Automated analysis of NMR assignments and structures for proteins

Hunter NB Moseley* and Gaetano T Montelione†

Recent developments in protein NMR technology have provided spectral data that are highly amenable to analysis by advanced computer software systems. Specific data collection strategies, coupled with these computer programs, allow automated analysis of extensive backbone and sidechain resonance assignments and three-dimensional structures for proteins of 50 to 200 amino acids.

Addresses
Center for Advanced Biotechnology and Medicine, and Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, New Jersey 08854-5638, USA
*e-mail: hunter@cabm.rutgers.edu
†e-mail: guy@cabm.rutgers.edu

Current Opinion in Structural Biology 1999, **9**:635–642

0959-440X/99/$ – see front matter

Abbreviations

2D	two-dimensional
3D	three-dimensional
COSY	scalar coupling correlated spectroscopy
HSQC	heteronuclear single quantum coherence correlated spectroscopy
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect correlated spectroscopy
rmsd	root mean square deviation
TOCSY	total scalar-coupling correlated spectroscopy

Introduction

A powerful feature of macromolecular structure analysis by NMR spectroscopy is its potential for automation. It has been recognized for some time that many of the interactive tasks carried out by an expert during the process of spectral analysis could, in principle, be carried out more efficiently and rapidly by computational systems. With the advent of multidimensional and triple-resonance strategies for determining resonance assignments and 3D structures, it became increasingly clear that the quality and information content of protein NMR spectra could allow largely automated analyses of assignments and structures for small proteins. Over the past few years, this potential has been realized to some degree and key steps in many production structure analyses are now carried out using automated methods. This advance has tremendous implications for the growing role of NMR spectroscopy as a powerful and accessible tool for biophysical chemistry, drug design and structural genomics. In this review, we summarize recent advances in automating the processes of determining 3D structures of proteins from NMR data.

Peak-picking algorithms

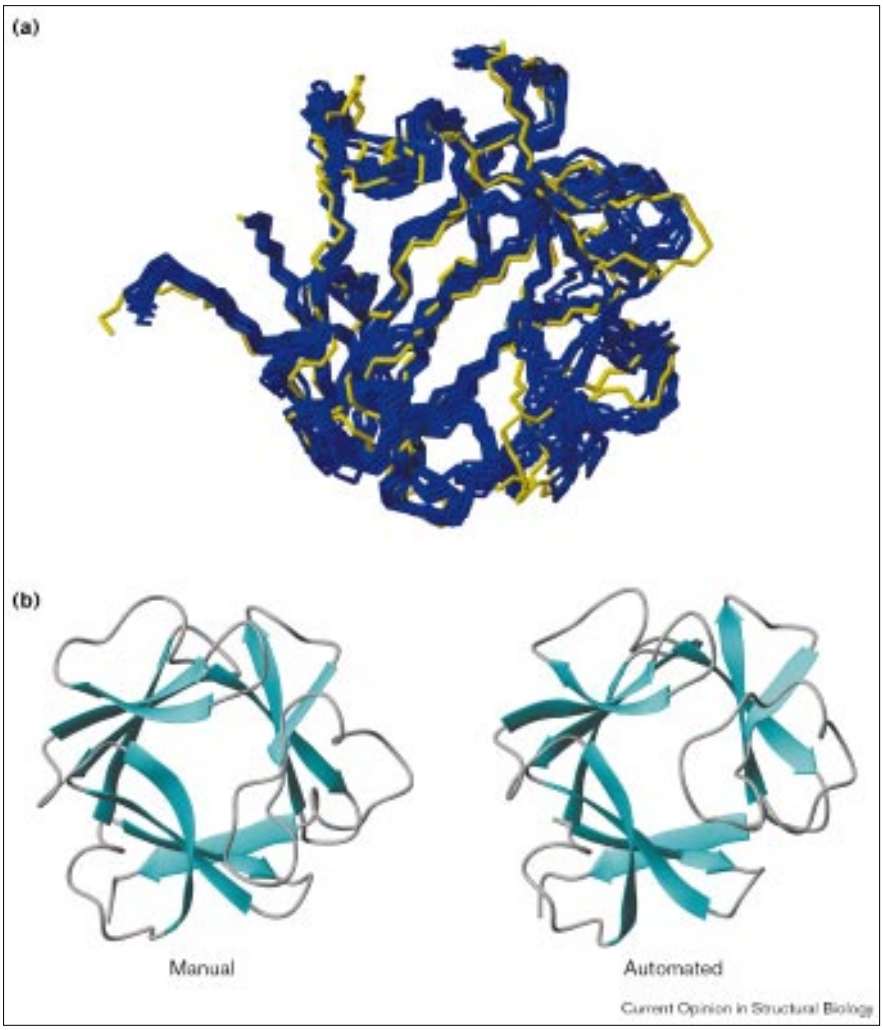
High-quality peak picking is crucial for successful automated spectral analysis. Programs that automate peak picking must deal with the problem of identifying real peaks and/or excluding artifactual peaks. Several high-quality peak-picking programs have been described over the past few years. Most allow for interactive validation and editing by a user. AUTOPSY [1•] is a comprehensive program for automating peak picking with facilities for determining noise level, segmenting the spectra into peak-containing regions, identifying well-separated peaks, resolving spectral overlap and integrating peaks. It is very robust and appears to have features that make it well suited to preparing input for automated analysis programs. Schulte *et al.* [2] have described a Bayesian statistical method that distinguishes between real and artifactual peaks that have been automatically picked by another program.

Automated analysis of resonance assignments

Resonance assignments form the basis for characterizing secondary structure, dynamics, intermolecular interactions and 3D structures of proteins. Significant progress has been made recently in the automated analysis of resonance assignments [3•–5•,6••,7••,8•], particularly using triple-resonance NMR data [9•]. Several laboratories are developing programs that automate either backbone or complete resonance assignments using a variety of final mapping methods (Table 1). These include programs developed by Lukin *et al.* [6••], Leutner *et al.* [8•] and Buchler *et al.* [4•], which use simulated-annealing-like methods, by Bartels *et al.* (GARANT) [3•], which uses genetic algorithms, and by Li and Sanctuary [5•] and Zimmerman *et al.* (AutoAssign) [7••], which use rules-based deterministic (best-first) algorithms.

Most automation programs use the same general analysis scheme: step 1, filter peaks (filtering) and relate resonances from different spectra (referencing); step 2, group resonances into spin systems (grouping); step 3, identify the amino acid type of spin systems (typing); step 4, find and link sequential spin systems into segments (linking); and step 5, map spin-system segments onto the primary sequence (mapping). Different automation programs implement each step with varying degrees of success; however, the overall robustness is dictated by the performance of the weakest step. Programs lacking key steps show significant limitations in robustness [3•,10–12]. Some laboratories focus their attention on specific steps of the scheme, providing computer-assisted assignment programs that show possible results after grouping [10], typing [13], linking [12,14•] or mapping (i.e. generating a list of candidate assignments, instead of definitive assignments) [11,15].

Figure 1



Automatic structure determination of basic fibroblast growth factor. **(a)** Superposition of 10-NMR-derived structures of basic fibroblast growth factor computed using the automated analysis program AutoStructure (Y Huang, R Tejero, GT Montelione, unpublished data) and comparison with one structure (yellow) from the ensemble determined by manual analysis of the same data [42]. Backbone conformations are shown only for residues 29–155, as the N-terminal polypeptide segment is not well defined in either the automated or manual analysis. For this portion of the structure, the backbone rmsds within the families of structures determined by AutoStructure and manual structural analysis are 0.6 Å and 0.3 Å, respectively. The backbone rmsd between the AutoStructure and manually determined NMR structures is 0.7 Å. **(b)** Comparison of overall chain folds, as determined by AutoStructure and manual analysis.

Step 1 – filtering and referencing

All programs filter peaks. Intensity is the primary filtering method used by all programs, either directly or indirectly. Resonance consistency across spectra is another important filtering method [3•–5•,6••,7••,8•,10]. Programs using multiple spectra generally reference the resonances across the spectra. Some programs assume that the referencing is correct. Others calculate self-consistent referencing using common isolated peaks found in multiple spectra [6••,7••].

Step 2 – grouping

Most programs also group resonances into spin systems that are related to a single amino acid or dipeptide. Many methods group resonances via common 'root' resonances found in all or most of the spectra [4•,6••,7••,8•,10]. Some grouping methods use bond patterns and/or bond-pattern templates to group resonances into spin systems [5•,10,11,15]. Bond-pattern methods are sensitive to incomplete peak lists and overlap, however, and can fail when peaks are missing. The program from Croft *et al.* [10] uses a set of pruning rules to limit the list of possible spin systems. The program from van Geerestein-Ujah *et al.* [12] uses a maximum common subgraph isomorphism algorithm to identify and group resonances into secondary structure segments. GARANT [3•,16] groups peaks into a graph representing peak relationships based upon common resonances.

Step 3 – typing

With the exception of GARANT [3•,16], all the automation programs classify amino acid spin systems with respect to possible amino acid types. One common typing method involves matching spin systems to bond-pattern templates [4•,5•,10]; however, spin-system typing from bond-pattern templates, like the bond-pattern methods described above, is very sensitive to the completeness of the data. Another common typing method involves statistical analysis of chemical shifts, specifically the C^{α} and C^{β} resonance shifts [6••,7••,8•]. Programs from both Zimmerman *et al.* [7••] and Lukin *et al.* [6••] use Bayesian statistical methods

Table 1

Recently developed software for the automated analysis of resonance assignments and/or 3D structures.

Program name or authors	Utility	Applicable spectral data
Li and Sanctuary [5•]	Complete resonance assignments	Any HN-detected triple-resonance experiment. Any heteronuclear TOCSY or COSY.
AutoAssign [7••]	Backbone resonance assignments	HNCO HN(CA)CO, HNCA, HN(CO)CA, HNCACB, HN(CO)CACB, HNHA, HN(CO)HA.
Lukin *et al.* [6••]	Backbone resonance assignments	HNCA, HN(CA)CO, HNHA, HNCACB, COCAH, HCA(CO)N, HNCO, HN(CO)CA, HN(CO)HA, CBCA(CO)NH.
PASTA [8•]	Backbone resonance assignments	Any experiment containing only HN, NH, CA, CB, CO and HA resonances.
Buchler *et al.* [4•]	Backbone resonance assignments	HNCA, HN(CO)CA, HNHA, HA(CO)NH, (H)CCH-TOCSY, (H)C(CACO)NH-TOCSY.
GARANT/ DYANA [3•,16,23•]	Complete resonance assignments, 3D structure determination.	Any NMR experiment
NOAH/DYANA/ DIAMOD [27•,28,30••]	3D structure determination	2D and/or 3D NOESY
ARIA/ X-PLOR [31••]	3D structure determination	2D, 3D and/or 4D NOESY
AutoStructure/ DYANA/ X-PLOR/CONGEN	3D structure determination	2D and/or 3D NOESY

combined with BioMagResDatabase statistics to classify spin-system types. Several papers describe the improvement of these statistical methods for spin-system typing by including resonance-phase-labeled experiments [17–19], spin-system edited experiments [20,21] and structural corrections based on C^{α} and C^{β} chemical shifts [22]. A program from Huang *et al.* [13] types spin systems via a neural network using 3D TOCSY-HSQC (total scalar-coupling correlated spectroscopy — heteronuclear single quantum coherence correlated spectroscopy) data.

Step 4 – linking

Most programs find and link sequential spin systems into segments. There are two major linking methods: deterministic best-first methods and energy optimization algorithms, like simulated annealing. The deterministic best-first methods require a one-time, all-to-all, full comparison between every spin system. With the results sorted, the best (most reliable) links are established first, thus reducing the size of the problem [5•,6••,7••,14•]. Optimization algorithms use a pseudo-energy function with simulated annealing to evaluate potential connections between nearest neighbor spin systems [4•,8•]. These optimization algorithms are generally rather slow and can be susceptible to becoming trapped in local minima that correspond to incorrect assignment configurations.

Step 5 – mapping

The final step is mapping spin-system segments onto the primary sequence. As with linking, the most common mapping methods are deterministic best-first methods [5•,7••] and energy optimization algorithms (mostly derivatives of simulated annealing) [3•,4•,6••,8•]. Frequently, the deterministic best-first methods use constraint propagation and other rule-based algorithms to improve accuracy and efficiency [5•,7••]. Some simulated-annealing-like energy optimization algorithms either simplify the mapping problem [6••] or smooth out the energy surface [4•] in order to improve performance. Other energy optimization algorithms, like genetic algorithms, naturally work well with the choppy energy surfaces that arise from discreet changes in mapping [3•]. The GARANT program [3•,16] uses a genetic algorithm to map expected peaks to observed peaks.

Each automation program has its own strengths and weaknesses arising from its specific implementation of the general scheme. The program from Li and Sanctuary [5•] performs automated complete resonance assignments. The program can use any standard triple-resonance and heteronuclear TOCSY and COSY (scalar coupling correlated spectroscopy) experiments; however, testing has focused on the following experiments: HNCO, HNCA, HCACO, HN(CO)CA, ^{15}N-edited TOCSY-HMQC, HCCH-COSY and HCCH-TOCSY. The program's methods include grouping via bond-pattern templates applied with a constraint partitioning algorithm, bond-pattern typing, and best-first linking and mapping. The reported testing of the program was limited to one protein, the first 90 residues of chicken skeletal troponin C. The result of the testing was the assignment of only about one-third of the residues. This result reflects the brittleness of bond-pattern methods.

The AutoAssign program [7••] performs automated backbone resonance assignments. AutoAssign can analyze input from eight triple-resonance experiments (although, generally, only five or six are required): HNCO, HN(CA)CO (not required), HNCACB, HN(CO)CACB, HNCA, HN(CO)CA, HNHA (not required) and HN(CO)HA (not required). AutoAssign's methods include intensity and resonance consistency filtering, isolated peak referencing, root resonance grouping, CA/CB Bayesian statistics typing, best-first linking and a best-first constraint propagation algorithm for mapping. Its testing has been very rigorous and included real data sets for 11 different proteins, with sizes ranging from 6 to 18.7 kDa. The results

of these tests were excellent. They have an average assignment rate of 96% and the lowest average error rate (< 0.5%) among the surveyed methods ([7••]; GT Montelione, HNB Moseley, unpublished data).

The program from Lukin *et al.* [6••] also performs automated backbone resonance assignments. The program uses input from 10 experiments: HNCA, HN(CA)CO, HNHA, HNCACB, COCAH, HCA(CO)N, HNCO, HN(CO)CA, HN(CO)HA and CBCA(CO)NH. The program's methods include intensity and resonance consistency filtering, a Bayesian statistics referencing, a Bayesian statistics approach to root resonance grouping, CA/CB/CO/N Bayesian statistics typing, best-first linking and simulated annealing mapping. It has been tested on real data sets for three proteins: calmodulin (148 residues), the CheY-binding domain of CheA (134 residues) and glutamine-binding protein (226 residues). The results of these tests are also excellent, providing an average assignment completeness of approximately 95% and an average error rate of approximately 2%.

The PASTA (Protein Assignment by Threshold Accepting) program from Leutner *et al.* [8•] performs automated backbone resonance assignments. PASTA uses input from any experiment containing only HN, N, CA, CB, CO and HA resonances. PASTA's methods include intensity and resonance consistency filtering, root resonance grouping, CA/CB random coil statistical typing and a threshold-accepting (simulated-annealing-like) algorithm for linking and mapping. Its testing was rigorous and involved four different simulated data sets (derived from published NMR data) and one real data set (human nonpancreatic synovial phospholipase A_2, with 124 residues). Many subsets of the simulated data sets were used to test PASTA's robustness. The results from these tests are excellent. PASTA had complete assignments, with no errors for its one real data set. Furthermore, it assigned over 90% of the simulated data sets with up to 50% of the signals missing. Its error rate was approximately 1%.

The program from Buchler *et al.* [4•] performs automated backbone resonance assignments using input from six experiments: HNCA, HN(CO)CA, HNHA, HA(CO)NH, (H)CCH-TOCSY and (H)C(CACO)NH-TOCSY. The program's methods include intensity and resonance consistency filtering, root resonance grouping, bond-pattern typing and mean-field simulated annealing for linking and mapping. Its testing was rigorous and included testing one protein (the 172-residue peptide-binding domain of the *Escherichia coli* heat-shock protein DnaK) and many data subsets of this protein. The results for this protein when using complete data were flawless; however, the program shows significant brittleness when given more realistic, imperfect data sets. Even 30% missing data can degrade the performance to approximately 55% assignment with reasonable data quality.

The GARANT program from Bartels *et al.* [3•,16] performs complete resonance assignments using homologous structures and/or homologous chemical shifts when available. The program uses input from any experiment, but testing included homonuclear TOCSY, COSY and NOESY (nuclear Overhauser effect correlated spectroscopy), heteronuclear NOESY and CBCA(CO)NH data. The program's methods include resonance consistency filtering, common resonance grouping around related resonances and mapping expected peaks to observed peaks using genetic algorithms with annealed recombination. Results have been reported for tests on three real protein data sets: Tendamistat (R19L) (74 residues), Antp (C39S/W56S) homeodomain (68 residues) and free cyclophilin A (165 residues) [16]. The results provide approximately 97% of the possible backbone resonance assignments, with an approximately 11% error rate. The sidechain assignments were roughly 80% complete, with an unknown error rate. These results, however, required chemical shifts from a homologous protein as additional input. GARANT also has been used in the automated analysis of NOESY cross-peak assignments in refining the 3D structure of the 150-residue peptide deformylase from *E. coli* [23•].

Automated three-dimensional structure determination

Significant progress has also been made in developing programs for automated 3D protein structure determination (Table 1). In protein NMR spectroscopy, structure-generation calculations are generally carried out using the following data as input: distance constraints based on the analysis of multidimensional NOESY spectra; constraints on dihedral angles derived from experimental and/or statistical data, including NOESY, chemical shift and scalar coupling constant data; and residual dipolar couplings. In some cases, disulfide and/or hydrogen bond distance constraints that have been derived from other experimental data are also included.

Several approaches have been described for identifying backbone and/or sidechain dihedral-angle constraints using simultaneous analysis of nuclear Overhauser effect (NOE), scalar coupling and/or chemical shift data. Gippert *et al.* [24•] have described two complementary approaches involving a systematic search in torsion-angle space for the generation of all conformations of polypeptides that satisfy the local conformational constraints. Protein backbone ϕ and ψ constraints have also been derived by comparing experimental chemical shifts with a database of high-resolution crystal structures for which resonance assignments are available [25•,26]. These methods provide automated approaches for generating both dihedral-angle constraints and starting conformations that are consistent with these local constraints.

One of the principal goals of automated structure determination programs involves the iterative analysis of multidimensional NOESY data in order to refine and extend the list of distance constraints. Owing to the extensive

degeneracy of protein proton resonances, absolutely unambiguous assignments can only be made for a very small fraction of 2D NOESY peaks. ^{13}C- and ^{15}N-edited 3D and 4D spectra provide some resolution of these proton resonance degeneracies, particularly by the identification of symmetric NOESY cross peaks. Even in these multidimensional spectra, however, when using matching to chemical shift data alone, most individual cross peaks are assigned to several possible pairs of interacting hydrogen atoms. The process of structural analysis and refinement involves resolving these ambiguities of NOESY cross-peak assignments.

A key aspect in the analysis of NOESY cross-peak ambiguities involves the 'match tolerances', used in matching the resonance frequencies of peaks in the NOESY spectra with frequencies in the resonance assignment table [27•]. If these tolerances are too loose, the number of potential cross-peak assignments can become intractably large; if the tolerances are too tight, it is possible to exclude the correct candidate from the resonance assignment table. Ideally, the spectra used for determining resonance assignments should be collected using identical sample conditions as for the NOESY spectra that will be analyzed. For practical reasons, this is rarely the case and larger than optimal 'match tolerance' values are required to account for nonsystematic variations in chemical shifts between the resonance assignment table and NOESY peak lists.

The secondary structures and/or protein folds generated during the initial cycles of the structure-refinement process or during homology modeling can often be used to resolve ambiguities in making NOESY cross peak assignments. Consider, for example, a NOESY cross peak that is assignable to one (or both) interactions A–B and A–C, owing to the degeneracy of the resonance frequencies of atoms B and C. If atoms A and B are nearby one another in the partially refined protein structure, while the distance between A and C is large, the cross peak can potentially be assigned to the A–B, rather than the A–C, interaction. Possible NOESY assignments are 'ruled in' if the corresponding distance is less than a defined value (e.g. 5 Å) for a certain percentage of the ensemble of structures computed at each cycle of the iterative structural analysis. Although generally quite useful, this reasoning can sometimes result in incorrect assignments when one or more of these distances has a large variance among the set of structures computed in the current cycle of analysis.

One successful strategy, used by the program NOAH, is referred to as the 'self-correcting distance geometry' (SEDOC) method [27•,28]. For each ambiguous NOESY cross peak, NOAH computes the violations associated with each candidate assignment for each member of the ensemble. These violations are then used to estimate P_{vio}, the percentage of structures in which the corresponding distance constraint (plus a tolerance distance) is violated. Values of P_{vio} are then used to 'rule in' and 'rule out' candidate NOESY cross-peak assignments. Incorrect assignments made in this process can, in principle, be identified as consistent violations in the next cycle of structure calculations. The NOAH analysis program has been combined with the structure-generation programs DYANA [29] and DIAMOD [30••] for the iterative analysis of NOESY cross-peak assignments and structure refinement. Extensive tests have been carried out on several real protein data sets [27•,28,30••]. Most of these tests have used a subset of manually assigned tertiary NOEs to initiate the iterative analysis process. In some cases, manually validated and manicured NOESY peak lists and/or 'adapted' chemical shift lists adjusted manually for each different NOESY spectrum were used in the input [27•]. The automated method assigned 70–90% of all the cross peaks in 2D or 3D NOESY spectra, which is, on average, only 10% less than the corresponding manual analyses, and generated 3D protein structures with backbone rmsds of 0.6 to 1.5 Å between automatically and manually generated average structures. Using real homonuclear 2D NMR data for the 46-residue protein crambin (S22/I25), together with a subset of manually assigned tertiary NOEs, NOAH/DIAMOD automatically generated a family of structures, the best 10 of which converged with a backbone rmsd of 1.5 Å [30••]. These automatically generated structures are also quite similar to the structure of same mutant of crambin as determined by X-ray crystallography, with backbone rmsds between the X-ray and NMR structures of 2.2 Å for the whole molecule and 1.2 Å for nonloop regions. Efforts to carry out the same automated analysis without a subset of manually assigned constraints to initiate the process provided a less reliable set of NOESY cross-peak assignments, but resulted in similar 3D structures [30••].

An alternative, and more widely used, approach for resolving ambiguities in NOESY cross-peak assignments, Ambiguous Restraints for Iterative Assignment (ARIA) [31••], has been implemented as part of the X-PLOR structure-generation program [32,33]. ARIA assigns ambiguous NOEs during the structure calculation using a combination of ambiguous distance constraints and an iterative assignment strategy. An ambiguous NOE corresponds to a summed distance D:

$$D = \left[\sum_{i=1}^{N} D_i^{-6} \right]^{-1/6}$$

where the index i runs through all N candidate cross-peak assignments and D_i is the distance between the corresponding pairs of protons [31••]. This distance D is then restrained between upper and lower bounds determined by the calibrated cross-peak intensity. In this way, every possible NOESY assignment contributes to the constraint, with weights that depend on the inverse-sixth power of the corresponding distance in the partially refined structure. All (assigned and ambiguous) NOEs can be used together throughout the course of the structure refinement. In most cases, ARIA is used together with an initial 3D protein

structure that has been generated from manually assigned NOEs. Although limited test results have been reported in the literature, the ARIA approach has been recently used successfully in several experimental structure determinations and refinements [34–36]. Particularly robust results have been obtained using 4D ^{13}C-edited NOESY data [35,36], which are difficult to analyze manually.

AutoStructure (Y Huang, R Tejero, GT Montelione, unpublished data) is an expert system for automatic iterative analysis of protein NOESY spectra and structure generation using one or more of the structure-generation programs DYANA [29], X-PLOR [32] or CONGEN [37]. AutoStructure uses rules for assignments that are similar to those used by an expert to generate an initial protein fold: the identification of secondary structural elements using chemical shift, scalar coupling, amide hydrogen exchange and automatically assigned NOESY data; the use of symmetry features of 3D ^{13}C-edited NOESY spectra to resolve some NOESY assignment ambiguities; distances computed for specific secondary structures to 'rule in' and 'rule out' candidate NOE assignments; and contact maps generated from assigned NOESY data to validate candidate assignments. Having generated an initial fold using these processes, AutoStructure uses iterative structure-generation calculations to assign additional NOEs using rules based on the distances between the pairs of protons corresponding to the candidate assignments. Consistently violated assignments based on ambiguous NOEs are identified in cycles of the iterative structure calculations and, where appropriate, reassigned. Figure 1 shows an example of an AutoStructure result based on NOESY data, scalar couplings and a list of hydrogen-bond pairs for the 155-residue protein basic fibroblast growth factor. Similar results have been obtained for a few other proteins using AutoStructure and NMR data.

Conclusions

Recent developments enable the automated analysis of NMR assignments and 3D structures for proteins ranging from about 50 to 200 amino acids. Although progress over the past few years is encouraging, more work is required, even for small proteins, before automated structural analysis is routine. In particular, general methods for the automated analysis of sidechain resonance assignments are not yet well developed and there are, as yet, no examples of completely automated protein assignments and structural analyses. Moreover, little work has focused on the specific problems associated with nucleic acid structures. On the other hand, when good quality data are available, automated analysis of protein NMR data can be very rapid. Many of the resonance assignment programs execute in tens of seconds [7••] and automated structure refinements are being carried out in tens of minutes using arrays of Pentium processors for course-grain parallel calculations (R Tejero, Y Huang, GT Montelione, unpublished data).

Future directions for the automated analysis of assignments involve the development of approaches that require fewer spectra for robust analysis and interactive methods [38] for visually validating and editing the results of automated analysis engines. The high speed of the automated structural analyses suggests that it will be possible to incorporate complete relaxation matrix calculations for the improved calibration of distance constraints, energy refinement and ensemble-averaging calculations as optional features of the automated refinement process. Residual dipolar coupling data should also greatly enhance the efficiency of the automated analysis process [39], particularly in the initial determination of the protein fold, and also provide higher resolution structures [39,40]. Overall, one can expect that, in the next few years, automated analysis of small-protein resonance assignments and 3D structures will become routine in many laboratories and will contribute significantly to the fields of molecular biophysics, structural biology and structural genomics [41].

Acknowledgements

We thank Roberto Tejero and Yuanpeng Huang for useful discussions and Rebecca Watson for her expert assistance in scientific editing. The authors' work on automated analysis is supported by grants from the National Institutes of Health (GM-47014), the National Science Foundation (MCB-9407569 and MCB-9357526) and by a New Jersey Commission on Science and Technology Research Excellence Award.

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. • Koradi R, Billeter M, Engeli M, Güntert P, Wüthrich K: **Automated peak picking and peak integration in macromolecular NMR spectra using AUTOPSY.** *J Magn Reson* 1998, **135**:288-297.
An excellent description of the problems associated with automating peak picking and some methods for their solution.

2. Schulte A-C, Görler A, Antz C, Neidig K-P, Kalbitzer HR: **Use of global symmetries in automated signal class recognition by a Bayesian method.** *J Magn Reson* 1997, **129**:165-172.

3. • Bartels C, Güntert P, Billeter M, Wüthrich K: **GARANT — a general algorithm for resonance assignment of multidimensional nuclear magnetic resonance spectra.** *J Comput Chem* 1997, **18**:139-149.
A powerful approach using genetic algorithms for the automatic analysis of resonance and NOESY cross-peak assignments is described.

4. • Buchler NEG, Zuiderweg ERP, Wang H, Goldstein RA: **Protein heteronuclear NMR assignments using mean-field simulated annealing.** *J Magn Reson* 1997, **125**:34-42.
A powerful resonance assignment program using optimization methods is described. The paper provides a detailed description of the problems associated with the rough pseudo-energy functions used in simulated annealing methods for mapping. It also describes the mean-field method for smoothing pseudo-energy functions.

5. • Li K-B, Sanctuary BC: **Automated resonance assignment of proteins using heteronuclear 3D NMR.** *J Chem Inform Comput Sci* 1997, **37**:467-477.
This paper describes different experimental strategies for sidechain assignments and linking them to the backbone. It also describes a constraint-partitioning algorithm applied to the mapping.

6. •• Lukin JA, Grove AP, Talukdar SN, Ho C: **Automated probabilistic method for assigning backbone resonances of (^{13}C,^{15}N)-labeled proteins.** *J Biomol NMR* 1997, **9**:151-166.
The authors present a powerful resonance assignment strategy using Bayesian statistics methods for both referencing between spectra and amino acid typing.

7. Zimmerman DE, Kulikowski CA, Huang Y, Feng W, Tashiro M,
•• Shimotakahara S, Chien C-Y, Powers R, Montelione GT: **Automated analysis of protein NMR assignments using methods from artificial intelligence.** *J Mol Biol* 1997, **269**:592-610.
This paper describes the algorithms and methodology of the AutoAssign program for determining backbone resonance assignments. This includes a description of a Bayesian statistics method for amino acid typing, a constraint-propagation algorithm for mapping and test results on several real protein data sets.

8. Leutner M, Gschwind RM, Liermann J, Schwarz C, Gemmecker G,
• Kessler H: **Automated backbone assignment of labeled proteins using the threshold accepting algorithm.** *J Biomol NMR* 1998, **11**:31-43.
A description of the PASTA program for resonance assignment and a threshold-accepting algorithm for improving simulated annealing as applied to mapping.

9. Montelione GT, Rios CB, Swapna GVT, Zimmerman DE: **NMR pulse**
• **sequences and computational approaches for automated analysis of sequence-specific backbone resonance assignments of proteins.** In *Biological Magnetic Resonance, Volume 17: Structure Computation and Dynamics in Protein NMR.* Edited by Krishna R, Berliner L. New York: Plenum; 1999:81-130.
A summary of practical considerations in collecting triple-resonance NMR data suitable for automated analysis of resonance assignments.

10. Croft D, Kemmink J, Neidig K-P, Oschkinat H: **Tools for the automated assignment of high-resolution three-dimensional protein NMR spectra based on pattern recognition techniques.** *J Biomol NMR* 1997, **10**:207-219.

11. Gronwald W, Wilard L, Jellard T, Boyko RF, Rajarathnam K, Wishart DS, Sönnichsen FD, Sykes BD: **CAMRA: chemical shift based computer aided protein NMR assignments.** *J Biomol NMR* 1998, **12**:395-405.

12. van Geerestein-Ujah EC, Mariani M, Vis H, Boelens R, Kaptein R: **Use of graph theory for secondary structure recognition and sequential assignment in heteronuclear (^{13}C,^{15}N) NMR spectra: application to HU protein from *Bacillus stearothermophilus.*** *Biopolymers* 1996, **39**:691-701.

13. Huang K, Andrec M, Heald S, Blake P, Prestegard JH: **Performance of a neural-network-based determination of amino acid class and secondary structure from ^{1}H-^{15}N NMR data.** *J Biomol NMR* 1997, **10**:45-52.

14. Szyperski T, Baneck B, Braun D, Glaser RW: **Sequential resonance**
• **assignment of medium-sized ^{15}N/^{13}C-labeled proteins with projected 4D triple resonance NMR experiments.** *J Biomol NMR* 1998, **11**:387-405.
A novel and very powerful approach using reduced-dimensionality triple-resonance experiments for determining resonance assignments and software for the automated analysis of these spectra. This approach has the potential to reduce the time required for collecting data for resonance assignments.

15. Malliavin TE, Pons JL, Delsuc MA: **An NMR assignment module implemented in the Gifa NMR processing program.** *Bioinformatics* 1998, **14**:624-631.

16. Bartels C, Billeter M, Güntert P, Wüthrich K: **Automated sequence-specific NMR assignment of homologous proteins using the program GARANT.** *J Biomol NMR* 1996, **7**:207-213.

17. Tashiro M, Rios CB, Montelione GT: **Classification of amino acid spin systems using PFG HCC(CO)NH-TOCSY with constant-time aliphatic ^{13}C frequency labeling.** *J Biomol NMR* 1995, **6**:211-216.

18. Feng W, Rios CB, Montelione GT: **Phase labeling of C–H and C–C spin system topologies: application in PFG-HACANH and PFG-HACA(CO)NH triple resonance experiments for determining backbone resonance assignments in proteins.** *J Biomol NMR* 1996, **8**:98-104.

19. Rios CB, Feng W, Tashiro M, Shang Z, Montelione GT: **Phase labeling of C–H and C–C spin-system topologies: application in constant-time PFG-CBCA(CO)NH experiments for discriminating amino acid spin-system types.** *J Biomol NMR* 1996, **8**:345-350.

20. Dötsch V, Matsuo K, Wagner G: **Amino-acid-type identification for deuterated proteins with a β-carbon-edited HNCOCACB experiment.** *J Magn Reson B* 1996, **112**:95-100.

21. Dötsch V, Wagner G: **Editing for amino-acid type in CBCACONH experiments based on the ^{13}Cβ-^{13}Cγ coupling.** *J Magn Reson B* 1996, **111**:310-313.

22. Iwadate M, Asakura T, Williamson MP: **Cα and Cβ carbon-13 chemical shifts in proteins from an empirical database.** *J Biomol NMR* 1999, **13**:199-211.

23. O'Connell JF, Pryor KAD, Grant SK, Leiting B: **A high quality nuclear**
• **magnetic resonance solution structure of peptide deformylase from *Escherichia coli*: application of an automated assignment strategy using GARANT.** *J Biomol NMR* 1999, **13**:311-324.
This paper describes a largely automated structure determination using automated analysis of resonance assignments and some NOESY cross-peak assignments using GARANT.

24. Gippert GP, Wright PE, Case DA: **Distributed torsion angle grid**
• **search in high dimensions: a systematic approach to NMR structure determination.** *J Biomol NMR* 1998, **11**:241-263.
The authors describe very general and efficient approaches for the automatic analysis of dihedral-angle constraints from NMR data.

25. 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.
This paper describes the TALOS program for the automatic analysis of dihedral-angle constraints. The program is based on a database of chemical shift data for proteins with high-resolution crystal structures.

26. Beger RD, Bolton PH: **Protein φ and ψ dihedral restraints determined from multidimensional hypersurface correlations of backbone chemical shifts and their use in the determination of protein tertiary structures.** *J Biomol NMR* 1997, **10**:129-142.

27. Mumenthaler C, Güntert P, Braun W, Wüthrich K: **Automated**
• **combined assignment of NOESY spectra and three-dimensional protein structure determination.** *J Biomol NMR* 1997, **10**:351-362.
A description of the NOAH/DYANA approach for the automatic analysis of protein structures from NOE data. Together with the results described in [30••], this represents one of the most successful approaches to automated analysis of NOESY data described in the literature to date.

28. 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.

29. Güntert P, Mumenthaler C, Wüthrich K: **Torsion angle dynamics for NMR structure calculation with the new program DYANA.** *J Mol Biol* 1997, **273**:283-298.

30. Xu Y, Wu J, Gorenstein D, Braun W: **Automated 2D NOESY**
•• **assignment and structure calculation of crambin (S22/I25) with the self-correcting distance geometry based NOAH/DIAMOD programs.** *J Magn Reson* 1999, **136**:76-85.
The application of the NOAH/DIAMOD approach for the automatic analysis of NOESY spectra and the determination of the 3D structure of crambin (S22/I25). Together with the results described in [27•], this represents one of the most successful approaches to automated analysis of NOESY data described in the literature to date.

31. Nilges M, Macias MJ, O'Donoghue SI, Oschkinat H: **Automated**
•• **NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the Pleckstrin homology domain from β-spectrin.** *J Mol Biol* 1997, **269**:408-422.
This paper provides a description of the ARIA approach for using ambiguous constraints in structure calculations with the X-PLOR program, together with tests and applications to real data. ARIA has been used successfully in several experimental NMR structure determinations.

32. Brünger AT: *X-PLOR, version 3.1: a System for X-ray Crystallography and NMR.* New Haven, CT: Yale University Press; 1992.

33. Badger J, Kumar RA, Szalma S: **New features and enhancements in the X-PLOR computer program.** *Proteins* 1999, **35**:25-33.

34. Fraternali F, Amodeo P, Musco G, Nilges M, Pastore A: **Exploring protein interiors: the role of a buried histidine in the KH module fold.** *Proteins* 1999, **34**:484-496.

35. Mizoue LS, Bazan JF, Johnson EC, Handel TM: **Solution structure and dynamics of the CX_3C chemokine domain of fractalkine and its interaction with an N-terminal fragment of CX_3CR1.** *Biochemistry* 1999, **38**:1402-1414.

36. Luh FY, Archer SJ, Domaille PJ, Smith BO, Owen D, Brotherton DH, Raine AR, Xu X, Brizuela L, Brenner SL, Laue ED: **Structure of the cyclin-dependent kinase inhibitor p19Ink4d.** *Nature* 1997, **389**:999-1003.

37. Bassolino-Klimas D, Tejero R, Krystek SR, Metzler WJ, Montelione GT, Bruccoleri RE: **Simulated annealing with restrained molecular dynamics using a flexible restraint potential: theory and evaluation with simulated NMR constraints.** *Protein Sci* 1996, **5**:593-603.

38. Görler A, Gronwald W, Neidig K-P, Kalbitzer HR: **Computer assisted assignment of ^{13}C or ^{15}N edited 3D-NOESY – HSQC spectra using back calculated and experimental spectra.** *J Magn Reson* 1999, **137**:39-45.

39. Prestegard JH: **New techniques in structural NMR-anisotropic interactions.** *Nat Struct Biol* 1998, **5(NMR supplement)**:517-522.

40. Tjandra N, Omichinski JG, Gronenborn AM, Clore GM, Bax A: **Use of dipolar ^{1}H-^{15}N and ^{1}H-^{13}C couplings in the structure determination of magnetically oriented macromolecules in solution.** *Nat Struct Biol* 1997, **4**:732-738.

41. Montelione GT, Anderson S: **Structural genomics: a keystone for a human proteome project.** *Nat Struct Biol* 1999, **6**:11-12.

42. Moy FJ, Seddon AP, Böhlen P, Powers R: **High-resolution solution structure of basic fibroblast growth factor determined by multidimensional heteronuclear magnetic resonance spectroscopy.** *Biochemistry* 1996, **35**:13551-13561.