Available online at www.sciencedirect.com
ScienceDirect
Journal of Magnetic Resonance 189 (2007) 173–181
JMR Journal of Magnetic Resonance
www.elsevier.com/locate/jmr

Protein assignments without peak lists using higher-order spectra ☆

Gregory Benison *, Donald S. Berkholz, Elisar Barbar

Department of Biochemistry and Biophysics, Oregon State University, Corvallis, OR 97331, USA

Received 22 June 2007; revised 5 September 2007
Available online 20 September 2007

Abstract

Despite advances in automating the generation and manipulation of peak lists for assigning biomolecules, there are well-known advantages to working directly with spectra: the eye is still superior to computer algorithms when it comes to picking out peak relationships from contour plots in the presence of confounding factors such as noise, overlap, and spectral artifacts. Here, we present constructs called higher-order spectra for identifying, through direct visual examination, many of the same relationships typically identified by searching peak lists, making them another addition to the set of tools (alongside peak picking and automated assignment) that can be used to solve the assignment problem. The technique is useful for searching for correlated peaks in any spectrum type. Application of this technique to novel, complete sequential assignment of two proteins (AhpFn and $IC74^{84–143}$) is demonstrated. The program "burrow-owl" for the generation and display of higher-order spectra is available at http://sourceforge.net/projects/burrow-owl or from the authors.

Keywords: Sequential assignment; Higher-order spectra; Proteins; Burrow-owl; Peak lists

1. Introduction

The starting point for all biomolecular NMR studies is the assignment problem: the mapping of specific atoms in a molecular structure to specific peaks in spectra. For small molecules, the chemical shift or multiplet structure of a peak is often sufficient for assignment. In protein work, however, multiplet structure is usually not observed, and accidental chemical shift degeneracy or near degeneracy are prevalent: there are usually many candidates for any particular assignment based on chemical shift alone. For this reason, assignment of NMR spectra is almost entirely based on corroboration via multiple peaks. Typically, the same atom participates in multiple (ambiguous) relationships due to its participation in multiple magnetization transfer pathways (either in the same experiment, or across several experiments). For example, it is common practice to search the CBCA(CO)NH spectrum for pairs of peaks sharing common H and N chemical shifts but with different C chemical shifts (one at C^{α} and one at C^{β}). If the C^{α} and C^{β} assignments are known, the pair of peaks can then be assigned, and hence the H and N assignments obtained. Often assigning either the C^{α} or the C^{β} peak individually is not possible due to the degeneracy of C^{α} and C^{β} chemical shifts, but assigning the pair of peaks is possible because of unique (C^{α}, C^{β}) combinations.

Making assignments through corroboration always involves finding peaks that share a common chemical shift in one or more dimensions. In two-dimensional spectra, identifying such groups of peaks can be accomplished effectively through visual examination of the spectrum as a contour plot. However, biomolecular studies usually demand three- or four-dimensional spectra to achieve sufficient resolution. Only small portions of these spectra can be visualized at once, which can make it impossible to identify groups of peaks through visual inspection alone. For this reason it has become standard practice to abstract such spectra into peak lists and then to search for correlations within those peak lists rather than within the spectra themselves (Fig. 1).

The search for correlations within peak lists is amenable to automation, and several programs have been developed

☆ This work was supported by NSF CAREER Grant MCB-0417181 to E.J.B. and an American Heart Association Fellowship (Award #0720019Z) to G.C.B.
* Corresponding author. Fax: +1 541 737 0481.
E-mail address: gbenison@gmail.com (G. Benison).

1090-7807/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2007.09.009

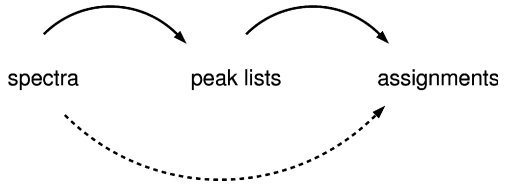


Fig. 1. The assignment process, with and without using peak lists.

for this purpose. AUTOASSIGN can rapidly obtain complete assignments for a protein backbone using standard NMR spectra, given peak lists of sufficiently high quality [1]. Another more recent example in the same genre is AUTOLINK [2]. The main limitation of these automated assignment protocols is that the quality of their output can only be as good as the quality of the peak lists that serve as their input. Peak lists can be generated by looking at sections of spectra as contour plots and picking peaks, but there have also been efforts to automate this time-consuming process. APART is a program that filters and refines peak lists based on expected relationships between peaks [3]. AUTOPSY is a program that abstracts peak information from spectra, taking advantage of the fact that peaks which are related by magnetization transfer pathways also have similar lineshapes [4].

Despite advances in automating peak picking, working directly with spectra rather than with peak lists has well-known advantages. In particular, the eye remains superior to computer algorithms for picking out peak relationships from contour plots in the presence of confounding factors such as noise, overlap, and spectral artifacts. The quality of peak lists (and in turn, the performance of automated assignment) generally improves with the amount of time spent manually grooming them. In practice, this task accounts for a large portion of the effort involved in modern biomolecular NMR, which negates somewhat the effectiveness of automated assignment.

In this paper, we describe constructs called higher-order spectra that make it possible to use contour plots to directly visualize relationships that otherwise could only be found by searching peak lists. Using higher-order spectra, one can uncover the many correlations present in common three- and four-dimensional biomolecular NMR experiments without sacrificing the advantages of direct visual examination. We describe an approach to sequential assignment that proceeds directly from spectra to assignments without an intervening peak-picking step, and we illustrate its application in obtaining complete backbone assignments of two proteins. Because it relies only on standard NMR experiments, our direct visualization approach based on higher-order spectra can be used independently or as a complement to approaches based on peak picking.

2. Theory

Our analysis is based on a set of simple primitive operations that take as arguments one or more spectra and return as a result another spectrum, where a spectrum is defined simply as an n-dimensional matrix of real numbers, with a reference frequency and digital resolution defined for each dimension. For simplicity, we do not consider the imaginary components of NMR data, because all of our examples and many other common cases require only the real component. The primitive operations have the important property that the output of one can serve as the input for any other. This allows them to be combined in any way and any number of times, making it easy to construct arbitrarily complex transformations. The five primitive operations are:

- *Transposition*. The dimensions are put in a new order by exchanging the first dimension with any later dimension, leaving the overall dimensionality unchanged. Through multiple applications of the transposition operator, the dimensions can be put into any order.

$$S(i_1, \ldots i_k \ldots) = S'(i_k, \ldots i_1 \ldots)$$

- *Projection*. A slice is taken perpendicular to the first dimension of the spectrum, decreasing the overall dimensionality by one.

$$S(i_2, \ldots) = S'(i_1^j, i_2, \ldots)$$

- *Convolution*. Two spectra are combined to form a higher-order spectrum by multiplying their values together in a manner similar to the direct product as defined in linear algebra. The dimensionality of the resulting spectrum is equal to the sum of the dimensionalities of the two spectra being convoluted.

$$S(i_1', i_2', \ldots, i_1'', i_2'', \ldots) = S'(i_1', i_2', \ldots) * S''(i_1'', i_2'', \ldots)$$

- *Diagonal projection*. Similarly to ordinary projection, only one point from the first dimension is retained, except rather than using a constant index to select this point, the index i_1^* is adjusted to retain the point nearest in chemical shift to the corresponding point in dimension 2. The result is a slice along a diagonal of the spectrum. The dimensionality of the spectrum is decreased by one. The operation works even when the first and second dimensions have different digital resolutions, as long as they have some overlapping chemical shift range.

$$S(i_2, \ldots) = S'(i_1^*, i_2, \ldots)$$

- *Integration*. All points in the first dimension are summed, resulting in a spectrum of one fewer dimensions.

$$S(i_2, \ldots) = \sum_{i1} S'(i_1, i_2, \ldots)$$

3. Implementation

The source code for the program burrow-owl is freely available. It currently accepts as input any spectrum in the NMRPipe [5] format, but it could be extended to accept

other formats. The primitive operations defined above can be applied in any order, any number of times, using any number of raw spectra as input. This can result in spectra of such high dimensionality that they would require an unreasonable amount of storage if calculated in their entirety. Instead, the program calculates planes and other subsections on the fly as they are viewed, trading processing power for storage. Modern desktop computers are fast enough to make this process transparent for the examples cited in this paper. A visual interface based on the GTK+ widget set (http://www.gtk.org) provides contour plots and basic annotations such as cursors and labels. Spectra and annotations can be made responsive to mouse and keyboard input, allowing for the creation of interactive visual interfaces. Both the calculation of higher-order spectra and the construction of visual displays are expressed in the Scheme programming language, which is a modern Lisp dialect [6]. Scheme, like other popular modern languages such as TCL and Perl, has many properties (particularly automatic memory management) that make it easy to learn and well-suited to rapid development. The list-handling capability of Scheme is also naturally suited to the management of lists of chemical shift assignments.

No effort has been made to provide a single all-purpose user interface; rather the intent is to make it easy to construct task-specific interfaces for any spectrum type. Sample scripts are provided for the cases presented in this study, such as sequential assignment via CBCA(CO)NH and HNCACB spectra.

4. Results

4.1. Example in one dimension

We will illustrate the higher-order spectrum approach with a one-dimensional assignment problem (Fig. 2) that is very simple but conceptually similar to more difficult problems of higher-dimensionality. In spectrum 2d, which is the point-by-point multiplication of spectra 2a and b, a common peak at 8 ppm is clearly identified, without any need to first abstract spectra 2a and b into peak lists. Since there are few peaks and few dimensions in this example, the shared peak is obvious from visual comparison alone, but that is not the case in more difficult examples.

The one-dimensional point-by-point multiplication operation can be implemented in terms of our primitives as a convolution operation (corresponding to the transformation 2a,b → 2c) followed by diagonal projection (2c → 2d). We note that it is sometimes simpler to think in terms of such point-by-point multiplication operations than in terms of our five primitives of transposition, projection, convolution, diagonal projection, and integration. However, there are useful applications of our primitives that cannot be expressed as point-by-point multiplications. Because point-by-point multiplication can be defined so easily in terms of our primitives, our choice does not preclude thinking (and coding) in terms of point-by-point multiplication when desired.

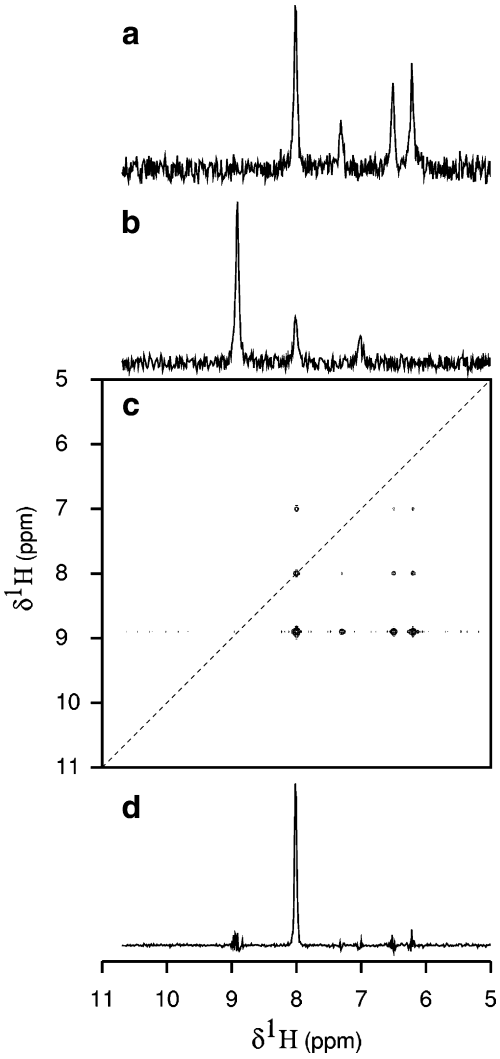


Fig. 2. Using convolution and projection to find common peaks in one-dimensional spectra. (a and b) Hypothetical 1D NMR spectra. (c) Convolution of (a) and (b). (d) Diagonal projection of (c) (also shown as a dashed line in (c)).

4.2. Relationship to indirect covariance spectroscopy

Indirect covariance spectroscopy is a way of viewing proton-detected spectra as carbon–carbon correlations, which are often more easily interpreted [7]. Application of a sequence of three higher-order spectrum operators—convolution, diagonal projection, and integration—yields a transformation equivalent to the covariance operation as originally defined, with the absence of the final square-root step (which does not change the positions of the peaks). To demonstrate the equivalence of the two approaches, we calculated a ^{13}C–^{13}C correlation spectrum (Fig. 3b) of a mixture of proline and leucine at natural abundance from a ^{1}H–^{13}C HSQC-TOCSY spectrum (Fig. 3a) using the steps in Fig. 3c.

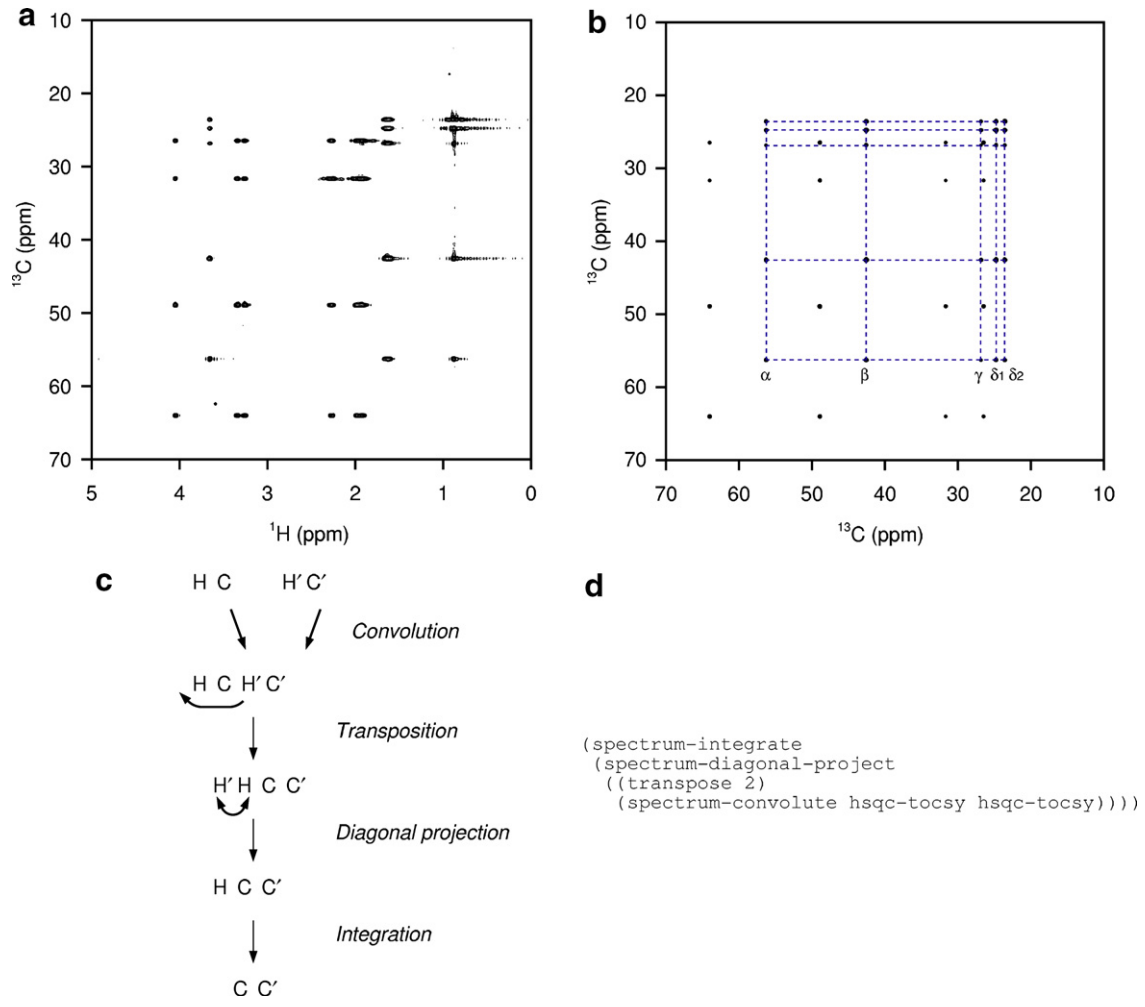


Fig. 3. Indirect covariance NMR spectroscopy implemented using the higher-order spectrum operators. (a) ^{1}H/^{13}C HSQC-TOCSY spectrum of a mixture of 100 mM proline and 100 mM leucine at natural abundance in water. (b) ^{13}C–^{13}C correlation experiment calculated from (a) with the transformation shown in (c), as implemented by the Scheme program shown in (d). The leucine spin system is highlighted by dashed lines.

4.3. Finding correlations in the CBCA(CO)NH experiment

The CBCA(CO)NH experiment, a workhorse for assigning protein backbones, contains signals at $(H_i, N_i, C^{\alpha}_{(i-1)})$ and $(H_i, N_i, C^{\beta}_{(i-1)})$ [8]. It is commonly used to assign H_i and N_i given assignments for residue $i-1$. This can be accomplished by abstracting the spectrum as a peak list, grouping the peaks into pairs with matching H_i and N_i chemical shifts, and then searching for a pair with the correct $C^{\alpha}_{(i-1)}$ and $C^{\beta}_{(i-1)}$ values. The same result can be achieved by searching for a single peak within an appropriate higher-order spectrum, constructed using the sequence of steps shown in Fig. 4. First, two copies of the CBCA(CO)NH spectrum are convoluted to yield a six-dimensional spectrum with two proton, two carbon, and two nitrogen dimensions. This spectrum contains a cross peak at $(H, N, C^{\alpha}_{(i-1)}, H, N, C^{\beta}_{(i-1)})$. Since this cross peak has an identical chemical shift (H) in the first and fourth dimensions, we use diagonal projection to eliminate one of these dimensions while retaining the peak of interest. Using the same reasoning, we eliminate one of the nitrogen dimensions, resulting in a four-dimensional (H, N, C_I, C_{II}) spectrum. By successively applying to this four-dimensional spectrum two projection operators, one at $C^{\alpha}_{(i-1)}$ and one at $C^{\beta}_{(i-1)}$, we obtain an easily visualized plane containing a peak at (H_i, N_i); i.e. the sequential assignment is obtained, making simultaneous use of C^{α} and C^{β} correlation, without the need to abstract the spectrum as a peak list.

In Fig. 5, we illustrate the technique using a CBCA(CO)NH spectrum of AhpFn,[1] the 202 residue N-terminal domain of the protein AhpF from the *Salmonella typhimurium* alkyl hydroperoxide reductase system [9]. The HSQC (spectrum 5a) shows all assignable amide resonances. A slice taken through the CBCA(CO)NH spectrum at T164 C^{β} (spectrum 5b) identifies (H, N) assignments for E165 consistent with $C^{\beta}_{(i-1)}$; similarly, a slice taken at T164 C^{α} (spectrum 5c) identifies assignments consistent with $C^{\alpha}_{(i-1)}$. Spectrum 5d is a two-dimensional projection at $C^{\alpha}_{(i-1)}$ and $C^{\beta}_{(i-1)}$ of the four-dimensional (H, N, C_I, C_{II}) spectrum,

[1] *Abbreviations used:* AhpFn, alkyl hydroperoxide reductase F N-terminal domain; IC74, cytoplasmic dynein intermediate chain.

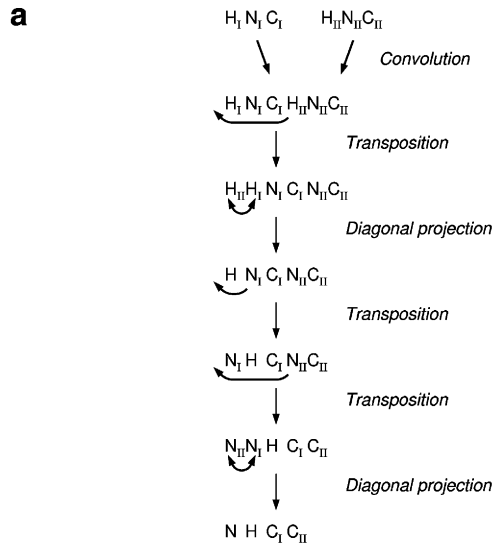


Fig. 4. (a) Steps for creating the 4D (H, N, C_I, C_{II}) spectrum from the 3D CBCA(CO)NH spectrum using convolution, transposition, and diagonal projection. Arrows show the rearrangement of dimensions for transposition or diagonal projection. (b) Corresponding scheme code to implement the transformation.

showing assignments consistent with both $C^{\alpha}_{(i-1)}$ and $C^{\beta}_{(i-1)}$. We note that spectrum 5d is equivalent to the point-by-point multiplication of spectrum 5b and spectrum 5c, and there is no reason it could not be constructed that way. However, we find the four-dimensional construct more convenient for creating interactive interfaces. Also, it is needed as the starting point for amino acid type filtering, described in the next section.

The purpose of the four-dimensional higher-order construct is to help search for correlations, not to remove ambiguities present in the underlying three-dimensional spectrum. For example, Fig. 5d contains four discernable peaks, rather than just the one true assignment, because of accidental chemical shift degeneracy of carbon chemical shifts. Such ambiguities are inherent in the CBCA(CO)NH spectrum and must be resolved by other means, such as corroboration by other spectra; this is true no matter which technique is being used for assignment.

4.4. Filtering by amino acid type

In addition to its utility in making sequential matches, the CBCA(CO)NH spectrum contains valuable information about amino acid type. Some residue types (such as Ala, Thr, and Ser) usually can be identified unambiguously on the basis of C^{α} and C^{β} chemical shifts alone. For other amino acid types, the identity can often be narrowed to a few candidates out of the 20 standard amino acids.

Just as it is common practice to use peak lists to make sequential matches, it is common practice to use peak lists to identify residue types: a pair of peaks (one for C^{α}, one for C^{β}) is assigned to an amino acid type by comparing the carbon chemical shifts to a database of standard chemical shift distributions by amino acid type (available from the BioMagResBank at http://www.bmrb.wisc.edu/ref_info/statsel.htm).

Here, we will show how information about amino acid type can be extracted from the CBCA(CO)NH experiment without the need for any peak picking through the construction of an appropriate higher-order spectrum. The full sequence of steps is shown in Fig. 6 and can be summarized as: (1) Create a simulated 1D spectrum containing a single Gaussian peak with the mean and standard deviation set to the distribution for C^{α} atoms of the residue type. (2) Use convolution and projection to multiply the simulated 1D spectrum against one of the carbon dimensions of the 4D (H, N, C_I, C_{II}) spectrum. (3) Integrate the carbon dimension that was just convoluted. (4) Repeat steps 1–3 with the other carbon dimension using the distribution of C^{β} shifts for the residue type.

The resulting 2D (N, H) plane retains only peaks derived from CBCA(CO)NH signals with carbon chemical shifts that fall within the distribution defined by the residue type. Because the CBCA(CO)NH spectrum correlates amide chemical shifts with the carbon atoms of the preceding residue, the remaining peaks in the 2D plane correspond to residues immediately following the amino acid type being

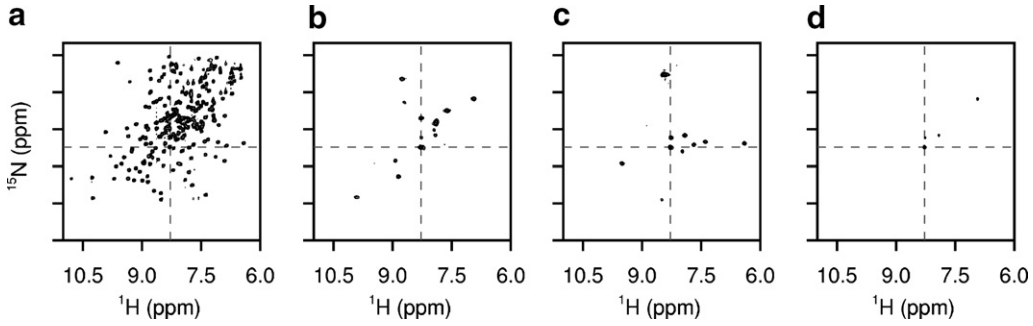
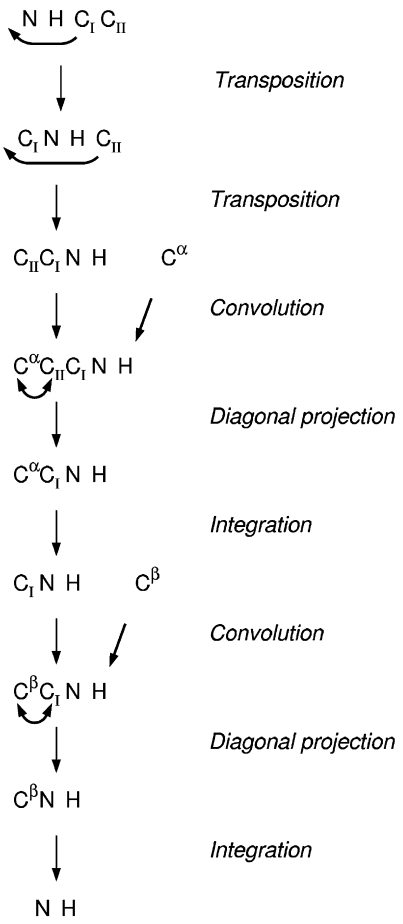


Fig. 5. (a) ^{1}H, ^{15}N HSQC spectrum of AhpFn. (b) Section of the CBCA(CO)NH spectrum of AhpFn taken at T164 C^{β} (68.9 ppm). (c) Section of the CBCA(CO)NH spectrum taken at T164 C^{α} (65.7 ppm). (d) Point-by-point product of (b) and (c), or equivalently a (^{1}H, ^{15}N) section of the 4D (H, N, C_I, C_{II}) higher-order spectrum taken at T164 C^{α} and T164 C^{β}. Dashed lines are drawn at the chemical shifts of E165 H and E165 N.

a



b

```
(define cbcaconh-aa-filtered
  (spectrum-integrate
   (spectrum-diagonal-project
    (spectrum-convolute
     cb-profile
     (spectrum-integrate
      (spectrum-diagonal-project
       (spectrum-convolute
        ca-profile
        ((transpose 3)
         ((transpose 2) cbcaconh-4d)))))))))
```

Fig. 6. (a) Steps for the creation of amino acid-type-filtered NH sections from the CBCA(CO)NH spectrum. The starting point is the 4D (H, N, C_I, C_{II}) spectrum from Fig. 4. C^α and C^β are simulated one-dimensional spectra containing a single peak with a Gaussian profile representing the distribution of chemical shifts for that amino acid type. (b) Scheme code to implement the transformation.

filtered for. Examples for Ala and Phe are shown in Fig. 7. The Ala-filtered spectrum identifies residues following Ala with near-perfect selectivity, whereas the selectivity of the Phe-filtered spectrum is somewhat lower because of the significant overlap between the chemical shift distributions of Phe, Tyr, Asn, and Asp.

4.5. Application to backbone assignments of AhpFn and IC74$^{84–143}$

We developed the following interactive protocol for sequential assignment of proteins based on triple resonance experiments. From an HNCACB strip plot of residue i, the user marks the C_i^α and C_i^β positions with cursors, referring to the corresponding CBCA(CO)NH strip to avoid mistakenly marking a $C_{(i-1)}^\alpha$ or $C_{(i-1)}^\beta$ peak. The program then takes a section at C_i^α and C_i^β through the (H, N, C_I, C_{II}) derivative of the CBCA(CO)NH spectrum, resulting in an (H, N) plane. The most intense peak in this plane is often the correct (H_{i+1}, N_{i+1}) assignment, even for a protein the size of AhpFn (202 residues). When multiple intense peaks are present, the ambiguity can be resolved through corroboration by other spectra: strip plots for residue i and the candidate $i+1$ assignment are presented side by side for any available amide-correlated experiments (CBCA(CO)NH, ^{15}N-edited NOESY, CCC-NH TOCSY [10], HNHA, etc.) All of the candidate assignments can be considered quickly in this way, because to investigate each one it is only necessary to move a cursor onto a peak in the (H, N) plane. The net output is an assignment for residue $i+1$ based on simultaneous C^α and C^β correlation, and the only input is the i assignment and the raw spectra—there is no need to abstract the 3D experiments into peak lists. By reversing the roles of the HNCACB and CBCA(CO)NH spectra, one can extend assignments in the opposite direction, from i to $i-1$.

We applied our approach successfully to backbone assignment of two proteins: AhpFn, and IC74$^{84–143}$, a 61-residue fragment of the dynein intermediate chain with a 17-residue His tag [12] (Fig. 8). Assigning these two proteins is of comparable difficulty because IC74$^{84–143}$, though smaller, is natively unfolded and thus has poor chemical shift dispersion of the amide protons, making the degree of accidental degeneracy and overlap similar to that seen in AhpFn. For AhpFn, we began with an initial set of assignments determined using Autoassign based on the HNCACB, CBCA(CO)NH, HNCA, HN(CO)CA, and HNCO experiments. We also recorded a 3D CCC-NH TOCSY experiment, which can be very helpful as input to automated assignment procedures [11]; however, as is not unusual for a protein of this size, it was not of sufficiently high quality to be useful. Autoassign generated assignments for 141 residues (73% of the sequence). To verify these assignments and to fill in gaps, we then used the visual interface implemented in burrow-owl described above, making use of the HNCACB, CBCA(CO)NH, and ^{15}N-edited NOESY spectra. We were able to fill in all of the missing assignments and to determine that 26 of the assignments generated by Autoassign were incorrect, mostly due to misplacement of short stretches lacking residues with distinctive chemical shifts (such as Thr or Ala) and occasionally due to the presence of spurious peaks in the input list from noise or artifacts. For IC74$^{84–143}$, we completed the sequential assignments from scratch using only the visual interface implemented in burrow-owl and the HNCACB, CBCA(CO)NH, and ^{15}N-edited NOESY spectra. We began by identifying short segments with distinctive amino acid types and then extended those segments using the techniques described above. No automated assignment program was used and no peak lists were gen-

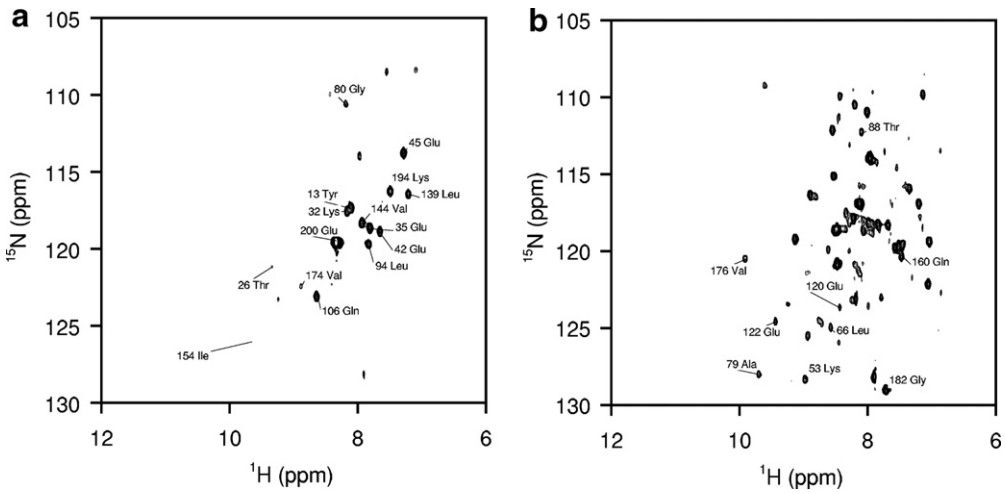


Fig. 7. Amino acid-type-filtered (H,N) sections derived from the CBCA(CO)NH spectrum of AhpFn according to the steps shown in Fig. 6. (a) Ala-filtered spectrum, with residues after Ala labeled; (b) Phe-filtered spectrum, with residues after Phe labeled.

erated. Complete backbone assignments for AhpFn and IC74$^{84–143}$ have been deposited in the BMRB under accession numbers 15264 and 15441, respectively.

5. Methods

5.1. Sample preparation

A ^{15}N-, ^{13}C-labeled sample of AhpFn was prepared as described previously [9] except full labeling was achieved by using ^{15}N-labeled ammonium chloride and ^{13}C-labeled glucose as the sole nitrogen and carbon sources, respectively, during growth. The sample was at pH 6.5 and contained 1 mM protein, 50 mM potassium chloride, 50 mM potassium phosphate, and 10% $^{2}H_2O$. A ^{15}N, ^{13}C-labeled sample of IC74$^{84–143}$ was prepared as described previously [12]. The sample was at pH 5.5 and contained 1 mM protein, 50 mM sodium phosphate, 50 mM sodium citrate, 50 mM sodium chloride, 10% $^{2}H_2O$, and a protease inhibitor cocktail (Roche).

5.2. NMR data collection

^{1}H–^{15}N HSQC, HNCACB, CBCA(CO)NH, and ^{15}N-edited NOESY experiments were recorded for both IC74$^{84–143}$ and AhpFn; for AhpFn, HNCO, HNCA, HN(CO)CA, and HCC(CO)NH-TOCSY experiments were recorded as well. All NMR experiments were recorded on a 600 MHz Bruker DRX spectrometer, at 20 °C for IC74$^{84–143}$ and 25 °C for AhpFn. Spectra were processed with NMRPipe [5].

6. Discussion

6.1. Automated vs. manual approaches

The development of computers, which has coincided with the development of NMR and X-ray crystallography, has at once made these methods practical and raised the issue of how computing power should best be used: what parts of analysis are better left to human judgement? Higher-order spectra bring out some of the best qualities of computers: they are calculated directly from raw data, and therefore very objective. Generating them involves significant computation, yet it is simple and well-defined. They allow the data to be viewed in a way not possible before the transformation. In addition, the insight is gained "for free" because the required computing power is so cheap. These same qualities are shared by other objective transformations such as the Patterson map and self-rotation function found in crystallography [13,14], and the Fourier transform, essential to both NMR and crystallography.

6.2. Comparison with peak list-based sequential assignment

The performance of automated assignment protocols can only be as good as the peak lists that serve as their input. For all but the best resolved and most intense spectra, abstraction of the spectrum into a peak list becomes a limiting factor, as illustrated by our experience with AhpFn: using Autoassign, we obtained candidates for 141 of 191 possible assignments, 115 of which were later shown to be correct. Through a visual approach based on higher-order spectra using the same raw data as input, we obtained complete assignments, showing that the success of Autoassign was not limited by the spectra themselves, but rather by their abstraction into peak lists.

Automated assignment typically involves an iterative cycle of peak picking, running the automated assignment protocol, and verification. In the initial stages, the process is very efficient: intense, well-resolved peaks are easily found and quickly assigned. The process becomes markedly less efficient in the final stages. Obtaining complete assignments requires identifying not only well-behaved peaks, but those that are weak, overlapped, or obscured by artifacts. Automatic peak pickers can miss such peaks,

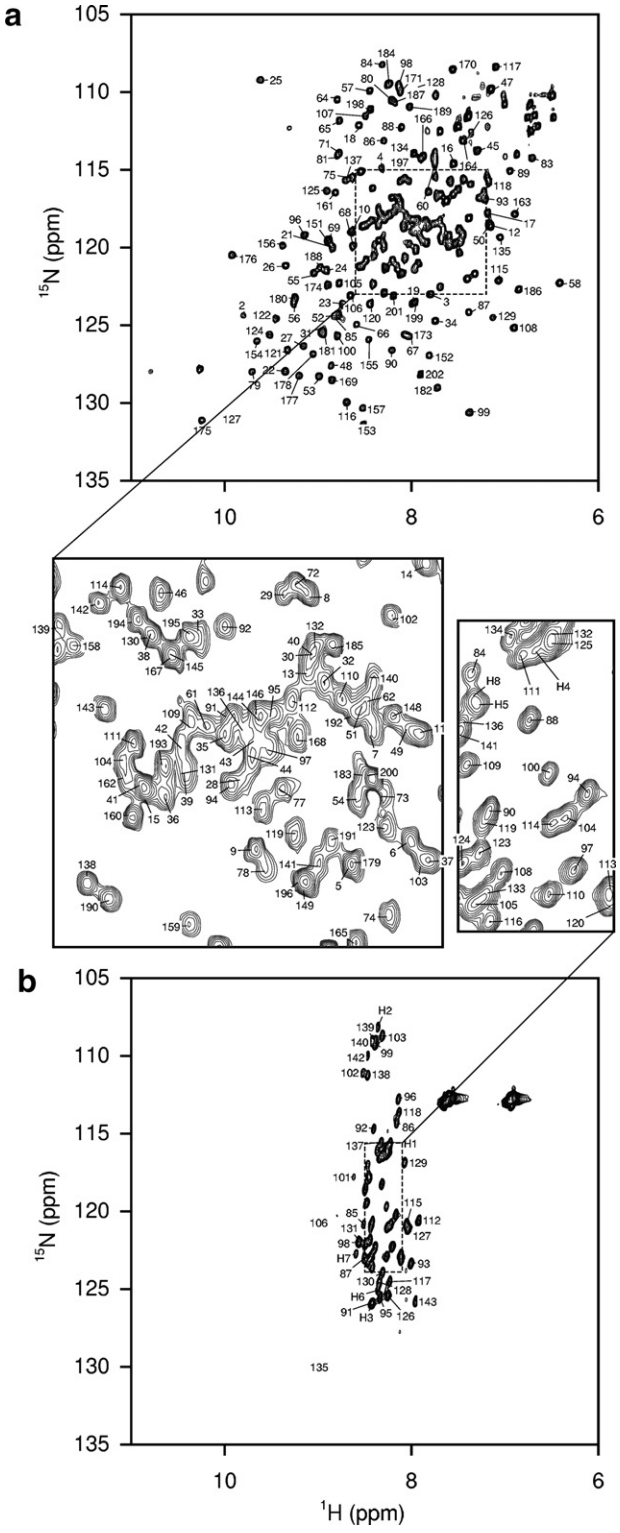


Fig. 8. ^{1}H–^{15}N HSQC spectra with assignments for (a) AhpFn and (b) IC74$^{84–143}$.

so instead time-consuming manual searches must be performed. It is difficult to search entire three-dimensional spectra for the least obvious peaks without any prior knowledge of their positions. It becomes likely in these final stages that spurious peaks due to noise or artifacts will be picked and included in the lists which then become input for the next round of automated assignment, making it quite possible that the result will become worse and not better with more effort devoted to peak picking. For this reason, visual verification (such as through strip plots) becomes critical.

The advantage of the higher-order spectrum approach as implemented by Burrow-owl is that it scales well to these more difficult final stages of assignment. Because all assignments are based on visual inspection of spectra, the all-important visual validation of the result is an integral part of the process. The higher-order spectrum approach is also much more incremental: a typical operation in the final stages is to extend an existing assigned fragment by one residue in either direction, searching for and validating several trial assignments in rapid succession. In doing so, there is no risk of introducing a spurious peak that will degrade the rest of the assignments, the ones already determined.

Our experience with IC74$^{84–143}$ demonstrates that the higher-order spectrum approach is useful not only in the final stages of the assignment process, but in the beginning stages as well, and indeed can be used independently for assigning an entire sequence. For example, the amino acid type-filtered CBCA(CO)NH and HNCACB constructs can be very useful for ''breaking into'' a sequence to begin a new assigned segment; the 4D constructs can then be used to extend that segment.

6.3. Other applications

We have described the application of higher-order spectra to the sequential assignment of proteins using standard CBCA(CO)NH and HNCACB spectra, but the technique is general and can be applied any time there is a need to search for correlated peaks. One possible application is in obtaining protein side-chain assignments from 3D HCCH-COSY type experiments. One commonly uses this spectrum to assign H^{α} and H^{β} knowing C^{α} and C^{β} by searching for a pair of peaks at (H^{β}, H^{α}, C^{α}) and (H^{α}, H^{β}, C^{β}); one could equivalently make a 4D higher-order construct and then search for a single peak at (H^{β}, H^{α}, C^{α}, C^{β}) which allows the H^{α} and H^{β} assignments to be seen directly without the need to search for a correlated pair of peaks. Only a few lines of Scheme code are required to construct such a spectrum.

Although we have presented higher-order spectra as a separate approach from peak picking, it does not need to be: one could generate a peak list from a higher-order spectrum in the same way one generates a peak list from any other spectrum. Peak picking may even work better in higher-order spectra than in the ordinary spectra used to construct them, because the higher-order spectra are typically rather sparse in comparison. The resulting peak lists could be used in an automated assignment protocol.

6.4. Relationship to reduced-dimensionality methods

Though the trend in biomolecular NMR has been to overcome chemical shift degeneracy through corroboration via

multiple correlations for each resonance, there is some indication that this trend may change with recent improvements in techniques such as filter diagonalization (FDM) and *g*-matrix transformation (GFT) for recording very high-dimensional (four or more) spectra [15,16]. These techniques can allow the recording of spectra that traditionally were impractical due to limitations of sensitivity and recording time. Corroboration may become less important in biomolecular NMR as these techniques become more popular. However, it is our impression that most protein sequences are still being assigned using the more traditional three-dimensional experiments based on C^{α} and C^{β} corroboration. Also, our higher-order spectrum approach is sufficiently flexible that it will find use in other contexts besides our example of sequential assignment based on C^{α} and C^{β}.

One of our primitives—the convolution operation—is also used in projection–reconstruction spectroscopy as the first step in reconstructing a full spectrum from reduced-dimensionality projections [17]. In the projection–reconstruction approach, two orthogonal projections are first convoluted to produce a spectrum of full dimensionality containing the peaks of interest plus artifacts which are then filtered out by recording non-orthogonal projections. We have argued here that the convolution operation is useful in other contexts as well, even without the subsequent filtering operation, and even when applied to pairs of spectra that cannot be considered as orthogonal projections of a larger spectrum that could actually be recorded.

7. Summary

In summary, higher-order spectra are a useful way of visualizing NMR data that can serve as a complement to existing assignment techniques or as a complete assignment strategy in its own right. The approach is based on five simple primitive operations that can be combined in ways tailored to specific assignment strategies. We have demonstrated their application to several common spectrum types, and we have provided a general-purpose implementation that would allow them to be applied to almost any other spectrum type. We hope this will serve as a starting point for the biomolecular NMR community to build up a library of applications suited to a variety of assignment tasks.

Acknowledgment

We thank Dr. David Horita of Wake Forest University for providing labeled samples and for recording CBCA(CO)NH and HNCACB spectra of AhpFn.

References

[1] D. Zimmerman, C. Kulikowski, Y. Huang, W. Feng, M. Tashiro, S. Shimotakahara, C. Chien, R. Powers, G. Montelione, Automated analysis of protein NMR assignments using methods from artificial intelligence, J. Mol. Biol. 269 (1997) 592–610.

[2] J. Masse, R. Keller, Autolink: automated sequential resonance assignment of biopolymers from NMR data by relative-hypothesis-prioritization-based simulated logic, J. Magn. Reson. 174 (2005) 133–151.

[3] N. Pawley, J. Gans, R. Michalczyk, APART: automated preprocessing for NMR assignments with reduced tedium, Bioinformatics 21 (2005) 680–682.

[4] R. Koradi, M. Billeter, M. Engeli, P. Güntert, K. Wüthrich, Automated peak picking and peak integration in macromolecular NMR spectra using AUTOPSY, J. Magn. Reson. 135 (1999) 288–297.

[5] F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, A. Bax, NMRPipe—a multidimensional spectral processing system based on Unix pipes, J. Biomol. NMR 6 (1995) 277–293.

[6] H. Abelson, G. Sussman, Structure and Interpretation of Computer Programs, MIT Press, 1996.

[7] F. Zhang, R. Bruschweiler, Indirect covariance NMR spectroscopy, J. Am. Chem. Soc. 126 (2004) 13180–13181.

[8] D.R. Muhandiram, L.E. Kay, Gradient-enhanced triple-resonance 3-dimensional NMR experiments with improved sensitivity, J. Magn. Reson. B 103 (1994) 203–216.

[9] L. Poole, A. Godzik, A. Nayeem, J. Schmitt, AhpF can be dissected into two functional units: tandem repeats of two thioredoxin-like folds in the N-terminus mediate electron transfer from the thioredoxin reductase-like C-terminus to AhpC, Biochemistry 39 (2000) 6602–6615.

[10] S. Grzesiek, J. Anglister, A. Bax, Correlation of backbone amide and aliphatic side-chain resonances in $^{13}C/^{15}N$-enriched proteins by isotropic mixing of ^{13}C magnetization, J. Magn. Reson. B 101 (1993) 114–119.

[11] D. Zimmerman, C. Kulikowski, L. Wang, B. Lyons, G. Montelione, Automated sequencing of amino acid spin systems in proteins using multidimensional HCC(CO)NH-TOCSY spectroscopy and constraint propagation methods from artificial intelligence, J. Biomol. NMR 4 (1994) 241–256.

[12] G. Benison, A. Nyarko, E. Barbar, Heteronuclear NMR identifies a nascent helix in intrinsically disordered dynein intermediate chain: implications for folding and dimerization, J. Mol. Biol. 362 (2006) 1082–1093.

[13] L. Tong, M. Rossmann, Rotation function calculations with GLRF program, Methods Enzymol. 276 (1997) 594–611.

[14] M. Rossmann, D. Blow, The detection of sub-units within the crystallographic asymmetric unit, Acta Cryst. 15 (1962) 24–31.

[15] J. Chen, D. Nietlispach, A. Shaka, V. Mandelshtam, Ultra-high resolution 3D NMR spectra from limited-size data sets, J. Magn. Reson. 169 (2004) 215–224.

[16] S. Kim, T. Szyperski, GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information, J. Am. Chem. Soc. 125 (2003) 1385–1393.

[17] E. Kupce, R. Freeman, Projection–reconstruction technique for speeding up multidimensional NMR spectroscopy, J. Am. Chem. Soc. 126 (2004) 6429–6440.