ELSEVIER

Automation of NMR structure determination of proteins

Amanda S Altieri[1] and R Andrew Byrd[2]

The automation of protein structure determination using NMR is coming of age. The tedious processes of resonance assignment, followed by assignment of NOE (nuclear Overhauser enhancement) interactions (now intertwined with structure calculation), assembly of input files for structure calculation, intermediate analyses of incorrect assignments and bad input data, and finally structure validation are all being automated with sophisticated software tools. The robustness of the different approaches continues to deal with problems of completeness and uniqueness; nevertheless, the future is very bright for automation of NMR structure generation to approach the levels found in X-ray crystallography. Currently, near completely automated structure determination is possible for small proteins, and the prospect for medium-sized and large proteins is good.

Addresses
Structural Biophysics Laboratory, Center for Cancer Research, National Cancer Institute, Frederick, Maryland 21702, USA
[1]e-mail: altieri@ncifcrf.gov
[2]e-mail: rabyrd@ncifcrf.gov

Current Opinion in Structural Biology 2004, **14**:547–553

This review comes from a themed issue on
Biophysical methods
Edited by Arthur G Palmer III and Randy J Read

Available online 15th September 2004

0959-440X/$ – see front matter

DOI 10.1016/j.sbi.2004.09.003

Abbreviations

BMRB	BioMagResBank
GFT	G-matrix Fourier transformation
NOE	nuclear Overhauser enhancement
NOESY	NOE spectroscopy
PCS	pseudocontact shift
PR	projection reconstruction
RD	reduced dimensionality
RDC	residual dipolar coupling
rmsd	root mean square deviation

Introduction

NMR is a prodigious technique that has impacted all aspects of protein structure investigations, even X-ray crystallography, whose structural targets may be selected and refined using NMR spectral properties. The availability of solution structures of proteins cannot be underestimated, even in cases in which crystal structures exist. The basic process of structure determination by NMR has not changed [1••], and includes distinct steps of data collection, data processing, peak picking and editing, resonance assignment, structural parameter assignment and calibration, structure calculation and, finally, structure validation. Initially, these analyses were done interactively, directed by human expertise. Automation of these time-consuming processes is hampered by the wealth of different information obtained from NMR spectra, and by the continued creativity of experimentalists in devising new approaches to separate and measure these parameters. Recent examples include residual dipolar couplings (RDCs) [2] and pseudocontact shifts (PCSs) [3–5]. RDCs provide global, orientational restraints for bond vectors, and PCSs provide both orientational and long-range translational restraints for paramagnetic H^N (or methyl-^{1}H, ^{15}N or ^{13}C atoms) vectors relative to a molecular coordinate system. New experimental restraints also include hydrogen bonds experimentally derived from *trans*-hydrogen bond scalar couplings, as well as isotope-directed sets of limited nuclear Overhauser enhancements (NOEs) [6•]. Regardless of the specific techniques used (specific labeling schemes or specialized pulse sequences), resonance assignment, restraint generation and structure calculation require parsing sets of numbers, multiply correlated in largely expected ways. Successful automation, then, comes down to designing the right tools to correlate, evaluate and properly utilize this data.

There is a critical need for data management, archiving and mining. The interchangeability of data, both raw unprocessed data and intermediate results, between software packages must be addressed before the goal of complete automation can be realized. The NMR community has not reached the same level of consensus in this area as the crystallographic community [7]. The CCPN project [8] has proposed a new all-encompassing data model, although acceptance has not been rapid. In the interim, the NMR-STAR format of the BioMagResBank (BMRB; http://www.bmrb.wisc.edu) is the primary program-independent data format. Cross-validation and comparison of automated procedures will be greatly facilitated by improved data exchange facilities. Two substantive efforts at developing a process-wide data management system have been reported [9,10]. These management systems will not be reviewed here; however, it is important to note that such efforts are underway. Each system has substantial requirements for computer resources and/or commercial licenses of relational database engines; nevertheless, such efforts are crucial to streamlining the overall process.

In this *Opinion*, we will address the recent developments in data collection that may significantly impact automa-

tion, the status of data processing and need for improved automated peak picking facilities, and the current approaches and requirements for automated resonance assignment. The area of automated structure determination that has received the greatest attention in recent years is the structure calculation procedure; we will summarize recent developments for the several available packages.

Data collection

The collection of triple-resonance, multidimensional NMR spectra for structure determination is driven by the need to achieve the highest level of completeness in resonance assignment. Even in larger proteins, for which deuteration is employed, the level of completeness is still pertinent to all sites for which a structural restraint can be determined, including RDCs, PCSs, chemical-shift-based torsion angles, NOEs and so on. A key bottleneck in this process is the lengthy time required to collect all of the raw data. Two significant developments have impacted the time requirement for data collection. First, the dramatic increase in sensitivity achieved by cryogenic NMR probes can greatly reduce the time required to collect conventional data sets. This sensitivity gain is universal, enhancing all data collection methods. Regrettably, this may represent the last dramatic sensitivity gain in NMR for some time. Therefore, further gains must be obtained in other aspects of data collection. The second significant development is a progression of ideas that challenge the traditional linear sampling and Fourier analysis of multidimensional NMR spectra [11••]. An early approach to speeding up data collection was to employ non-linear data sampling in the indirect dimensions, coupled with maximum entropy data processing [12]. This idea has been superseded by the general approach of reduced dimensionality (RD) experiments, pioneered by Szyperski with significant contributions from Marion and Gronenborn (see Kim and Szyperski for a discussion [13]). The RD experiments permit the shortening of data acquisition times by combining the chemical shift evolution of two indirect dimensions into one. The experiments adhere to linear sampling of all frequency dimensions and gain time by reducing the dimensionality from four to three dimensional, or three to two dimensional. The concept was extended by introducing the idea of joint evolution of two indirect dimensions and the acquired data was unraveled using a G-matrix Fourier transformation (GFT) [13,14••]. This concept is very powerful and provides the potential to remove spectral overlap by obtaining higher dimensionality spectra in a significantly reduced time [15], yielding a precise frequency list that is suitable for automated resonance assignment. Recently, the concept of joint time evolution has been extended and combined with the early imaging principles of projection reconstruction (PR) [16] to enable the very efficient acquisition of multidimensional spectra, with high resolution in all indirect dimensions, and permit the reconstruction of conventional-format NMR spectra [17•,18•,19,20]. The PR method has been generalized and applied to a five-dimensional experiment that was recorded in only 100 min [21••]. The PR methods provide spectral data in a conventional format, enabling the use of all existing spectral analysis tools. This feature may be crucial to recognition of artifacts, beneficial to validation of the methods, and to development of automated analysis procedures. Coupled with hardware improvements, these methods may reduce the time required to collect all of the resonance assignment data to 1–2 days, or less.

Data processing and peak picking

After data collection, the processing of data into frequency domain spectra is rather straightforward and, for Fourier analyses, the tools have evolved over time. In the context of automation, Montelione and co-workers illustrate one approach in their AutoProc program [22], which acts as an interface for NMRPipe [23], generating processing scripts based on acquisition parameters that accompany the data. The same principles could be applied to other script-driven processing packages. The new GFT and PR methods require additional processing beyond the standard analysis of the acquired two-dimensional planes, and it is anticipated that this software, which is currently available from the original authors, will be incorporated into the standard processing packages.

Peak picking remains a tricky problem for automation. The critical issues are overlap of resonances, the presence of spectral artifacts and determination of peak thresholds in spectra with a wide range of intensities, as is often observed in NOE spectroscopy (NOESY) spectra. There are several different approaches to peak picking and these have been recently reviewed [1••]. Two recent developments should provide software developers with tools to improve peak picking. First, the ability to lift degeneracy by increasing dimensionality via GFT and PR methods will permit testing of algorithms in correlation spectra. Secondly, the PASD algorithm for structure calculation (see below) is capable of dealing with the very large amount of erroneous data that could arise from poor automated peak picking. The algorithm identifies the bad data, thus providing a means of monitoring and refining the automated peak-picking procedures. In addition, the ATNOS program, discussed below, combines peak identification and assignment iteratively with structure calculation, again with the aim of eliminating the effects of poorly picked data.

Automated resonance assignment

The process of automated resonance assignment has been evolving for several years. The correctness and completeness of the assignment table are critical to the success of automated NOE assignment and structure calculation

(see below). There are several packages available (see related reviews and discussions [1••,24,25•,26••]). The general principle is to automate the rule-based approach that an expert would use to assign a series of related multidimensional spectra. The nature of requiring a specific set of experiments inherently refers back to an overall data management protocol for complete automation. The rules programmed are clearly dependent on the list of spectra required by the package, with limited options for different experiments. Successful assignment relies on observation of all expected resonances within the spectra, as well as the degree of resonance overlap. Automation of backbone resonance assignments is more straightforward than automated sidechain assignments; however, progress is being made [1••,27•]. Although there have been some impressive demonstrations, there remains a general consensus that the automated procedures are a good starting point, but still require some manual assignment and verification. This premise forms the basis of script-driven tools for interactive analysis software packages, such as ANSIG, XEASY, Sparky and NMRView. A good example of a semi-automated package is Smartnotebook [25•]. A recent software package, AVS [28], performs a careful statistical and connectivity check of assignments based on known data within the BMRB. This package operates on standard NMR-STAR files, and should be useful to validate both manual and automated assignment lists. The potential exists to improve automated assignment methods through improved resolution of high-dimensional PR spectra. The completeness may be positively impacted by the improved sensitivity of cryogenic probes; however, it is probable that gaps and miss-assignments will persist in many cases, and must be screened carefully.

Generation of constraints for non-NOE data

Secondary structure analysis and torsion angle constraints come directly from the automated assignment phase by linking to programs such as TALOS [29] and HYPER [30]. The AutoStructure [31•] program uses pattern analysis of specific secondary structure NOE contacts, in addition to chemical shifts, scalar coupling constants and slow amide proton exchange data, to generate dihedral angle and hydrogen bond restraints, as well as distance restraints. Newer structural restraints, such RDCs, have been used in some instances with automated structure calculations by inclusion as an additional external restraint file. Some tools are beginning to appear (e.g. the ipap.tcl script in the NMRPipe package) that enable rapid generation of these tables of RDCs based on known resonance assignments. These tools can easily be adapted to provide similar functionality for PCSs.

NOESY restraint parsing and structure calculation

Several programs were published in the late 1990s for automated or semi-automated structure calculation using NMR data. Most of these programs have been described and reviewed previously [1••,24,26••]. The recent improvements to many of these will be highlighted briefly here.

Two recent reports explore the use of deuteration and the limits of the amount of data necessary for an accurate structure determination, as well as which types of data are most facile to automation. The AutoStructure [31•] program was used to demonstrate that medium-resolution structures can be automatically assigned and calculated quickly for systems with limited NOE data (e.g. from partially deuterated proteins) [6•]. Modifications were made to AutoAssign to include spin system type assignments (STACs) and the distances used in AutoStructure were adjusted for the longer mixing times used with NOESY spectra of deuterated proteins. In a related demonstration [32], the NOAH/DIAMOD suite was used to determine a minimal data set required for accurate fold determination. Improvements to the automation of NOAH were also introduced in this paper and a new graphical user interface (GUI) to NOAH was recently announced [33].

The most recent version of ARIA [34,35] incorporates a correction for spin diffusion using relaxation matrix analysis [36•]. ARIA includes a quick method to calculate the NOE matrix from an ensemble of structures, for which most off-diagonal elements are close to zero. A distance cut-off derived from the ensemble is used to determine which elements are sufficiently small to be set to zero, thereby reducing the number of matrix elements calculated. To accommodate spin diffusion, the cut-off criteria are tailored to expected spin diffusion pathways for one or two intervening spins. The distance corrections are calculated from the ratio between the calculated volume and the isolated spin pair approximate volume. An option to do the final refinement in explicit water [37] is fully integrated into the current version of ARIA.

The programs CLOUDS, SPI and BACUS [38••,39,40] define a new method for automated protein structure determination using NMR data that can be categorized, more or less, as an 'assignment-free' approach. By grouping resonances into connected spin systems, unambiguous NOE identities can be accomplished in an automated manner. Bayesian inference is used to obtain the likelihood of backbone interatomic distances, which is then combined with chemical shifts to calculate probabilities of sequential connectivity. The fundamental concept is that the local environment is restricted, as it is mostly constrained by covalent bonds encoded in the J-connectivities of correlation spectroscopy/total correlation spectroscopy (COSY/TOCSY)-type spectra. For example, if two protons are close to each other, the rest of their J-coupled or NOE-linked spin systems are also likely to occur within the local neighborhood; this

can be used to modify the cross-peak matching probabilities. Quantification of spin system matching probabilities is done using Bayesian inference with systematic tracking of the likelihood of each grouping hypothesis. BACUS automatically establishes probabilistic identities of NOESY cross-peaks in terms of chemical shifts provided by SPI.

CANDID [41] carries out multiple cycles of iterative cross-peak assignment and structure calculation using the torsion angle dynamics program DYANA. CANDID uses network anchoring and constraint combination to obtain a correct *de novo* fold in the first cycle. Network anchoring is based on the premise that any network of correct NOE cross-peak assignments forms a self-consistent subset in an overall set of distance constraints that is sufficiently dense to determine a three-dimensional protein structure. The generalized relative contribution is determined from chemical shift tolerance, cross-peak symmetry, covalent structure compatibility, and the convergence of network anchoring and three-dimensional structure compatibility. Constraint combination is designed to reduce the impact of artifactual NOE upper distance constraints by combining assignments for two or several peaks into a single (virtual) upper limit distance constraint; this lowers the probability that an artifact peak will influence the output of the structure calculation. ATNOS [42] incorporates analysis of the raw NOESY data into the automated structure calculation, and provides feedback between the protein structure, NOE assignments and experimental NOESY spectra. RADAR is a very recent package that combines and tightly merges both CANDID and ATNOS. More information on RADAR should be available soon.

Another method was recently published, PASD [43••], that uses a probabilistic method to automatically calculate high-resolution structures from data in which as much as 80% of the long-range NOE assignments are incorrect (i.e. that were assigned automatically by matching NOE peaks to a chemical shift list). In PASD, the NOE assignments are not dependent on the structure from the previous cycle, therefore reducing the chance that the automated assignments and structure calculation are funneled down the wrong path from an incorrect global fold. The algorithm is designed to be error tolerant and incorporates three main features: first, the restraints are represented as a linear function, which maximizes the number of restraints that are simultaneously satisfied during simulated annealing; second, the algorithm avoids local minima from multiple possible assignments for an NOE by treating each one separately; and third, probabilistic inactivation and reactivation of all NOE restraints, such that no restraints are permanently removed from the calculation. The decision to turn a restraint on or off is evaluated randomly and a large number of evaluations are required for adequate sampling, resulting in simple probabilities to assess the final assignments.

Assignment completeness and chemical shift tolerances

All programs, except for CLOUDS/BACUS, require a resonance assignment list. The precision and accuracy of the resonance assignments, and the assignment completeness are critical. The precision of the assignments is rooted in the consistency of experimental set-up and control during data acquisition to ensure accurate chemical shift referencing and registration across a large number of multidimensional experiments [1••]. Assignment using a database analysis of all occurrences of each chemical shift across the entire set of spectra is a useful approach to assessing the experimental tolerances. Automated NOESY cross-peak assignment is dependent on both the precision (or tolerances) of the chemical shift assignments and the completeness. The assignment completeness that is required has been examined using an experimental data set with randomly excluded restraints [44•]. Up to 10% of the assignments could be excluded without deleterious effect on the resultant structures, provided heteronuclear data were used. The requirement for 90% assignment completeness is contrasted with the assignment completeness of NOESY spectra, for which up to 50% of the NOESY cross-peaks could be excluded without compromising the structures. The affect of incomplete assignment of aromatic residues, however, can be much more dramatic. Exclusion of all aromatic assignments for a protein, constituting only 6% of the total, resulted in 2 Å rmsd from the reference structure. However, when 20% (1.6% overall) of the aromatic residues were omitted for another protein, significant structural deviations occurred.

Hence, the effect of incomplete NOESY assignments depends on the role of the missing residues in establishing the architecture of the protein. *A priori*, it is not possible to predict the impact of completeness and all efforts should be made to improve resonance assignment completeness. Naturally, this includes sidechain assignments and places increased significance on the accuracy of automated assignment tools. Mapping NOESY cross-peak frequencies with those from the assignment list creates the initial list of NOESY assignments. Hence, the chemical shift tolerances used dramatically affect the output list of possible NOESY restraints. If the tolerance placed on the cross-peak assignments for a given spectrum is too tight, incorrect assignments can result, as well as missed assignments. Tolerances that are too loose result in restraints with an excessive level of ambiguity that must be filtered out by the structure calculation protocols. Future advances will combine improvements in assignment completeness and precision with improvements in the robustness of calculation protocols to manage ambiguous and erroneous NOESY assignments [43••].

Extent of automation

The extent to which these programs are automated varies from mostly automated, beginning with data processing through to structure calculation, to programs that are automated only to the extent of performing a specific task. In practice, the resultant assignments and calculated structures from automated programs should be analyzed, at least to some extent, by the user, as wide variation in spectral results from different proteins ensures that no method will work equally well with every system. In addition, most programs require some manual input, by inspection, data conversion or modification of program run-time parameters. Automated structures from all of the programs can be improved by refinement with other data (e.g. RDCs) or inclusion of explicit solvent. The most comprehensively automated program suite to date is AutoProc/AutoAssign/AutoStructure, which creates a project flow from data processing to sequential assignment and structure calculation, including streamlined input/output of data to and from a variety of data formats. Sidechain assignments, however, are not yet automated in this package. The ARIA program is highly automated once sequential assignments have been completed, and has a browser-driven user-friendly interface with input script generation and analysis scripts. ARIA also includes an interface to several interactive assignment programs to facilitate inspection during the assignment and structure calculation process. The programs NOAH [32,45,46] and SANE [47] both assign cross-peak lists from NOESY spectra using a chemical shift list, generate restraints and then parse these restraints iteratively using previously calculated structures. NOAH has recently been updated to more extensively automate this process, by reducing the number of user-supplied parameters and optimizing for combinations of two- and three-dimensional spectra. The CANDID/ATNOS suite is highly automated, again, once a sequential assignment list has been completed. It is unique in combining NOESY peak picking and assignment with iterative structure calculations, with the inclusion of extensive filters and algorithms to validate the cross-peaks. The newest program, PASD, in its current form, parses the list of possible NOESY restraints and calculates structures in three iterations.

The assumptions behind and approaches to automated structure calculation vary among the currently published methods. Most programs rely on a good global fold in the first or second iteration to guide the remaining assignments (ARIA, NOAH, SANE, CANDID and AutoStructure); however, BACUS and PASD do not. CANDID, PASD and CLOUDS/BACUS assume that incorrect restraint assignments are random and do not lead to self-consistency, whereas ARIA assumes errors are correlated. ARIA assumes that many NOE cross-peaks will have multiple contributions based on assignment tolerances (ambiguity), whereas the PASD method assumes a given NOE cross-peak arises from a single interaction, which can have multiple assignments. The two algorithms differ in how the multiple assignments are treated during the calculation, and each program provides tools for analyzing the final assignments or likelihood of assignments, respectively. CANDID does not fare well with large amounts of incorrect data (poorly picked), but does perform well using limited NOE data. PASD is robust to large amounts of incorrect cross-peaks; however, there is a computational cost for extensive statistical sampling. The next step for automation of structure calculations using NMR data can focus on improvements in generating NOE peak lists (data sampling, assignment completeness, assignment accuracy, spectral peak picking), or can further expand and develop methods that avoid pitfalls from inaccurate peak lists.

Conclusions

We have discussed the overall process of automation in protein NMR structure determination and highlighted the problems that the field faces in the near future. The issue of interchange of data appears to be significant in efforts to solicit and encourage groups uninvolved in software development to test and validate different approaches or combinations of approaches. Furthermore,

Table 1

Summary of programs for automated structure calculation.

Program	References	MD engine	Utility
ARIA	[34,35,36•,37]	CNS XPLOR	Ambiguous NOE restraint generation, spin diffusion correction, iterative structure calculation, analysis
AutoStructure	[28,31•]	XPLOR CNS DYANA	NOE, torsion angle and hydrogen bond restraint generation, NOESY assignment, iterative structure calculation, analysis
BACUS/CLOUDS	[38••,39,40]		NOESY assignment, distance matrix calculation
CANDID/ATNOS	[41,42]	DYANA	NOESY peak analysis, NOESY peak assignment, restraint generation, iterative structure calculation
NOAH	[32,33]	DIAMOD DIANA	NOESY assignment, NOE restraint generation, torsion angle restraints, iterative structure calculation
SANE	[47]	AMBER DYANA	NOESY assignment, restraint generation, structure calculation
PASD	[43••]	XPLOR-NIH	Probability analysis of NOE restraints and simultaneous structure calculation

MD, molecular dynamics.

the revolution in data collection afforded by the new GFT and PR methods holds the potential to increase the accuracy and completeness of assignments. This will naturally translate into improved automated methods for picking and assigning NOE spectra. The current status of automated structure calculation packages is very promising (Table 1). Progress in dealing with incorrect data (PASD), and tighter coupling between the spectral restraint data and structure calculation (AutoProc/AutoAssign/AutoStructure and ATNOS/CANDID/RADAR) highlight the developments and future requirements for automated structure calculation. Overall, the potential is indeed bright for the future of automated protein NMR structure determination.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Baran MC, Huang YJ, Moseley HN, Montelione GT: **Automated analysis of protein NMR assignments and structures**. *Chem Rev* 2004, **104**:3541-3556.

•• A thorough review of the problems and current strategies for automated structure determination. The description is more detailed for AutoStructure and affiliated programs; however, together with [25•], a clear state of the art emerges.

2. Bax A: **Weak alignment offers new NMR opportunities to study protein structure and dynamics**. *Protein Sci* 2003, **12**:1-16.

3. Gaponenko V, Altieri AS, Li J, Byrd RA: **Breaking symmetry in the structure determination of (large) symmetric protein dimers**. *J Biomol NMR* 2002, **24**:143-148.

4. Gaponenko V, Sarma SP, Altieri AS, Horita DA, Li J, Byrd RA: **Improving the accuracy of NMR structures of large proteins using pseudocontact shifts as long-range restraints**. *J Biomol NMR* 2004, **28**:205-212.

5. Ikegami T, Verdier L, Sakhaii P, Grimme S, Pescatore B, Saxena K, Fiebig KM, Griesinger C: **Novel techniques for weak alignment of proteins in solution using chemical tags coordinating lanthanide ions**. *J Biomol NMR* 2004, **29**:339-349.

6. Zheng DY, Huang YPJ, Moseley HNB, Xiao R, Aramini J, Swapna GVT, Montelione GT: **Automated protein fold determination using a minimal NMR constraint strategy**. *Protein Sci* 2003, **12**:1232-1246.

• This paper describes the extension of AutoAssign and AutoStructure to the case of perdeuterated methyl protonated proteins (e.g. labeled isoleucine, leucine and valine). This is an important problem for determining the folds of either intermediate-sized proteins or larger proteins using automation procedures.

7. Westbrook JD, Bourne PE: **STAR/mmCIF: an ontology for macromolecular structure**. *Bioinformatics* 2000, **16**:159-168.

8. Fogh R, Ionides J, Ulrich E, Boucher W, Vranken W, Linge JP, Habeck M, Rieping W, Bhat TN, Westbrook J *et al.*: **The CCPN project: an interim report on a data model for the NMR community**. *Nat Struct Biol* 2002, **9**:416-418.

9. Baran MC, Moseley HN, Sahota G, Montelione GT: **SPINS: standardized protein NMR storage. A data dictionary and object-oriented relational database for archiving protein NMR spectra**. *J Biomol NMR* 2002, **24**:113-121.

10. Zolnai Z, Lee PT, Li J, Chapman MR, Newman CS, Phillips GN Jr, Rayment I, Ulrich EL, Volkman BF, Markley JL: **Project management system for structural and functional proteomics: Sesame**. *J Struct Funct Genomics* 2003, **4**:11-23.

11. Freeman R, Kupce E: **New methods for fast multidimensional NMR**. *J Biomol NMR* 2003, **27**:101-113.

•• An excellent review of new data acquisition methods, including filter diagonalization and Hadamard methods, which were not discussed here.

12. Schmieder P, Stern AS, Wagner G, Hoch JC: **Improved resolution in triple-resonance spectra by nonlinear sampling in the constant-time domain**. *J Biomol NMR* 1994, **4**:483-490.

13. Kim S, Szyperski T: **GFT NMR experiments for polypeptide backbone and C-13(beta) chemical shift assignment**. *J Biomol NMR* 2004, **28**:117-130.

14. Kim S, Szyperski T: **GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information**. *J Am Chem Soc* 2003, **125**:1385-1393.

•• The first detailed description of GFT NMR, setting forth a procedure to extract frequency lists of peaks corresponding to high dimensionality (e.g. up to five dimensions). The process is described in general and is already being applied by numerous groups.

15. Xia YL, Arrowsmith CH, Szyperski T: **Novel projected 4D triple resonance experiments for polypeptide backbone chemical shift assignment**. *J Biomol NMR* 2002, **24**:41-50.

16. Hounsfield GN: **Computerized transverse axial scanning (tomography). 1. Description of system**. *Br J Radiol* 1973, **46**:1016-1022.

17. Kupce E, Freeman R: **Projection-reconstruction of three-dimensional NMR spectra**. *J Am Chem Soc* 2003, **125**:13958-13959.

• This paper provides the first illustration of reconstructing protein three-dimensional spectra and describes the equivalence of the spectral format to conventional Fourier transform methods, thus illustrating the use of existing graphical analysis tools.

18. Kupce E, Freeman R: **Reconstruction of the three-dimensional NMR spectrum of a protein from a set of plane projections**. *J Biomol NMR* 2003, **27**:383-387.

• A good explanation of the principles behind the PR method and illustration of the increased resolution that is obtainable, concomitant with the time savings.

19. Kupce E, Freeman R: **Fast reconstruction of four-dimensional NMR spectra from plane projections**. *J Biomol NMR* 2004, **28**:391-395.

20. Kupce E, Freeman R: **Projection-reconstruction technique for speeding up multidimensional NMR spectroscopy**. *J Am Chem Soc* 2004, **126**:6429-6440.

21. Coggins BE, Venters RA, Zhou P: **Generalized reconstruction of n-D NMR spectra from multiple projections: application to the 5-D HACACONH spectrum of protein G B1 domain**. *J Am Chem Soc* 2004, **126**:1000-1001.

•• This paper describes a generalization of the PR method described by Kupce and Freeman [17•,18•,19,20] to any dimensionality. A five-dimensional experiment is illustrated, similar to what has been reported by GFT [13]. A key advantage is the software to enable reconstruction of any n-dimensional subspectrum from a limited number of two-dimensional planes (e.g. 12 two-dimensional planes allow reconstruction of any subset of the five-dimensional spectrum).

22. Monleon D, Colson K, Moseley HN, Anklin C, Oswald R, Szyperski T, Montelione GT: **Rapid analysis of protein backbone resonance assignments using cryogenic probes, a distributed Linux-based computing architecture, and an integrated set of spectral analysis tools**. *J Struct Funct Genomics* 2002, **2**:93-101.

23. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A: **NMRPipe: a multidimensional spectral processing system based on UNIX pipes**. *J Biomol NMR* 1995, **6**:277-293.

24. Moseley HN, Montelione GT: **Automated analysis of NMR assignments and structures for proteins**. *Curr Opin Struct Biol* 1999, **9**:635-642.

25. Slupsky CM, Boyko RF, Booth VK, Sykes BD: **Smartnotebook: a semi-automated approach to protein sequential NMR resonance assignments**. *J Biomol NMR* 2003, **27**:313-321.

• An example of a script-driven semi-automated approach to resonance assignment used in conjunction with NMRView; similar procedures could be developed for other visualization programs. This type of analysis might be used to rapidly validate or correct assignments from an automated routine, and could benefit from the AVS program output [28].

26. Guntert P: **Automated NMR protein structure calculation**.
•• *Prog Nucl Magn Reson Spectrosc* 2003, **43**:105-125.
A comprehensive review of automated structure calculation, including discussion of chemical shift assignment and peak picking issues. A detailed discussion and comparison of the NOAH, ARIA, AutoStructure and CANDID programs is given.

27. Malmodin D, Papavoine CH, Billeter M: **Fully automated**
• **sequence-specific resonance assignments of hetero-nuclear protein spectra**. *J Biomol NMR* 2003, **27**:69-79.
An exciting illustration of automatic assignment of backbone and side-chain resonances via a combination of the early programs AUTOPSY and GARANT with a new program, PICS.

28. Moseley HNB, Sahota G, Montelione GT: **Assignment validation software suite for the evaluation and presentation of protein resonance assignment data**. *J Biomol NMR* 2004, **28**:341-355.

29. Cornilescu G, Delaglio F, Bax A: **Protein backbone angle restraints from searching a database for chemical shift and sequence homology**. *J Biomol NMR* 1999, **13**:289-302.

30. Tejero R, Monleon D, Celda B, Powers R, Montelione GT: **HYPER: a hierarchical algorithm for automatic determination of protein dihedral-angle constraints and stereospecific (CH2)-H-beta resonance assignments from NMR data**. *J Biomol NMR* 1999, **15**:251-264.

31. Huang YPJ, Swapna GVT, Rajan PK, Ke HP, Xia B, Shukla K,
• Inouye M, Montelione GT: **Solution NMR structure of ribosome-binding factor A (RbfA), a cold-shock adaptation protein from *Escherichia coli***. *J Mol Biol* 2003, **327**:521-536.
This paper provides a description of structure determination using the AutoStructure program, including statistics for the NOESY cross-peak assignment results.

32. Oezguen N, Adamian L, Xu Y, Rajarathnam K, Braun W: **Automated assignment and 3D structure calculations using combinations of 2D homonuclear and 3D heteronuclear NOESY spectra**. *J Biomol NMR* 2002, **22**:249-263.

33. Xu Y, Oezguen N, Braun W: **A GUI based NOAH/DIAMOD programs for automated NOESY spectra assignment and protein structure determination**. *Biophys J* 2003, **84**:275A.

34. Nilges M, Macias MJ, Odonoghue SI, Oschkinat H: **Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from beta-spectrin**. *J Mol Biol* 1997, **269**:408-422.

35. Linge JP, O'Donoghue SI, Nilges M: **Automated assignment of ambiguous nuclear Overhauser effects with ARIA**. In *Nuclear Magnetic Resonance of Biological Macromolecules*, part B. Edited by James TL, Dotsch V, Schmitz U. London: Academic Press; 2001:71-90. [Book series: Methods in Enzymology, vol 339.]

36. Linge JP, Habeck M, Rieping W, Nilges M: **Correction of spin**
• **diffusion during iterative automated NOE assignment**. *J Magn Reson* 2004, **167**:334-342.
A lucid description of the relaxation matrix analysis method used to correct for spin diffusion.

37. Linge JP, Williams MA, Spronk C, Bonvin A, Nilges M: **Refinement of protein structures in explicit solvent**. *Proteins* 2003, **50**:496-506.

38. Grishaev A, Llinas M: **BACUS: a Bayesian protocol for the**
•• **identification of protein NOESY spectra via unassigned spin systems**. *J Biomol NMR* 2004, **28**:1-10.
The third in a set of unique programs to determine a three-dimensional protein structure before any sequential chemical shift assignments. The method is computationally quick, and does not require an initial structural model or iterative calculations.

39. Grishaev A, Llinas M: **CLOUDS, a protocol for deriving a molecular proton density via NMR**. *Proc Natl Acad Sci USA* 2002, **99**:6707-6712.

40. Grishaev A, Llinas M: **Sorting signals from protein NMR spectra: SPI, a Bayesian protocol for uncovering spin systems**. *J Biomol NMR* 2002, **24**:203-213.

41. Herrmann T, Guntert P, Wuthrich K: **Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA**. *J Mol Biol* 2002, **319**:209-227.

42. Herrmann T, Guntert P, Wuthrich K: **Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS**. *J Biomol NMR* 2002, **24**:171-189.

43. Kuszewski J, Schwieters CD, Garrett DS, Byrd RA, Tjandra N,
•• Clore GM: **Completely automated, highly error-tolerant macromolecular structure determination from multidimensional nuclear Overhauser enhancement spectra and chemical shift assignments**. *J Am Chem Soc* 2004, **126**:6258-6273.
This paper describes a new algorithm for automatic assignment of NOESY restraints integrated with structure calculation. The method begins with a random structure and is computationally intensive; however, it is highly error tolerant.

44. Jee J, Guntert P: **Influence of the completeness of chemical**
• **shift assignments on NMR structures obtained with automated NOE assignment**. *J Struct Funct Genomics* 2003, **4**:179-189.
A detailed analysis of the effects of chemical shift assignment completeness on structure calculations, including the serious but occasionally overlooked hazard of incomplete aromatic resonance assignments. The topic of NOESY assignment completeness is also presented.

45. Mumenthaler C, Guntert P, Braun W, Wuthrich K: **Automated combined assignment of NOESY spectra and three-dimensional protein structure determination**. *J Biomol NMR* 1997, **10**:351-362.

46. Mumenthaler C, Braun W: **Automated assignment of simulated and experimental NOESY spectra of proteins by feedback filtering and self-correcting distance geometry**. *J Mol Biol* 1995, **254**:465-480.

47. Duggan BM, Legge GB, Dyson HJ, Wright PE: **SANE (Structure assisted NOE evaluation): an automated model-based approach for NOE assignment**. *J Biomol NMR* 2001, **19**:321-329.