two correlations, the  $(C_i^\alpha, N_{i+1})$  peak appears lower in intensity or may be absent. This can be recovered from the 2D hNcoCA spectrum, which gives a correlation between  $C^\alpha$  of residue  $i$  with  $^{15}\text{N}$  of only residue  $i+1$  ( $C_i^\alpha, N_{i+1}$ ) (Fig. 1b). The 2D hNcoCA thus helps distinguish between  $N_i$  and  $N_{i+1}$  peaks in the 2D hNCA for a given  $C_i^\alpha$ . The 2D spectra are analysed in a combined manner yielding identification/assignment of a long peptide in a directional manner which can be mapped onto the primary sequence. The connected stretch of amino acids is categorized

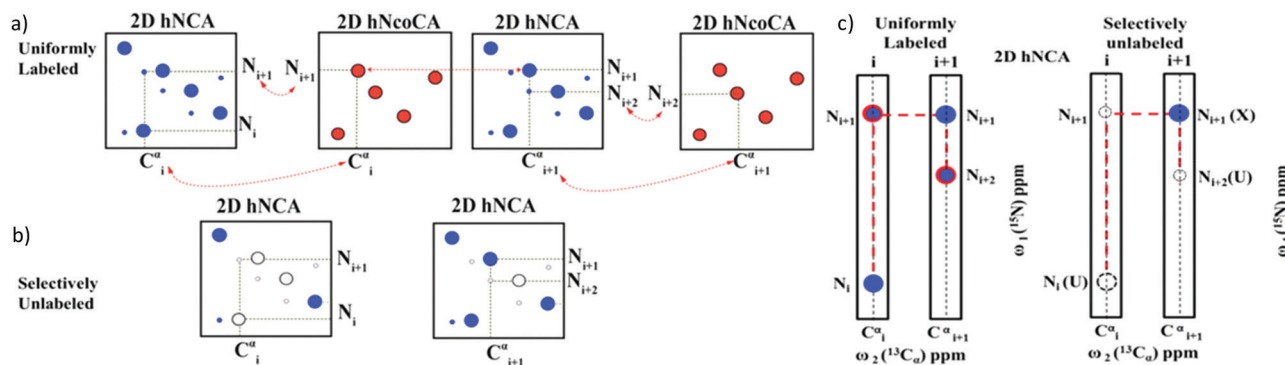
into specific patterns of X and U (X-labeled residue and U-unlabeled residue) with the help of selective unlabeled. If the pattern is unique in the sequence, then it is directly mapped onto the protein primary sequence to obtain backbone resonance assignment. The amino acid types corresponding to X and U need not be known.

Two samples are used: a uniformly  $^{13}\text{C}$ ,  $^{15}\text{N}$  labeled sample, and a ( $^{13}\text{C}$ ,  $^{15}\text{N}$ ) labeled sample with simultaneous unlabeled of multiple amino acid types chosen optimally.

The assignment strategy is depicted in Fig. 2a and b. For any given residue  $C_i^\alpha$  as a starting point, 2D hNCA gives  $(C_i^\alpha, N_i)$  and  $(C_i^\alpha, N_{i+1})$  correlations. The 2D hNcoCA helps identify the  $(C_i^\alpha, N_{i+1})$  correlation. Next, using the  $N_{i+1}$  shift thus identified, a search is carried out in 2D hNCA to identify the  $(C_{i+1}^\alpha, N_{i+1})$  correlation (as indicated by the horizontal dotted line in Fig. 2a). Having identified  $C_{i+1}^\alpha$  from 2D hNCA, the  $N_{i+2}$  can be identified based on 2D hNcoCA at the  $C_{i+1}^\alpha$  position as shown, and the assignment can be continued. In the same manner, starting from  $(C_i^\alpha, N_i)$  one can also proceed in the reverse direction.

With the help of these two experiments we will be able to obtain a sequential connectivity between the peaks (Fig. 2c). These sequentially connected amino acids form a stretch which has to be directly mapped onto the primary sequence. We have to identify the type of amino acid residues for mapping onto the sequence. The 2D hNCA spectrum of the selectively unlabeled sample helps identify the amino acid type. A peak can be categorized into X (labeled) or U (unlabeled) based on whether it is present or absent in the 2D hNCA spectrum of the selectively unlabeled sample (Fig. 2b). Note that 2D hNcoCA of the selectively unlabeled sample is not necessary, because the residues that are selectively labeled or unlabeled can be identified from the hNCA spectrum itself. Thus, only three 2D spectra are recorded (2 on the uniformly labeled sample and one on the selectively unlabeled sample).

The connected stretch of resonances as shown in Fig. 2c can be written in terms of X and U. The primary sequence of amino acids can also be translated to the sequence of X and U based



**Fig. 2** The assignment strategy using the 2D hNCA (blue) and 2D hNcoCA (red) spectra is illustrated for a tri-peptide stretch: U ( $i$ )-X ( $i+1$ )-U ( $i+2$ ), where X corresponds to a labeled residue and U corresponds to any selectively unlabeled residue. (a) A sequential connectivity between the peaks can be established by linking them as shown. Peaks depicted by large and small circles correspond, respectively, to self  $(C_i^\alpha, N_i)$  and sequential  $(C_i^\alpha, N_{i+1})$  correlations. (b) Based on the presence (shown by filled circles) or absence (empty circles) of  $(C_i^\alpha, N_i)/(C_i^\alpha, N_{i+1})$  correlations in the 2D hNCA spectrum recorded on the selectively unlabeled sample, a particular residue can be identified as belonging to labeled (X) or unlabeled (U) amino acid category. (c) A strip plot shows the connectivity: U ( $i$ )-X ( $i+1$ )-U ( $i+2$ ) established using (a) and (b).

on the choice of amino acids for selective unlabeled. The sequentially connected stretch of amino acids forms a pattern of X and U (e.g., UXU). If this pattern is unique in the primary sequence (translated to the sequence of X and U), then we can directly map it onto the sequence to obtain backbone resonance assignments as shown below (also see Fig. S1, ESI†).

The choice of amino acids to be unlabeled is decided based on an in-house developed python-based computer program (available at <http://nrc.iisc.ernet.in/hsa/software.htm>) and based on the possibilities of isotope scrambling of  $^{15}\text{N}$  or  $^{14}\text{N}$  among amino acids during expression of the protein in *E. coli*.<sup>12</sup> We have previously shown that based on the amino acid composition, a subset of amino acids from the following combinations: {Arg, Lys, Asn, Thr, Gly, Ser, Ala} can be chosen in any given protein for optimal selective unlabeled.<sup>13</sup> These amino acids either do not cross-metabolize to other amino acid residues in the bacterial biosynthetic pathway or do so within the same set (e.g., Gly and Ser cross-metabolize to each other). This avoids the unlabeled of undesired residues. The program tries these combinations or their subsets taking the primary sequence of a protein as input. It then translates the amino acid sequence into a sequence of X, U and 1, where X is a labeled residue, U is a selectively unlabeled residue and 1 is assigned for proline. The program searches along the sequence for identifying unique amino acid stretches of different lengths: hexa, penta, tetra, tri, and di-peptide stretches and outputs the amino acid residues that can be assigned. This helps in selecting the best combination of amino acids to be unlabeled. The algorithm of the program is depicted in Fig. S2 of the ESI.† The set {Thr, Gly, Ser, Ala} was found to be optimal for selective unlabeled in  $\alpha$ -synuclein; this constitutes 36.4% of residues in the protein.

Fig. 3 shows the 2D spectra obtained for  $\alpha$ -synuclein (complete acquisition parameters are provided in Table S1, ESI†). Fig. 4a shows an example of sequential connection among a set of residues in  $\alpha$ -synuclein. This was carried out using the strategy depicted in Fig. 2 (i.e.,  $i - i + 1 - i + 2 \dots$ ). Based on the presence/absence of peaks in the 2D hNCA spectrum obtained for the selectively unlabeled sample (Fig. 4b), the stretch of amino acids is classified as XUXU. The XUXU pattern is unique in the protein sequence from residue number 63 to 68 (V63, T64, N65, V66, G67, G68). This stretch of amino acids can thus be directly assigned. The results of this method were validated by comparing the results with the assignment done on the protein using the traditional assignment strategy.<sup>16</sup>

Using this approach 121 spin systems were identified out of expected 134 spin systems at 283 K. The stretches of residues assigned along the polypeptide are depicted in Fig. S3 of the ESI.† Since  $\alpha$ -synuclein is an intrinsically disordered protein and its aggregation rate (40–70 hours) is high under physiological conditions,<sup>17</sup> the assignment of this protein using traditional methods is a difficult task. Our method is more effective in this case because it is fast and due to the high dispersion of peaks in  $^{13}\text{C}$ -detected experiments.<sup>18</sup>

Rapid protein assignments are of high utility where measurement time is crucial because of limited sample stability. Our method is rapid since it uses only two 2D NMR experiments



Fig. 3 (a) The complete 2D hNCA spectrum of the uniformly labeled sample, (b) 2D hNCA of selectively unlabeled sample and (c) 2D hNCoCA spectrum of the uniformly labeled sample. Note that in this approach, the 2D hNCoCA spectrum of the selectively unlabeled sample is not needed (see text).

which together can be completed in less than 20 hours of experiment time. The experimental time can be further reduced using non-uniform sampling.<sup>19</sup> This method is useful not only in the case of IDPs, but also for structured proteins or for assignments of partially folded structures/intermediates formed during the course of protein folding or aggregation.<sup>20–22</sup> Further the method is applicable to (per)deuterated proteins. The process of peak picking, sequential connection, pattern identification and sequence mapping can be accelerated with the help of automation.

A drawback of  $^{13}\text{C}$  detection with respect to  $^1\text{H}$  detection is its intrinsically lower sensitivity due to the lower gyromagnetic ratio of  $^{13}\text{C}$  with respect to that of  $^1\text{H}$ . Further loss of sensitivity can occur due to  $^{13}\text{C}^\alpha$  transverse ( $T_2$ ) relaxation. In the two experiments, the  $^{13}\text{C}^\alpha$  resides in the transverse plane during detection and for  $\sim 25$  ms within the pulse sequence.<sup>23</sup> However, due to their high flexibility, the rotational correlation time of IDPs is lower compared to that of structured proteins of comparable size leading to higher sensitivity and sharper peaks. For instance, the rotational correlation time of 14.5 kDa  $\alpha$ -synuclein is 4.3 ns at 283 K,<sup>24</sup> which is equivalent to that of an  $\sim 8$  kDa globular structured protein. Thus, the 2D experiments are expected to exhibit sufficient sensitivity for larger size IDPs. Sensitivity can be further enhanced by deuteration and using cryogenic probes optimized for  $^{13}\text{C}$ -detection.

