

# AUTOBA: Automation of backbone assignment from HN(C)N suite of experiments

Aditi Borkar · Dinesh Kumar · Ramakrishna V. Hosur

Received: 16 November 2010 / Accepted: 12 May 2011 / Published online: 29 May 2011  
© Springer Science+Business Media B.V. 2011

**Abstract** Development of efficient strategies and automation represent important milestones of progress in rapid structure determination efforts in proteomics research. In this context, we present here an efficient algorithm named as AUTOBA (Automatic Backbone Assignment) designed to automate the assignment protocol based on HN(C)N suite of experiments. Depending upon the spectral dispersion, the user can record 2D or 3D versions of the experiments for assignment. The algorithm uses as inputs: (i) protein primary sequence and (ii) peak-lists from user defined HN(C)N suite of experiments. In the end, one gets  $H^N$ ,  $^{15}N$ ,  $C^\alpha$  and  $C'$  assignments (in common BMRB format) for the individual residues along the polypeptide chain. The success of the algorithm has been demonstrated, not only with experimental spectra recorded on two small globular proteins: ubiquitin (76 aa) and M-crystallin (85 aa), but also with simulated spectra of 27 other proteins using assignment data from the BMRB.

**Keywords** Structural proteomics · HN(C)N · Backbone assignment · Start points · Checkpoints · Automated assignment · AUTOBA

## Abbreviations

NMR	Nuclear magnetic resonance
HSQC	Heteronuclear single quantum correlation
BMRB	Biological magnetic resonance data bank
CARA	Computer aided resonance assignment (a software for NMR data analysis)
AUTOBA	Automatic backbone assignment

## Introduction

Biomolecular NMR spectroscopy has expanded dramatically in recent years and is now a powerful technique in proteomics research. The key step in this process is the sequence specific backbone assignment and getting this information fast is crucial, especially for proteins that are unstable in solution or tend to precipitate in matter of days. With this view, methodological developments are occurring continuously and many new pulse sequences have been developed to save on experimental time (Lescop and Brutscher 2007; Kupce and Freeman 2003b; Kupce and Freeman 2003a; Freeman and Kupce 2003; Kim and Szyperski 2003; Frydman et al. 2002; Jaravine et al. 2008; Rovnyak et al. 2004; Felli and Brutscher 2009; Kupce and Freeman 2003c; Atreya and Szyperski 2005; Lescop et al. 2007; Schanda et al. 2006; Hiller et al. 2005; Fiorito et al. 2006). The concomitant developments in NMR hardware have further increased the sensitivity and resolution, which not only help in saving time, but provide spectral data that are highly amenable for automated analysis by computer

Aditi Borkar and Dinesh kumar have contributed equally.

**Electronic supplementary material** The online version of this article (doi:10.1007/s10858-011-9518-0) contains supplementary material, which is available to authorized users.

A. Borkar · D. Kumar · R. V. Hosur (✉)  
Department of Chemical Sciences, Tata Institute of Fundamental Research (TIFR), 1, Homi Bhabha Road, Colaba,  
Mumbai 400005, India  
e-mail: hosur@tifr.res.in

R. V. Hosur  
UM-DAE Centre for Excellence in Basic Sciences, Mumbai  
University Campus, Kalina, Santa Cruz, Mumbai 400098, India

programs. As a result, in the last few years, several automated algorithms have been introduced for fast data analysis (Jung and Zweckstetter 2004; Moseley et al. 2004; Coggins and Zhou 2003; Atreya et al. 2000; Friedrichs et al. 1994; Lukin et al. 1997; Lescop and Brutscher 2009; Crippen et al. 2010). In this context, we present here an efficient algorithm named as AUTOBA (*Automatic backbone assignment*) which has been designed to automate the assignment protocols based on HN(C)N suite of experiments that exploit the generally good  $^{15}\text{N}$  chemical shift dispersion (Panchal et al. 2001; Chatterjee et al. 2002; Chatterjee et al. 2006; Kumar et al. 2009; Kumar et al. 2010b; Kumar and Hosur 2010) seen in folded and unfolded proteins alike. The performance of AUTOBA has also been demonstrated here not only with experimental spectra recorded on two globular proteins: ubiquitin (76 aa) and M-crystallin (85 aa), but also with simulated spectra of 27 other proteins using assignment data from the BMRB.

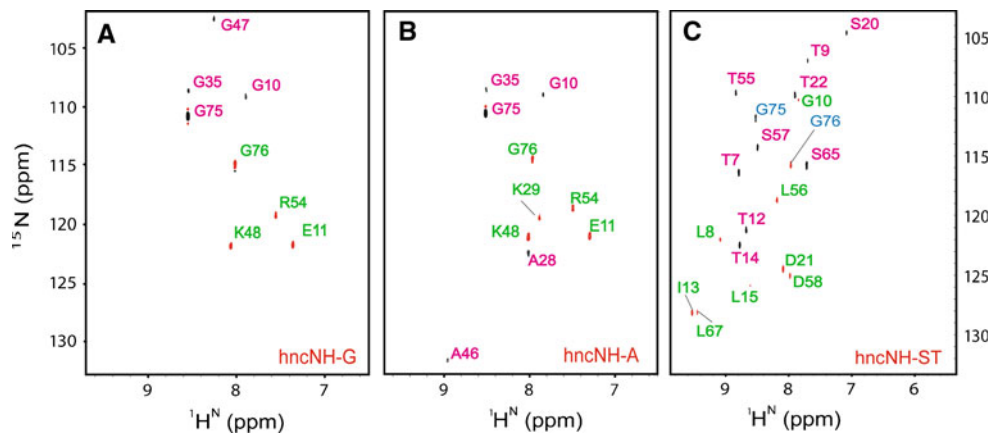
AUTOBA has been written in C language code and uses only standard C libraries. It can be compiled with  $g^{++}$  and no additional downloads are required for compiling the program and thus it can be run on multiple platforms. The AUTOBA program has also been given JAVA based graphical user interface (GUI) which helps to browse and upload the input files and finally to run the program. The software package can be downloaded freely from here: <http://www.tifr.res.in/~hosur/download.htm>. We have also developed the web-based server for AUTOBA usage, which can be accessed from here (<http://www.tifr.res.in/~hosur/AUTOBA.html>).

### NMR experiments used in AUTOBA

As the AUTOBA algorithm automates the backbone assignment strategy recently described by (Kumar and Hosur 2010), it uses just the (i) protein primary sequence, and (ii) peak-lists from user defined HN(C)N suite of experiments as inputs. These experiments can be the 2D-projections of the basic 3D experiments: 3D-HN(C)N (Chatterjee et al. 2002), 3D-hncoCANH (Kumar et al. 2010b), and 3D-hnCOcaNH (Kumar et al. 2010b) or the original 3D experiments themselves. Depending upon the spectral dispersion, the user can record 2D or 3D versions of the experiments and submit the peaks list into AUTOBA for assignment. For example, in case of small well folded proteins, all the information can be derived from 5 to 6 2D spectra only (total data collection time ranges from 8 to 12 h only), and thus will be of great value when the small protein under investigation is unstable. On the other hand, in more complex systems (e.g. large MW proteins or unfolded proteins), where the spectral dispersions in the 2D experiments described above may be insufficient, the

original 3D experiments (e.g. 3D-HN(C)N (Chatterjee et al. 2002), 3D-hncoCANH (Kumar et al. 2010b), and 3D-hnCOcaNH (Kumar et al. 2010b) would be required for the backbone assignment. The various experiments required for AUTOBA are:

1.  $^1\text{H}$ - $^{15}\text{N}$  projection plane of 3D-CBCACONH (Grzesiek and Bax 1992): This spectrum referred to here as 2D-cbcacoNH produces  $\text{H}^{\text{N}}$ - $^{15}\text{N}$  HSQC type spectrum but with exclusively backbone correlations.
2.  $F_2(^{15}\text{N})$ - $F_3(^1\text{H})$  plane of HN(C)N spectrum projected down the  $F_1(^{15}\text{N})$  axis (Kumar and Hosur 2010): This particular plane has been designated as 2D-hncNH spectrum (Kumar and Hosur 2010) and is used to identify the HSQC cross peaks corresponding to glycines, alanines, serines/threonines and the residues following them in the sequence (hereafter referred to as checkpoints). The spectrum can be recorded from the basic HN(C)N experiment by simply avoiding the  $t_1$  evolution (Kumar and Hosur 2010). Depending upon the variant of the 3D-HN(C)N experiment used, HN(C)N-G (Panchal et al. 2001) or HN(C)N-A (Chatterjee et al. 2006) or HN(C)N-ST (Kumar et al. 2009), these planes are designated as, 2D-hncNH-G, 2D-hncNH-A, and 2D-hncNH-ST, respectively, (Fig. 1). The 2D-hncNH-G spectrum (Fig. 1a) provides identification of peaks belonging to glycines (referred to as primary checkpoints here), and residues following them in the sequence (referred to as secondary checkpoints here); the 2D-hncNH-A spectrum (Fig. 1b) shows peaks belonging to alanines and glycines (primary checkpoints) and residues following them in the sequence (secondary checkpoints); the 2D-hncNH-ST spectrum (Fig. 1c) shows peaks belonging to serines/threonines (primary checkpoints) and residues following them in the sequence (secondary checkpoints). In these spectra, the primary and secondary checkpoints are distinguished by their opposite signs; the primary checkpoints appear positive in sign and the secondary checkpoints appear negative in sign. However, it turns out that if a primary (Glycines, Alanines, and Serines/Threonines) or secondary (residues following Glycines, Alanines, and Serines/Threonines) checkpoint residue in the protein sequence is followed by a proline residue or if it is present at the end of C-terminal arm, the characteristic sign patterns will be altered because of the alterations in peak intensity additions when projections are taken (Kumar and Hosur 2010). Further, if the protein sequence contains stretches like GG, [G/A][G/A] or [S/T][S/T], the peak for second G, G/A or S/T will be absent, in 2D-hncNH-G, 2D-hncNH-A, and 2D-hncNH-ST spectra, respectively. All these features are evident in Fig. 1. In such



**Fig. 1** 2D-hnCNH-G (a), 2D-hnCNH-A (b), and 2D-hnCNH-ST (c) spectra of ubiquitin (1.6 mM, 76 aa) recorded on a 800 MHz Bruker spectrometer (equipped with cryo-probe) at 298 K. The black (positive peak sign) and red (negative peak sign) contours represent primary and secondary checkpoints, respectively. The primary and secondary checkpoints have been shown by pink and green labels,

respectively. In the 2D-hnCNH-G spectrum (a), the peak for Ile36 (following Gly35) is missing as it is followed by a proline (Pro37) in the sequence. In the 2D-hnCNH-A spectrum (b) Gly47 peak is missing as it is preceded by Ala46 in the sequence. Similarly, in the 2D-hnCNH-ST spectrum (c) Thr66 peak is missing as it is preceded by Ser65 in the sequence

situations the variants of 2D-(HN)NH spectrum (Kumar et al. 2010a) are preferred over the variants of 2D-hnCNH spectrum.

3.  $F_1(^{15}\text{N})-F_3(^1\text{H})$  plane of HN(C)N spectrum projected down the  $F_2(^{15}\text{N})$  axis (Kumar and Hosur 2010): This particular plane has been designated as 2D-hNcnH spectrum (Kumar and Hosur 2010) and provides intra- ( $H_i, N_i$ ) and inter- ( $H_i, N_{i+1}$ ) residue amide correlations directly on the HSQC type spectra. These intra- ( $H_i, N_i$ ) and inter- ( $H_i, N_{i+1}$ ) residue amide correlations can be easily identified by their opposite sign: intra-residue ( $H_i, N_i$ ) correlation peaks appear positive in sign and inter-residue ( $H_i, N_{i+1}$ ) correlation peaks appear negative in sign. This particular spectrum can be recorded from the basic HN(C)N experiment by avoiding the  $t_2$  evolution (Kumar and Hosur 2010). Depending upon the variant of 3D-HN(C)N experiment used -either HN(C)N-G (Panchal et al. 2001) or HN(C)N-A (Chatterjee et al. 2006) or HN(C)N-ST (Kumar et al. 2009)- these planes are designated as: 2D-hnCNH-G, 2D-hnCNH-A, and 2D-hnCNH-ST, respectively. These variants of 2D-hNcnH spectrum carry with them the triplet specific peak patterns of the basic HN(C)N which help to identify certain specific triplet sequences along the sequence, and thus provide checkpoints for the sequential assignment process. However, in case when more than one residue has identical amide  $^1\text{H}^{\text{N}}$  chemical shifts, the identification of checkpoint becomes ambiguous. In order to avoid this problem, the checkpoints are identified from variants of 2D-hnCNH/2D-(HN)NH spectrum (Kumar and Hosur 2010, Kumar et al. 2010a).
4.  $F_1(^{13}\text{C})-F_3(^1\text{H})$  planes of 3D hnCOcANH and 3D hnCOcANH spectra (Kumar et al. 2010b) projected

down the  $F_2(^{15}\text{N})$  axis: These particular planes have been designated as 2D-hnCOcANH and 2D-hnCOcANH spectra and provide intra- ( $H_i, C_i$ ) and inter- ( $H_i, C_{i-1}$ ) residue correlations directly on the  $^1\text{H}-^{13}\text{C}$  HSQC type spectra. These correlations can be easily distinguished because of their opposite signs: intra-residue ( $H_i, C_i$ ) correlation peaks appear positive in sign and inter-residue ( $H_i, C_{i-1}$ ) correlation peaks appear negative in sign. The 2D-hnCOcANH and 2D-hnCOcANH spectra can be recorded from the basic 3D hnCOcANH and 3D hnCOcANH experiments, respectively, by avoiding the  $t_2(^{15}\text{N})$  evolution.

## AUTOBA protocol and working

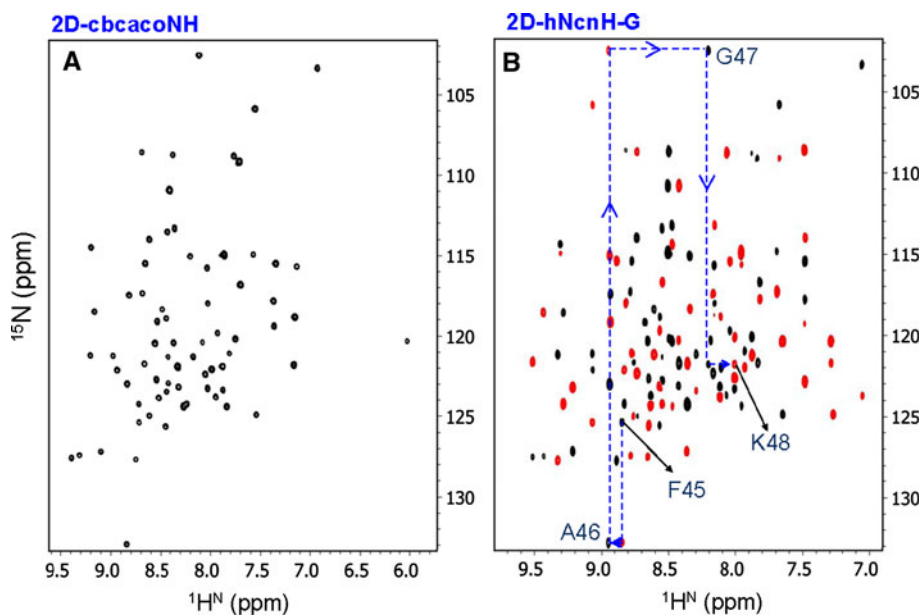
1. AUTOBA input files: The algorithm starts by reading in the input data. This comprises of the protein primary sequence in standard single letter code and the peak lists from the HN(C)N suite of experiments. These peak lists are: (i) 2D-cbcacoNH peak-list for backbone only amide chemical shifts (Fig. 2a), (ii) peak-list derived from 2D-hNcnH or 3D-HN(C)N spectrum for sequential amide correlations (Fig. 2b), (iii) peak-lists derived from variants of 2D-hnCNH spectrum for identifying spin systems corresponding to glycines, alanines, and serines/threonines and the residues following them in the sequence (Fig. 1); this information can also be given manually in the form a checkpoint file (the format of input file can be seen from demo input data provided along with the software package) using the notations for various amino acid types given in (Fig. 3) and (iv)

the peak-lists derived from 2D-hnCOcanH and 2D-hncoCANH spectra or from 3D-hnCOcanH and 3D-hncoCANH spectra (Fig. 4) for extracting assignments of  $^{13}\text{C}^\alpha$  and  $^{13}\text{C}'$  chemical shifts. The peak-lists used in AUTOBA are in the format as generated by peak picking module of CARA (Keller 2004). The procedure of generating various input files has been explained in detail at this site: “<http://www.cbmr.res.in/autoba.html>”. Note that, the 2D peak list file should contain the chemical shift as well as intensity information (to distinguish positive and negative peaks) for all the peaks. The format of input files can also be seen in their respective panels on the server page.

- Identifying self peaks: The AUTOBA algorithm mainly works on the 2D-hNcnH or the 3D-HN(C)N peak list input. If 2D-hNcnH peak list is submitted, the algorithm first distinguishes all self peaks from the cross peaks in the spectrum. This is done by matching the chemical shifts of the peaks in the 2D-hNcnH spectrum with those in the 2D-cbcacoNH (Fig. 2a). The matches are identified as “self” peaks and the remainder is identified as “cross” peaks. In case of 3D-HN(C)N, the self and cross peaks are readily identified from the  $^{15}\text{N}$  shifts in the  $F_1$  and  $F_2$  planes.
- Identifying checkpoints: The chemical shifts from the 2D-hncNH-G, 2D-hncNH-A and 2D-hncNH-ST are matched with the self peaks identified from the 2D-hNcnH or 3D-HN(C)N spectrum from step 2. Note

that both the checkpoint peak ( $i$ ) and the sequential to this checkpoint peak ( $i + 1$ ) (the secondary checkpoint) is seen in the checkpoint peak list but are distinguished from each other by their opposite sign of the intensities. Thus, for each match between the spectra, the intensity of the checkpoint peak is examined. If the sign of the intensity is equal to that of the checkpoint, the matched 2D-hNcnH or 3D-HN(C)N peak is identified as the primary checkpoint and is given a new index number; otherwise it is identified as a secondary checkpoint. The secondary checkpoints are given the index number = index number of the primary checkpoint +26. The index number is chosen as the number of the single letter code of the amino acid in the English alphabets. For example, if a peak is identified as a Serine or Threonine, (U), it is given an index 21 and its sequential will be indexed as 47.

- Establishing sequential connectivities: Sequential amide correlations are established from high-resolution 2D-hNcnH or 3D-HN(C)N spectra following the path from intra-residue correlation [ $^1\text{H}_i^{\text{N}} - ^{15}\text{N}_i$ ] via inter-residue correlation [ $^1\text{H}_i^{\text{N}} - ^{15}\text{N}_{i+1}$ ] to the next intra-residue correlation [ $^1\text{H}_{i+1}^{\text{N}} - ^{15}\text{N}_{i+1}$ ]; i.e. starting from an  $i^{\text{th}}$  peak in the 2D-hNcnH spectrum, it is connected to a cross peak if both their  $^{\text{N}}\text{H}$  chemical shifts are matched with some limit of the chemical shift tolerance. This



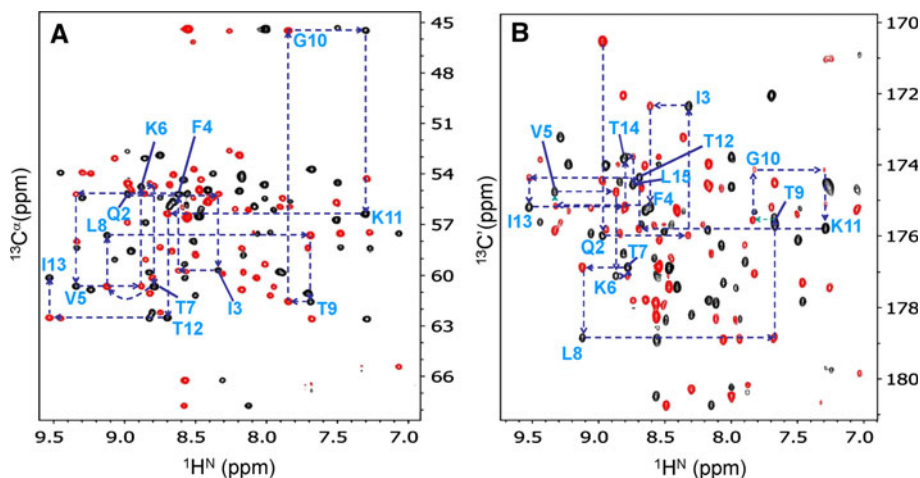
**Fig. 2** **a** 2D-cbcacoNH spectrum of ubiquitin showing backbone only  $^{15}\text{N}$ - $^1\text{H}$  correlation peaks and **b** the 2D-hNcnH spectrum of ubiquitin showing intra-residue ( $H_i, N_i$ ) and inter-residue ( $H_i, N_{i+1}$ ) correlations. The illustrative sequential walk (F45 to K48) following the path: ( $H_i, N_i$ )  $\rightarrow$  ( $H_i, N_{i+1}$ )  $\rightarrow$  ( $H_{i+1}, N_{i+1}$ ) has been shown on

the spectrum in blue dotted line. Black and red colors represent positive and negative signs, respectively. In 2D-hNcnH spectrum, black and red colors represent self (except for residues which are present next to glycines) and sequential (except for residues which are present next to glycines) peaks, respectively



**Fig. 3** Ubiquitin primary sequence and its transformation to a format where all amino acids are replaced by letter X except for glycines, alanines and serines/threonines which are replaced, respectively, by letters G, A, and U

**Fig. 4** **a** 2D-hncoCAnH spectrum of 1.6 mM ubiquitin (76 aa) and illustrative stretches of sequential walk on this spectrum (Q2 to I13) shown with dotted blue lines, **b** 2D-hnCOcanH spectrum of 1.6 mM ubiquitin and illustrative stretches of sequential walk on this spectrum (Q2 to L15) shown with dotted blue lines. Black and red colors represent self and sequential peaks, respectively, except for residues which are present next to glycines and hence can easily be identified



cross peak is then connected to other self peak ( $i + 1$ )th if their amide N chemical shifts are matched with some limit of the chemical shift tolerance.

5. Selecting true positives and eliminating false branching: From step 4, each  $i$  may have multiple connections. However, if any self peak has been identified as a checkpoint, then it should be connected only to its sequential. Thus, in case of multiple connections emanating from a checkpoint peak in the 2D-hNcnH or 3D-HN(C)N spectra, all the  $i + 1$  peaks that are not identified as a corresponding secondary checkpoint to the primary checkpoint are deleted (Fig. 1). For example, if a Serine or Threonine, (U), is connected to three other self peaks, then the peaks with index 47 amongst these three peaks only will be stored and the others would be deleted.

Similarly the  $C'$  and  $C^\alpha$  peaks can be identified by matching the H chemical shifts from the 2D-hNcnH or 3D-HN(C)N spectrum with those from 2D-hnCOcanH or 2D-hncoCAnH spectra, respectively within the limits of chemical shift tolerance (Fig. 4). In case of multiple connections in the 2D-hNcnH or 3D-HN(C)N peaks from step 4, a match between the corresponding  $^{13}\text{C}$  chemical shifts will confirm the false positives and true positives amongst these connections. Using this discrimination in addition to the checkpoint information limits the branching and ambiguities in the final assignments.

6. Performing sequential walk: After making all the possible  $i \rightarrow i + 1$  connectivities, a sequential walk is performed such as  $i \rightarrow i + 1 \rightarrow i + 2 \rightarrow i + 3$  and so on to form a stretch. Each stretch is then coded such that only the checkpoints retain their single letter code letters while all the other amino acids are replaced by the letter “X”. For example, if the stretch formed represents the sequence GKALDGRRLSDY, then it will be coded as GXAXXGXUXUXX. The protein sequence too is coded in a similar fashion apart from the fact that prolines additionally are represented with letter “P” in the protein code as depicted in Fig. 3. Thus, the primary sequence will now look like —XXXXGXXXXAXXPXXUXXUXXXXU—. While a sequential walk is being performed, the stretch is coded in parallel and compared with the coded protein sequence, and if a match is found the stretch is stored. The sequential walk halts when (i) the stretch reaches a Proline or some other dead end due to missing peaks, or (ii) the stretch becomes redundant because one spin system in a coded stretch may connect to another spin system in other coded stretches or (iii) the coded stretch does not match with any part of coded protein sequence.
7. Removing ambiguous assignments: It might be possible that the stretches obtained during the sequential walk process from step 6 may match to multiple regions of the primary protein sequence giving rise to

ambiguities in final peak assignments. This would be particularly true for shorter stretches. For example a “UXUX” stretch can match to TITL (aa 12–15), SDTI (aa 20–23) and TLSL (aa 55–58) in the ubiquitin sequence. Thus, intelligent choices are made between these branches by AUTOBA by method of elimination. First, the longest stretch is identified which uniquely matches to a particular region of the protein. All other smaller stretches that overlap with this stretch are eliminated from the database. For example, “XXXXXUXUGXUXXXXX” is the longest stretch uniquely matching to QIFVKTLTGKTITLEVE (aa 2–18) region of ubiquitin. Thus, TITL will be eliminated as a possible match for “UXUX”. Iterations of this procedure with the next longest stretch and so on will finally eliminate many more branches.

8. Writing outputs: Finally, (i) the  $i \rightarrow i + 1$  connectivities, (ii) all the branches and (iii) all the peak assignments are written as output files. The peak assignments are written both as a text file and in the BMRB format. In case there are still any ambiguous assignments, all of them will be listed in the text format of the peak assignments file. For an example, refer to Tables S2 and S3 in supplementary information.

Thus, complete assignment for backbone  $^1\text{H}^{\text{N}}$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}^{\alpha}$ , and  $^{13}\text{C}'$  resonances can be established. The working of AUTOBA has been shown in the form of a flow chart in Fig. 5.

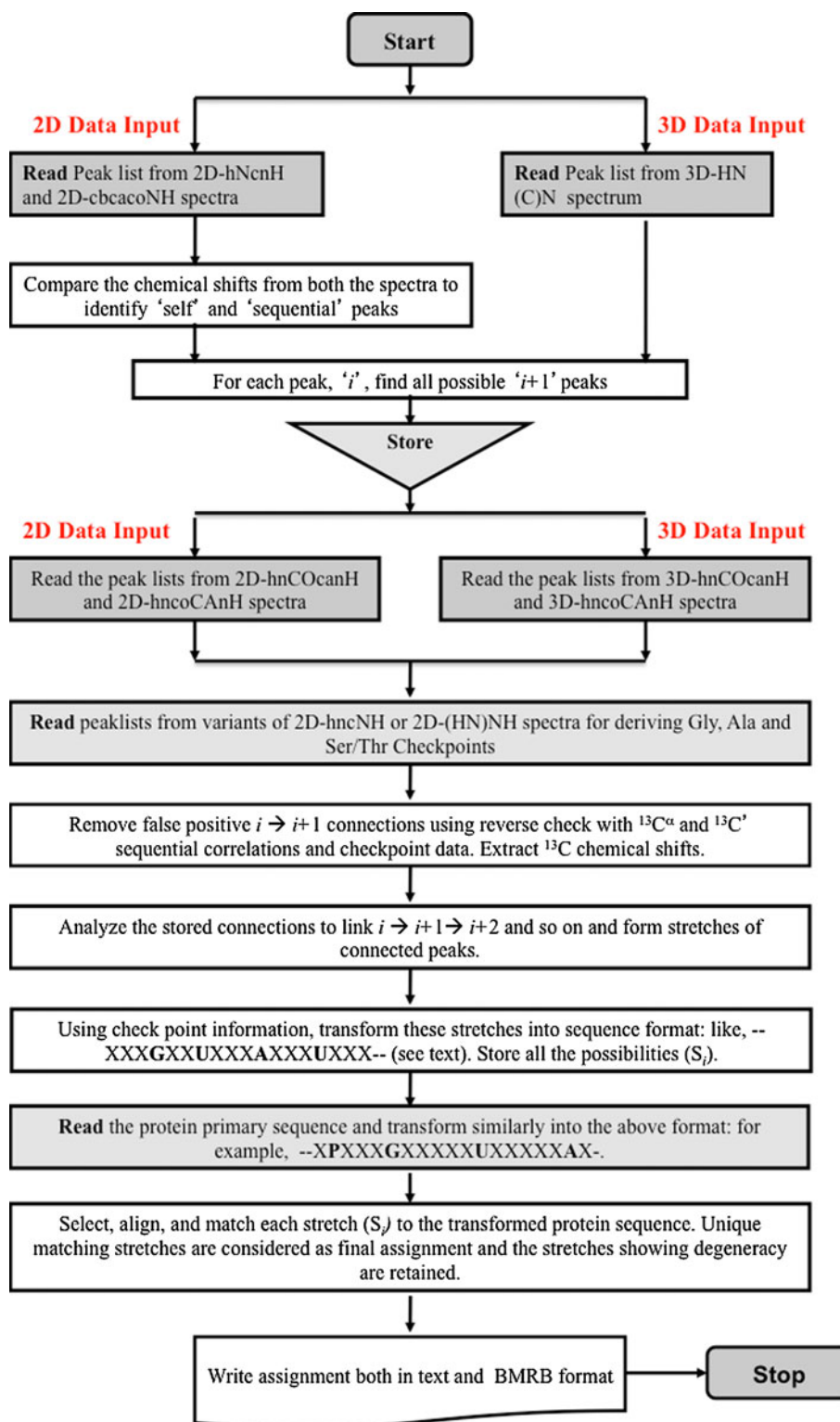
### Pitfalls and remedies

Having described the way AUTOBA works, it is prudent to consider now the possible pitfalls, how to detect wrong assignments and possible solutions. When the HSQC spectrum truly represents the finger print of the protein, with exactly the expected number of peaks which are well dispersed with no degeneracy of chemical shifts, the sequential connectivities in the experimental spectra would be unique, the assignments will be unambiguous and the confidence level can be 100%. However, this is an ideal situation. Often the number of peaks will not be as expected. There can be missing peaks for various reasons or there could be extra peaks due to slow conformational exchange or artifacts. There could be severe overlaps of peaks or degeneracies of some chemical shifts and then there could be errors in the assignments. In such a situation depending upon the distribution of the checkpoints along the sequence, the confidence levels of assignments can be different in different segments of the protein chain. If the checkpoints are closely placed, and all the sequential connections in the intervening regions are clearly seen in

the spectra, then the confidence level of assignment in those regions can be 100%. It is clear that a large number of checkpoints that are well dispersed along the length of the chain would surely reduce the errors. While AUTOBA automatically determines the checkpoints from the experiments described in this paper, it also accepts other checkpoint information derived from other experiments such as (Barnwal et al. 2008; Chugh and Hosur 2008; Dotsch et al. 1996; Grzesiek and Bax 1993; Lescop et al. 2008; Pantoja-Uceda and Santoro 2008; Schubert et al. 2001a; Schubert et al. 2001b); these will have to be added manually to the checkpoint list. The need for these can be readily determined at the beginning itself by looking at the distribution of glycines, alanines, serines and threonines along the chain. We may mention here that while working with the simulated spectra of twenty proteins from BMRB for testing the AUTOBA performance (see below) we did not need any other checkpoint information. The different sources of error and possible remedies are briefly described below.

- (i) Missing peaks: Peaks can be missing in the HSQC spectra due to chemical exchange mediated line broadening. Such a residue will be equivalent to a proline so far as AUTOBA is concerned. Sequential connections will stop at that residue, unless there is a degeneracy of NH or N chemical shifts and AUTOBA may continue with another peak on the same line. In the 2D hNcnH spectra, peaks can also be missing if a sequential peak overlaps with an HSQC peak since they have opposite signs. The latter can be readily detected by comparison with the HSQC spectrum. In the former situation, the errors in the sequential walk can be detected in most situations by recording spectra under more than one solution conditions such as temperature, salt, pH etc., ensuring at the same time that these changes do not radically alter the protein structure. Such information should be already available from previous biophysical studies on stability, thermal denaturation etc. on the protein. Temperature coefficients for different amide protons are generally different and thus small degeneracies will get lifted when the temperature is changed; sequentially connected pairs of peaks move together and the extent of this movement would be different for different residues. Recording the spectra at another higher field would also lift the degeneracies. Thus AUTOBA analysis should routinely be done with spectra recorded at more than one conditions.
- (ii) Overlapping peaks: The assignment breakdown is easier to notice in this case because AUTOBA tends to output multiple ambiguous assignments of small stretches (typically 3–5 residues) of the protein

**Fig. 5** Flowchart depicting the functioning of AUTOBA program



sequence for such input spectra. Overlaps can sometimes be resolved by changing the solution conditions. If that is not successful, the best way to minimize the errors would be to increase the resolution by increasing the field or by using TROSY

- (iii) Extra peaks: The HSQC spectra may show extra peaks due to slow conformational equilibria in certain segments of the chain. In such a situation

more than one set of connectivities would be found for the same segment. Extra peaks due to artifacts can be readily distinguished since they would not show any systematic connectivities.

Finally, after eliminating the peaks for which confidence level of assignment is 100%, it can happen that some uncertainties may remain when AUTOBA is not able to decide on a unique segment for a particular subset of input peaks even though connectivities between them are well established. AUTOBA would then output several possibilities. Since each of them would be equally probable, the confidence level here would be  $100/n$  where  $n$  is the number of possible matches. This would happen when the checkpoint information is not enough. In such a situation it would

become necessary to obtain side chain assignments corresponding to those HSQC peaks using the well established standard methods. If this is also insufficient, which can happen in flexible domains of proteins, checkpoints can be introduced in those stretches of the sequence by site directed mutagenesis or by selective labeling/unlabeling schemes.

### Performance of AUTOBA

The performance of AUTOBA was tested initially using simulated spectra from BMRB data sets (27 proteins, Table 1). The possibility of peak overlaps was checked using a cumulative chemical shift margin of 0.05 ppm (Eq. 1).

**Table 1** The performance of AUTOBA tested by executing automation procedure on simulated BMRB 2D data sets taken for twenty seven well folded globular proteins

	BMRB ID	# of residues	# of peaks	AUTOBA performance				% Accurate assignments
				2D data set				
				Accurate	Ambiguous	Wrong	Unassigned	
1	16260	126	121	118	0	0	8	97.52
2	16175	125	112	84	14	1	26	75.00
3	16233	123	111	102	4	0	17	91.89
4	6902	121	108	70	15	14	22	64.81
5	16147	115	100	82	9	2	22	82.00
6	4036	114	109	103	3	2	6	94.50
7	4964	110	100	91	6	0	13	91.00
8	4113	108	101	94	3	1	10	93.07
9	16713	104	103	95	1	0	8	92.23
10	4109	100	97	91	6	0	3	93.81
11	6304	100	93	93	0	0	7	100.00
12	15194	97	97	91	0	0	6	93.81
13	6801	96	85	80	4	1	11	94.12
14	4001	92	73	62	0	0	30	84.93
15	4105	92	86	53	0	16	23	61.63
16	15743	92	87	85	4	0	3	97.70
17	7202	90	87	81	4	0	5	93.10
18	16261	90	86	82	3	1	4	95.35
19	4931	89	88	87	0	0	2	98.86
20	16706	89	57	56	5	0	28	98.25
21	6903	85	80	80	0	0	5	100.00
22	6904	85	80	80	0	0	5	100.00
23	16177	80	69	65	0	7	8	94.20
24	4769	76	68	67	0	0	9	98.53
25	6697	75	41	37	0	0	38	90.24
26	4023	71	67	62	0	0	9	92.54
27	7280	58	58	57	0	0	1	98.28

The categories, accurate, ambiguous, wrong and unassigned in the Table, are mutually exclusive. The percentage assignment obtained in each case is shown in the last column. Following this, all the tested data has been kept here (<http://www.tifr.res.in/~hosur/AUTOBA.html>) for future demonstration and testing of the AUTOBA



**Table 2** The performance of AUTOBA tested by executing automation procedure on simulated BMRB 3D data sets taken for twenty seven well folded globular proteins

	BMRB ID	# of residues	# of peaks	AUTOBA performance				% Accurate assignments
				3D data set				
				Accurate	Ambiguous	Wrong	Unassigned	
1	16260	126	121	118	0	0	8	97.52
2	16175	125	112	83	15	2	25	74.11
3	16233	123	111	106	0	0	7	95.50
4	6902	121	108	81	11	9	20	75.00
5	16147	115	100	99	0	0	16	99.00
6	4036	114	109	106	0	0	8	97.25
7	4964	110	100	91	6	0	13	91.00
8	4113	108	101	101	0	0	7	100.00
9	16713	104	103	96	0	0	8	93.20
10	4109	100	97	97	0	0	3	100.00
11	6304	100	93	93	0	0	7	100.00
12	15194	97	97	91	0	0	6	93.81
13	6801	96	85	80	4	1	11	94.12
14	4001	92	73	62	0	13	17	84.93
15	4105	92	86	67	9	7	9	77.91
16	15743	92	87	86	0	3	3	98.85
17	7202	90	87	85	0	0	5	97.70
18	16261	90	86	81	4	2	3	94.19
19	4931	89	88	87	0	0	2	98.86
20	16706	89	57	56	0	5	28	98.25
21	6903	85	80	80	0	0	5	100.00
22	6904	85	80	80	0	0	5	100.00
23	16177	80	69	65	0	8	7	94.20
24	4769	76	68	67	0	0	9	98.53
25	6697	75	41	37	0	0	38	90.24
26	4023	71	67	64	0	0	7	95.52
27	7280	58	58	57	0	0	1	98.28

The categories, accurate, ambiguous, wrong and unassigned in the Table, are mutually exclusive. Following this, all the tested data has been kept here (<http://www.tifr.res.in/~hosur/AUTOBA.html>) for future demonstration and testing of the AUTOBA

$$\Delta\delta^{cum} = \sqrt{(\delta_H^i - \delta_H^j)^2 + (\delta_N^i - \delta_N^j)^2} \quad (1)$$

where  $i$  and  $j$  represent self and sequential peaks, respectively. The results showed that the overlap possibility of self and sequential peaks in these simulated 2D-hNcnH spectra is on the average 2.75%. Following this, the AUTOBA automation was run on all these proteins for evaluating the performance of the procedure. The results are summarized in Table 1 and Table 2. In many of the cases AUTOBA worked very well and the percentage assignment established was more than 90% after full optimization; in some cases 100% assignment was achieved (Tables 1, 2).

The performance of AUTOBA was then tested with actual experimental data on two small well-folded proteins:

ubiquitin (1.6 mM) and M-crystallin (1.4 mM). Ubiquitin was purchased from Cambridge Isotope Laboratories, Inc., ([www.isotope.com](http://www.isotope.com)) as  $^{13}\text{C}/^{15}\text{N}$  uniformly labeled standard sample, 1.6 mM in concentration, in 50 mM sodium phosphate buffer containing 300 mM NaCl at pH 6.5. M-crystallin was a kind gift from Dr. Ravi Pratap Branwal (Barnwal et al. 2006). All the spectra were recorded on an 800 MHz spectrometer equipped with a cryogenically cooled triple-resonance probe. The various experimental parameters used for recording these spectra are listed in Table S1. In each of the two cases, the whole exercise (including experimentation and data analysis) took less than a day.

For identification of self and sequential peaks in 2D-hNcnH spectra, the chemical shifts for different spin

systems from 2D-cbcacoNH spectra are matched with those in 2D-hNcnH spectra. This requires introduction of limits for chemical shift tolerances which are chosen based on the resolution present along respective dimensions e.g. here 0.01 ppm chemical shift tolerance was used along  $^1\text{H}$  and 0.1 ppm was used along  $^{15}\text{N}$  dimensions. Hence, the precision and accuracy of the chemical shifts obtained from the spectral data and the choice of these limits is of utmost importance since errors result in multiple and erroneous assignments. The performance of AUTOBA on these 2D experimental spectra is described below.

### Ubiquitin

In case of Ubiquitin (76 aa), all expected 70 NH-cross peaks were seen in 2D-cbcacoNH spectrum in agreement with the results reported before (BMRB accession number 15410). Among the six missing peaks, three are prolines, one is N-terminal residue, and the other two are Glu24 and Gly53. AUTOBA analysis of variants of 2D-hncNH spectrum yielded amino-acid type identification for 29 spin systems out of expected 33 (5 glycines, 2 alanines, 10 serine/threonine and the residues following them in the sequence). The identification was absent for residues Ile23, Ile36, Gly53, and Thr66. The residues Ile23 and Gly53 were missing because of conformational line broadening, Ile36 was missing as it is followed by Pro37 in the sequence (Kumar and Hosur 2010). Peak list from 2D-(HN)NH-ST spectrum (Kumar et al. 2010a) was used to identify Thr66. Based on this data, AUTOBA run was able to assign 55 out of 70 peaks in Ubiquitin. Further, when the program was fed with peak-lists from 2D-hncoCAnH and 2D-hnCOcanH spectra, the AUTOBA run was able to assign 68 peaks unambiguously. Along with this, AUTOBA also assigned the  $^{13}\text{C}^\alpha$  (65 out of 68, 96%) and  $^{13}\text{C}'$  (64 out of 68, 94%) chemical shifts.

### M-crystallin

Unlike Ubiquitin, M-crystallin (85 aa) is an example of a small protein with repetitive glycines, alanines, and serines/threonines in the sequence (see Fig. S1). As clear from the Figure S1, the sequence is quite rich in glycines, alanines, and serines/threonines (total 27 in number: 10 glycines, 5 alanines, 12 serines/threonines). Moreover, they are well dispersed all along the sequence. Thus, in this case we have used only the primary checkpoints for the assignment process derived from variants of 2D-(HN)NH (Kumar et al. 2010a) spectrum. In the 2D-cbcacoNH spectrum of M-crystallin, 78 NH-cross peaks (or spin-systems) out of expected 80 NH correlations (reported in BMRB accession number 6904) were observed. The AUTOBA analysis of all the three variants of 2D-(HN)NH (Kumar et al. 2010a)

spectrum yielded amino-acid type identification for 23 spin systems out of expected 27. The identification was missing for one glycine (Gly44, overlapping with Phe82) and three alanines (Ala3, Ala4, and Ala21 which do not even show up in the HSQC spectrum). AUTOBA run was able to assign 55 out of 78 peaks in M-Crystallin. Further, when the program was fed with peak-lists from 2D-hncoCAnH and 2D-hnCOcanH spectra, the AUTOBA-2D run was able to assign 64 peaks unambiguously. Along with this, AUTOBA also assigned 62 out of 78  $^{13}\text{C}^\alpha$  chemical shifts (88%) and 62 out of 78  $^{13}\text{C}'$  chemical shifts (87%). The various 2D spectra of M-Crystallin have been shown in Fig. S1: 2D-hNcnH (Fig. S1A), 2D-hncoCAnH (Fig. S1B), and 2D-hnCOcanH (Fig. S1C).

The performance of AUTOBA was then tested with 3D experimental data acquired on ubiquitin (1.6 mM). AUTOBA was able to assign 68 peaks accurately without using the  $^{13}\text{C}$  information from 3D-hncocaNH and 3D-hnCOcaNH peak lists. Finally, when the program was fed with 3D-hncocaNH and 3D-hnCOcaNH peak-lists the complete backbone ( $^1\text{H}^N$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}^\alpha$ , and  $^{13}\text{C}'$ ) assignment was obtained as reported earlier in BMRB accession number 15410.

### Discussion

We have described here an automated strategy to obtain complete assignment of backbone ( $^1\text{H}^N$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}^\alpha$ , and  $^{13}\text{C}'$ ) resonances from HN(C)N suite of experiments [viz. 3D-HN(C)N, 3D-hncoCAnH, 3D-hnCOcaNH, 2D-hNcnH, 2D-hncoCAnH, 2D-hnCOcanH, and variants of 2D-hncNH spectrum] in  $^{15}\text{N}/^{13}\text{C}$  labeled proteins. The success of the automation has been demonstrated on experimental data recorded on two globular proteins: Ubiquitin and M-Crystallin. In each case, complete assignment was established as reported earlier in BMRB accession numbers: 15410 (Ubiquitin, 76 aa) and 6904 ( $\text{Ca}^{2+}$  loaded holo-form of M-crystallin, 85 aa). The performance of the algorithm has also been demonstrated using BMRB data on 27 other proteins.

It is also noteworthy that in some cases like BMRB IDs 16147, 4109 etc., AUTOBA performed better on the simulated 3D data set rather than the 2D data set (Tables 1, 2). This is due to the fact that in a 3D spectrum peak list, the self and cross peaks are readily distinguished. Hence no ambiguities arise when AUTOBA is identifying the self peaks (as described in AUTOBA protocol and working: Step 2) and many of the possible branching in data is outright eliminated. On the other hand, some cases like BMRB IDs 16175, 6902 etc., the performance of AUTOBA is poor on both simulated 2D and 3D datasets. This is the case when the input data has lot of missing or

overlapping peaks. Secondly, although the Tables 1 and 2 list proteins containing number of residues from 58 to 126, there is no correlation between the percentage of accurate assignments obtained from AUOTBA and the number of residues. In fact, AUTOBA reports 97.5% accurate assignments for a well dispersed spectrum of BMRB ID 16260 with 126 residues and a dismal 61.6% for BMRB ID 4105 with 92 residues.

The output given by AUTOBA already accounts for elimination of most of the ambiguously assigned peaks that are correctly assigned in a long stretch. However, in case of overlapping spectra, there are many short stretches as well as a few equally long stretches assigned to two different parts of the protein. In such a scenario, it becomes impossible for AUTOBA to pin down a unique assignment and it outputs all the alternative assignments as ambiguous. For example, consider the output given in the supplementary material for BMRB ID 16175. None of the ambiguously assigned peaks for Residues 5 to 10 occur in the correctly assigned long stretch.

AUTOBA has been trained on the simulated BMRB data to minimize wrong assignments. In the BMRB data analysis (Tables 1, 2) the wrong assignments reported are generally for residues whose true peaks are missing in the spectrum (we could identify these assignments as wrong since we already know the assignments here). For example, in case of BMRB ID 16147, both wrong assignments in Table 1 are for residues not present in the spectrum. The genuinely ‘correct’ residues for those peaks are reported as ‘unassigned’. Hence, in a new protein, after all the internal checks for consistency, wrong assignments would result in “unassigned” residues in the AUTOBA output. Presence of lot of unassigned residues even with absence of any ambiguous assignments indicates speculative assignments. Thus the ratio of accurate/(ambiguous + unassigned) residues is a good indicator of the AUTOBA performance on a particular protein spectrum. A higher ratio means that the AUTOBA assignment is quite good whereas a lower ratio means that the user should be cautious of the results.

Another problem worth cautioning is a possibility of peak cancellation in the 2D-hNcnH spectrum when two conditions, namely, (a) the  $^1\text{H}^{\text{N}}$  chemical shift of residue  $j$  is identical to that of residue  $i$  and (b)  $^{15}\text{N}$  chemical shift of  $j$  is identical to that of  $i + 1$ , are simultaneously satisfied. But, considering that the amide proton and  $^{15}\text{N}$  dispersions are usually very good in folded proteins, the chance of both the criteria getting satisfied at the same time can be expected to be small, and will be even lower at higher magnetic fields. Secondly, as mentioned earlier, in the variants of 2D-hncNH spectrum if two consecutive residues  $i$  and  $i + 1$  are of the same type (GG in hncNH-G or (G/A)(G/A) in hncNH-A or (S/T)(S/T) in hncNH-ST) then the peak for  $i + 1$  residue will be missing in the

spectrum. In such a situation one has to acquire variants of 2D-(HN)NH experiment (Kumar et al. 2010a) which do not show this problem.

The success of the automation is dictated by proper amino acid type identification. Any mistake in assigning the amino acid type will lead to wrong assignment. The results presented here have been obtained using glycines, alanines, and serines/threonines checkpoints only. Compared to other amino acids, these four have very characteristic  $^{13}\text{C}^{\alpha}$  (for glycines) and  $^{13}\text{C}^{\beta}$  (for alanines and serines/threonines) chemical shifts (Chugh et al. 2008), which help to minimize the mistakes. In addition, AUTOBA strategy also makes use of the residues following glycines, alanines, and serines/threonines in the sequence as secondary checkpoints (identified from variants of 2D-hncNH spectrum, Fig. 1). On an average, the percentage composition of these four residues (Gly, Ala, and Ser/Thr) put together in any given protein is as much as  $25.8 \pm 5.8\%$  (Barnwal et al. 2008). Thus we believe that AUTOBA should accomplish successful sequence specific resonance assignment in most cases. We may mention here that alternative strategies for identification of Ala, Gly, Ser/Thr residues have also been reported before (Atreya et al. 2000; Friedrichs et al. 1994; Grzesiek and Bax 1993; Lukin et al. 1997) and have been used in automated assignment algorithms (Atreya et al. 2000; Lescop and Brutscher 2009). A number of other methods are also available for identifying amino-acid types directly on the HSQC spectrum (Dotsch et al. 1996; Grzesiek and Bax 1993; Pantoja-Uceda and Santoro 2008; Lescop et al. 2008; Kumar et al. 2010a; Schubert et al. 2001b; Schubert et al. 2001a; Chugh and Hosur 2008). AUTOBA can indeed take these inputs as well.

(Lescop and Brutscher 2009) recently introduced the so-called «BATCH» strategy: a fast and highly automated backbone resonance assignment protocol based on 2D spectra (Lescop and Brutscher 2009). In this strategy, amino-acid-type identification is obtained from a HADAMAC experiment (Lescop et al. 2008) that can be recorded in  $\sim 30$  min, which provides a significantly higher amino-acid-type discrimination than the usual statistical analysis of  $^{13}\text{C}$  chemical shifts. The MUSIC (multiplicity selective in-phase coherence transfer) experiments (Schubert et al. 2001a; Schubert et al. 2001b) also enables amino acid type identification. Indeed, AUTOBA can also make use of the checkpoints derived from such other NMR methods available for amino-acid type identification (Dotsch et al. 1996; Grzesiek and Bax 1993; Pantoja-Uceda and Santoro 2008; Lescop et al. 2008; Kumar et al. 2010a; Chugh and Hosur 2008; Schubert et al. 2001a; Schubert et al. 2001b). Combining the different approaches: (i) amino-acid-type identification from MUSIC/HADAMAC or other experiments (ii) amide sequential correlations from 2D-hNcnH experiment (Kumar and Hosur 2010), and (iii) backbone

$^{13}\text{C}$  ( $^{13}\text{C}^{\alpha}$  and  $^{13}\text{C}'$ ) correlations from 2D-hncoCANH, and 2D-hnCOcanH experiments (Kumar et al. 2010b), the speed of backbone assignment can be enhanced further. Further, the experiments used here (2D-hNcnH, 2D-hncNH, 2D-hncoCANH, and 2D-hnCOcanH) are fully compatible with other fast NMR methods like BEST NMR (Schanda et al. 2006; Lescop et al. 2007) and non-linear sampling (Rovnyak et al. 2004). Thus, the total experiment time can be reduced further which will be useful in cases when the given protein sample cannot be concentrated beyond a certain value or when the protein under investigation is unstable in solution.

The pulse sequences HN(C)N, hncoCANH, and hnCOcanH and the software-program AUTOBA described here will be available upon request or can be downloaded from this link <http://www.tifr.res.in/~hosur/download.htm>. The AUTOBA server can be accessed freely from here: <http://www.tifr.res.in/~hosur/AUTOBA.html>.

**Acknowledgments** We thank the Government of India for providing financial support to the National Facility for High Field NMR at Tata Institute of Fundamental Research, India.

## References

- Atreya HS, Szyperski T (2005) Rapid NMR data collection. *Meth Enzymol* 394:78–108
- Atreya HS, Sahu SC, Chary KV, Govil G (2000) A tracked approach for automated NMR assignments in proteins (TATAPRO). *J Biomol NMR* 17:125–136
- Barnwal RP, Jobby MK, Sharma Y, Chary KV (2006) NMR assignment of M-crystallin: a novel  $\text{Ca}^{2+}$  binding protein of the  $\beta\gamma$ -crystallin superfamily from *Methanosarcina acetivorans*. *J Biomol NMR* 36:32
- Barnwal RP, Rout AK, Atreya HS, Chary KV (2008) Identification of C-terminal neighbours of amino acid residues without an aliphatic  $^{13}\text{C}$ gamma as an aid to NMR assignments in proteins. *J Biomol NMR* 41:191–197
- Chatterjee A, Bhavesh NS, Panchal SC, Hosur RV (2002) A novel protocol based on HN(C)N for rapid resonance assignment in ( $^{15}\text{N}$ ,  $^{13}\text{C}$ ) labeled proteins: implications to structural genomics. *Biochem Biophys Res Commun* 293:427–432
- Chatterjee A, Kumar A, Hosur RV (2006) Alanine checkpoints in HNN and HN(C)N spectra. *J Magn Reson* 181:21–28
- Chugh J, Hosur RV (2008) Spectroscopic labeling of A, S/T in the  $^1\text{H}$ – $^{15}\text{N}$  HSQC spectrum of uniformly ( $^{15}\text{N}$ – $^{13}\text{C}$ ) labeled proteins. *J Magn Reson* 194:289–294
- Chugh J, Kumar D, Hosur RV (2008) Tuning the HNN experiment: generation of serine-threonine checkpoints. *J Biomol NMR* 40:145–152
- Coggins BE, Zhou P (2003) PACES: protein sequential assignment by computer-assisted exhaustive search. *J Biomol NMR* 26:93–111
- Crippen GM, Rousaki A, Revington M, Zhang Y, Zuiderweg ER (2010) SAGA: rapid automatic mainchain NMR assignment for large proteins. *J Biomol NMR* 46:281–298
- Dotsch V, Matsuo H, Wagner G (1996) Amino-acid-type identification for deuterated proteins with a b-carbon-edited HNCOCACB experiment. *J Magn Reson B* 112:95–100
- Felli IC, Brutscher B (2009) Recent advances in solution NMR: fast methods and heteronuclear direct detection. *Chemphyschem* 10:1356–1368
- Fiorito F, Hiller S, Wider G, Wuthrich K (2006) Automated resonance assignment of proteins: 6D APSY-NMR. *J Biomol NMR* 35:27–37
- Freeman R, Kupce E (2003) New methods for fast multidimensional NMR. *J Biomol NMR* 27:101–113
- Friedrichs MS, Mueller L, Wittekind M (1994) An automated procedure for the assignment of protein  $^1\text{HN}$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}$  alpha,  $^1\text{H}$  alpha,  $^{13}\text{C}$  beta and  $^1\text{H}$  beta resonances. *J Biomol NMR* 4:703–726
- Frydman L, Scherf T, Lupulescu A (2002) The acquisition of multidimensional NMR spectra within a single scan. *Proc Natl Acad Sci USA* 99:15858–15862
- Grzesiek S, Bax A (1992) Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J Am Chem Soc* 114:6293
- Grzesiek S, Bax A (1993) Amino acid type determination in the sequential assignment procedure of uniformly  $^{13}\text{C}/^{15}\text{N}$ -enriched proteins. *J Biomol NMR* 3:185–204
- Hiller S, Fiorito F, Wuthrich K, Wider G (2005) Automated projection spectroscopy (APSY). *Proc Natl Acad Sci USA* 102:10876–10881
- Jaravine VA, Zhuravleva AV, Permi P, Ibragimov I, Orekhov VY (2008) Hyperdimensional NMR spectroscopy with nonlinear sampling. *J Am Chem Soc* 130:3927–3936
- Jung YS, Zweckstetter M (2004) Mars-robust automatic backbone assignment of proteins. *J Biomol NMR* 30:11–23
- Keller RLJ (2004) Optimizing the process of nuclear magnetic resonance spectrum analysis and computer aided resonance assignment. Thèse de doctorat, ETH Zurich Thesis No 15947, Switzerland
- Kim S, Szyperski T (2003) GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information. *J Am Chem Soc* 125:1385–1393
- Kumar D, Hosur RV (2010) An efficient high-throughput protocol based on 2D-HN(C)N for unambiguous HN and  $^{15}\text{N}$  backbone assignment in small folded proteins in less than a day. *Res Commun: Curr Sci* 99:1581–1586
- Kumar D, Chugh J, Hosur RV (2009) Generation of serine/threonine checkpoints in HN(C)N spectra. *J Chem Sci* 121:955–964
- Kumar D, Paul S, Hosur RV (2010a) BEST-HNN and 2D-(HN)NH experiments for rapid backbone assignment in proteins. *J Magn Reson* 204:111–117
- Kumar D, Reddy JG, Hosur RV (2010b) hnCOcanH and hncoCANH pulse sequences for rapid and unambiguous backbone assignment in ( $^{13}\text{C}$ ,  $^{15}\text{N}$ ) labeled proteins. *J Magn Reson* 206:134–138
- Kupce E, Freeman R (2003a) Fast multi-dimensional Hadamard spectroscopy. *J Magn Reson* 163:56–63
- Kupce E, Freeman R (2003b) Fast multi-dimensional NMR of proteins. *J Biomol NMR* 25:349–354
- Kupce E, Freeman R (2003c) Projection-reconstruction of three-dimensional NMR spectra. *J Am Chem Soc* 125:13958–13959
- Lescop E, Brutscher B (2007) Hyperdimensional protein NMR spectroscopy in peptide-sequence space. *J Am Chem Soc* 129:11916–11917
- Lescop E, Brutscher B (2009) Highly automated protein backbone resonance assignment within a few hours: the “BATCH” strategy and software package. *J Biomol NMR* 44:43–57
- Lescop E, Schanda P, Brutscher B (2007) A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J Magn Reson* 187:163–169
- Lescop E, Rasia R, Brutscher B (2008) Hadamard amino-acid-type edited NMR experiment for fast protein resonance assignment. *J Am Chem Soc* 130:5014–5015

- Lukin JA, Gove AP, Talukdar SN, Ho C (1997) Automated probabilistic method for assigning backbone resonances of ( $^{13}\text{C}$ ,  $^{15}\text{N}$ )-labeled proteins. *J Biomol NMR* 9:151–166
- Moseley HN, Sahota G, Montelione GT (2004) Assignment validation software suite for the evaluation and presentation of protein resonance assignment data. *J Biomol NMR* 28:341–355
- Panchal SC, Bhavesh NS, Hosur RV (2001) Improved 3D triple resonance experiments, HNN and HN(C)N, for HN and  $^{15}\text{N}$  sequential correlations in ( $^{13}\text{C}$ ,  $^{15}\text{N}$ ) labeled proteins: application to unfolded proteins. *J Biomol NMR* 20:135–147
- Pantoja-Uceda D, Santoro J (2008) Amino acid type identification in NMR spectra of proteins via beta- and gamma-carbon edited experiments. *J Magn Reson* 195:187–195
- Rovnyak D, Frueh DP, Sastry M, Sun ZY, Stern AS, Hoch JC, Wagner G (2004) Accelerated acquisition of high resolution triple-resonance spectra using non-uniform sampling and maximum entropy reconstruction. *J Magn Reson* 170:15–21
- Schanda P, Van MH, Brutscher B (2006) Speeding up three-dimensional protein NMR experiments to a few minutes. *J Am Chem Soc* 128:9042–9043
- Schubert M, Oschkinat H, Schmieder P (2001a) MUSIC and aromatic residues: amino acid type-selective ( $^1\text{H}$ –( $^{15}\text{N}$ ) correlations, III. *J Magn Reson* 153:186–192
- Schubert M, Oschkinat H, Schmieder P (2001b) MUSIC, selective pulses, and tuned delays: amino acid type-selective ( $^1\text{H}$ –( $^{15}\text{N}$ ) correlations, II. *J Magn Reson* 148:61–72