REVIEW

Automated structure determination from NMR spectra

Peter Güntert

Abstract Automated methods for protein structure determination by NMR have increasingly gained acceptance and are now widely used for the automated assignment of distance restraints and the calculation of three-dimensional structures. This review gives an overview of the techniques for automated protein structure analysis by NMR, including both NOE-based approaches and methods relying on other experimental data such as residual dipolar couplings and chemical shifts, and presents the FLYA algorithm for the fully automated NMR structure determination of proteins that is suitable to substitute all manual spectra analysis and thus overcomes a major efficiency limitation of the NMR method for protein structure determination.

Keywords Protein structure · NMR assignment · Automated assignment · Chemical shift · FLYA algorithm

Introduction

When the NMR method for protein structure determination in solution was introduced in the early 1980s, all analysis of the two-dimensional (2D) spectra was done manually with the help of large paper plots. The memoirs of pioneers in volume 41, issue S1 of Magnetic Resonance in Chemistry afford a vivid picture of this period. Rulers were used to check the frequency alignment of peaks; assignments and other information were stored in hand-written notebooks or marked on the spectra. Only the initial and the final step of the analysis were in the domain of computers: the processing of the raw NMR data by Fourier transformation and the actual calculation of the 3D structure, after initial attempts by interactive model building guided by the NMR data had been unsuccessful. The manual spectra analysis required many months or even years of work by an experienced spectroscopist to solve the structure of a small protein. Gradually the situation has changed over the years. Tools that facilitate the interactive assignment procedures have been introduced that make use of computer graphics and allow to store and manage the relevant data on the computer (Bartels et al. 1995; Delaglio et al. 1995; Eccles et al. 1991; Goddard and Kneller 2001; Johnson and Blevins 1994; Keller 2004; Kobayashi et al. 2007; Kraulis 1989; Neidig et al. 1995). Since the beginning of NMR structure determination it was expected and promised that steps of the spectra analysis can be automated. Soon automated algorithms for peak picking and partial assignment of the chemical shifts appeared, but were not widely used in practice. On the other hand, procedures for automated NOESY assignment proved sufficiently robust to widely replace the earlier manual approach (Herrmann et al. 2002b; Nilges et al. 1997).

The complete automation of protein structure determination is one of the challenges of biomolecular NMR spectroscopy that has, despite of early optimism (Pfändler et al. 1985), proved difficult to achieve. The unavoidable imperfections of experimental NMR spectra and the

P. Güntert (✉)
Institute of Biophysical Chemistry,
Goethe-University Frankfurt am Main,
Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany
e-mail: guentert@em.uni-frankfurt.de

P. Güntert
Frankfurt Institute for Advanced Studies,
Ruth-Moufang-Str. 1, 60438 Frankfurt am Main, Germany

P. Güntert
Graduate School of Science, Tokyo Metropolitan University,
1-1 Minami-Osawa, Hachioji, Tokyo 192-0397, Japan

intrinsic ambiguity of peak assignments that results from the limited accuracy of frequency measurements turn the tractable problem of finding the chemical shift assignments from ideal spectra into a formidably difficult one under realistic conditions. A variety of automated algorithms tackling different parts of NMR protein structure analysis have been developed and reviewed (Altieri and Byrd 2004; Baran et al. 2004; Gronwald and Kalbitzer 2004). However, only recently a purely computational algorithm has been published that is capable of determining the 3D structure of proteins on the basis of uninterpreted spectra (López-Méndez and Güntert 2006).

Fully automated NMR structure determination is more demanding than automating individual parts of NMR structure analysis because the cumulative effect of imperfections at successive steps can easily render the overall process unsuccessful. For example, it has been demonstrated that reliable automated NOE assignment and structure calculation requires around 90% completeness of the chemical shift assignment (Herrmann et al. 2002b; Jee and Güntert 2003), which is not straightforward to achieve by unattended automated peak picking and automated resonance assignment algorithms. Present systems designed to handle the whole process therefore generally require certain human interventions (Gronwald and Kalbitzer 2004; Huang et al. 2005). The interactive validation of peaks and assignments, however, still constitutes a time-consuming obstacle for high-throughput NMR protein structure determination. The crucial indicator for a fully automated NMR structure determination method is the accuracy of the resulting 3D structures when real experimental input data is used and any human interventions at intermediate steps are avoided. Even "small" manual corrections, or the use of idealized input data, can lead to substantially altered conclusions, and prejudice the assessment of different methods.

This review comprises three parts. (1) An overview of the "classical" approach to automated protein structure analysis by NMR that consists of replacing manual steps in NOESY based NMR structure determination by automated algorithms. (2) A survey of alternative approaches that do either not require chemical shift assignments or rely on other data than NOEs to define the 3D structure. (3) A presentation of the FLYA algorithm for the fully automated NMR structure determination of proteins.

Automated spectrum analysis algorithms

The NMR structure determination of a protein conventionally involves the preparation of (typically uniformly ^{13}C/^{15}N-labeled) soluble protein, the acquisition of a set of 2D and 3D NMR experiments, NMR data processing, peak picking, chemical shift assignment, NOE assignment and

Peak picking	→	Signal frequencies
Shift assignments	→	Chemical shifts
NOESY assignment	→	Structural restraints
Structure calculation	→	3D structure
Refinement, validation	→	Final structure

Fig. 1 Steps of a NMR protein structure determination and their resulting data

collection of conformational restraints, structure calculation, refinement and validation (Fig. 1). Virtually all of the more than 5000 NMR protein structures in the Protein Data Bank (Berman et al. 2000) have been determined by this approach.

A variety of computational approaches have been introduced to provide automation for specific parts of an NMR structure determination. A recent review documents close to 100 such algorithms and programs (Gronwald and Kalbitzer 2004). Automated procedures are widely accepted for the assignment of NOE distance restraints and the structure calculations. The automation of the preceding steps of peak picking and resonance assignment has also been the subject of intensive research. Nevertheless, manual or semi-automated approaches still prevail, especially for the assignment of the side-chain chemical shifts. This chapter gives an overview of the algorithms used for different tasks in "classical" NMR protein structure determination.

Automated peak picking

The identification of the NMR signals in two- and higher-dimensional spectra, often referred to as "peak picking", is the first step in the analysis of NMR spectra. Guided by the ongoing assignment process, an experienced spectroscopist can often identify crucial peaks with virtual certainty and, if necessary, make an assignment on the basis of a single, uniquely identified peak. Automated approaches to NMR spectra analysis on the other hand generally have to cope with a lower reliability of peak identification than a spectroscopist who visually inspects the spectra. In compensation, the operation of automated methods can be enhanced by redundancy, e.g. the availability of multiple peaks for a given atom. This can be achieved by recording a set of spectra that provide complementary information for the assignment of a given atom or group of atoms, such that the algorithm can determine their resonance assignment from various pieces of data without relying on the certain identification of any specific peak (Bartels et al. 1997). A variety of algorithms for automated peak picking have been developed, relying on

rule based feature recognition (Antz et al. 1995; Dancea and Günther 2005; Garrett et al. 1991; Herrmann et al. 2002a; Huang et al. 2005; Johnson 2004; Kleywegt et al. 1990; Koradi et al. 1998; Moseley et al. 2004; Rouh et al. 1994), neural networks (Carrara et al. 1993; Corne et al. 1992), antiphase fine structure pattern detection (Meier et al. 1984; Neidig et al. 1990; Pfändler et al. 1985), etc. Nevertheless, even sophisticated recognition methods often fail for complex spectra, mainly because of strong peak overlap, noise, and artifacts such as spurious signals, baseline distortions, and phase distortions. A weakness of many automated approaches is the fact that they analyze only the data points around a local maximum that is part of a potential peak. When interpreting spectra manually, an experienced spectroscopist will make use also of non-local information. In this context it is important that multidimensional spectra typically contain multiple peaks that have the same line shape and the same chemical shift in one frequency domain.

The program NMRView includes a representative, often used example of a simple and rapid automated algorithm for locating peaks that is robust in the absence of overlap (Johnson 2004). Peaks are considered points of local maxima, i.e. points that have a higher intensity than all adjacent points. When NMRView locates peaks, it also identifies the peak bounds, i.e. the width of the peak at the level of the intensity threshold, estimates the peak width at half-height, determines whether the peak is on the edge of the spectrum or adjacent to other peaks, and calculates the center position by interpolating the intensities of the adjacent data points.

The program AUTOPSY is an example of a sophisticated algorithm for automated peak picking of multi-dimensional protein NMR spectra with overlapping peaks (Koradi et al. 1998). The main elements of this program are a function for local noise level calculation, the use of symmetry considerations, and the use of line shapes extracted from well-separated peaks for resolving groups of overlapping peaks. The algorithm generates lists with the frequency positions and integrals of peaks, and a reliability measure for the recognition of each peak.

Automated chemical shift assignment

In de novo 3D structure determinations of proteins by NMR, the key conformational data are upper distance limits derived from nuclear Overhauser effects (NOEs) (Kumar et al. 1980; Macura and Ernst 1980; Neuhaus and Williamson 1989; Solomon 1955). In order to extract distance restraints from a NOESY spectrum, its cross peaks have to be assigned, i.e. the pairs of interacting hydrogen atoms have to be identified. The assignment of NOESY cross peaks requires as a prerequisite the knowledge of the chemical shifts of the spins from which NOEs are arising.

Aside from structure determinations, chemical shift assignments represent crucial information in protein NMR studies on dynamics or binding, for instance in NMR-based ligand screening in drug discovery.

There have been several attempts to automate this chemical shift assignment step that has to precede the collection of conformational restraints and the structure calculation. These methods have been reviewed recently (Altieri and Byrd 2004; Baran et al. 2004; Gronwald and Kalbitzer 2004; Moseley and Montelione 1999), and will not be discussed in detail here. Many automated approaches target the question of assigning the backbone and C^{β} chemical shifts, usually on the basis of triple resonance experiments that delineate the protein backbone through one- and two-bond scalar couplings, using exhaustive, heuristic, or data base searches, Monte Carlo, or simulated annealing methods (Andrec and Levy 2002; Atreya et al. 2000, 2002; Bailey-Kellogg et al. 2000, 2005; Bernstein et al. 1993; Bhavesh et al. 2001; Buchler et al. 1997; Chatterjee et al. 2002; Chen et al. 2005; Coggins and Zhou 2003; Friedrichs et al. 1994; Güntert et al. 2000; Hare and Prestegard 1994; Kamisetty et al. 2006; Kjaer et al. 1994; Leutner et al. 1998; Li and Sanctuary 1997a; Lin et al. 2005; Lukin et al. 1997; Masse and Keller 2005; Moseley et al. 2001; Olson and Markley 1994; Vitek et al. 2005, 2006; Volk et al. 2008; Wang et al. 2005; Wu et al. 2006; Xu et al. 2002, 2006; Zimmerman et al. 1997). Others algorithms are concerned with the more demanding problem of assigning the backbone *and* side-chain chemical shifts (Bartels et al. 1996, 1997; Choy et al. 1997; Croft et al. 1997; Eghbalnia et al. 2005; Gronwald et al. 1998; Hitchens et al. 2003; Li and Sanctuary 1997b; Masse et al. 2006; Pristovšek et al. 2002; Xu et al. 1993, 1994). In most cases, these algorithms require peak lists from a specific set of NMR spectra as input, and produce lists of chemical shifts of varying completeness and correctness, depending on the quality and information content of the input data and the capabilities of the algorithm.

One of the most general and often used chemical shift assignment algorithms is the program GARANT (Bartels et al. 1996, 1997). It has three principal elements. The first is the representation of resonance assignments as an optimal match between experimentally observed peaks and peaks expected based on the amino acid sequence and the magnetization transfer pathways in the spectra used (Fig. 2). Any set of 2D, 3D and 4D homonuclear and heteronuclear NMR spectra can be used. A main advantage of the GARANT algorithm is its ability to analyze the peak lists from all available spectra simultaneously, e.g. to simultaneously assign the backbone and side-chain resonances. The second key element is a scoring function that evaluates the match between observed and expected peaks in order to distinguish between correct and incorrect
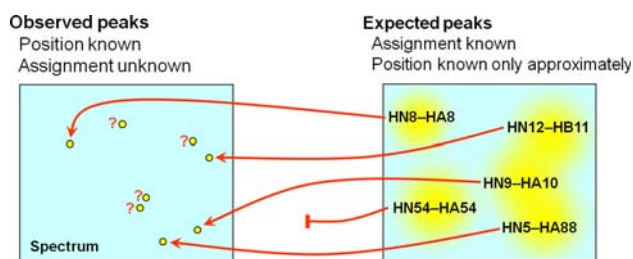
Fig. 2 Scheme of automated chemical shift assignment with the program GARANT

resonance assignments. The score captures the essential features of a correct resonance assignment, i.e. the presence of expected peaks in the spectra, the positional alignment of peaks that originate from the same atoms and the statistical agreement of the assigned resonance frequencies with a chemical shift data base compiled from the known resonance assignments of many proteins. The third key element is the optimization of the score by an evolutionary algorithm combined with a local optimization routine. GARANT is an important part of the FLYA algorithm for fully automated NMR structure analysis, described below.

Automated NOE assignment

Obtaining a comprehensive set of distance restraints from a NOESY spectrum is in practice by no means straightforward. Resonance and peak overlap turn NOE assignment into an iterative process in which preliminary structures, calculated from limited numbers of distance restraints, serve to reduce the ambiguity of the cross peak assignments. Additional difficulties may arise from spectral artifacts and noise, and from the absence of expected signals because of fast relaxation. These inevitable shortcomings of NMR data collection are the main reason why laborious interactive procedures have dominated this central step of NMR protein structure determination for a long time. Automated procedures follow the same general scheme as the interactive approach but do not require manual intervention during the assignment/structure calculation cycles. Two main obstacles have to be overcome by an automated method starting without any prior knowledge of the structure: First, the number of cross peaks with unique assignment based on chemical shift alignment alone is in general not sufficient to define the fold of the protein (Güntert 2003). An automated method must therefore have the capability to use also NOESY cross peaks that cannot (yet) be assigned unambiguously. Second, the automated program must be able to cope with the erroneously picked or inaccurately positioned peaks and with the incompleteness of the chemical shift assignment of typical experimental data sets. An automated

procedure needs devices to substitute for the intuitive decisions made by an experienced spectroscopist in dealing with the imperfections of experimental NMR data.

Besides semi-automatic approaches (Duggan et al. 2001; Güntert et al. 1993; Meadows et al. 1994), several algorithms have been developed for the automated analysis of NOESY spectra given the chemical shift assignments of the backbone and side-chain resonances, namely NOAH (Mumenthaler and Braun 1995; Mumenthaler et al. 1997), ARIA (Habeck et al. 2004; Linge et al. 2003a; Nilges et al. 1997; Rieping et al. 2007), AUTOSTRUCTURE (Huang et al. 2006), KNOWNOE (Gronwald et al. 2002), CANDID (Herrmann et al. 2002b) and a similar algorithm implemented in CYANA (Güntert 2004), PASD (Kuszewski et al. 2004), and a Bayesian approach (Hung and Samudrala 2006). Automated NOE assignment algorithms generally require a high degree of completeness of the backbone and side-chain chemical shift assignments (Jee and Güntert 2003).

Ambiguous distance restraints (Nilges 1995) provide a powerful concept for handling ambiguities in the initial, chemical shift-based NOESY cross peak assignments. Prior to the introduction of ambiguous distance restraints in the ARIA algorithm (Nilges et al. 1997), in general only unambiguously assigned NOEs could be used as distance restraints in the structure calculation. Since the majority of NOEs cannot be assigned unambiguously from chemical shift information alone, this lack of a general way to include ambiguous data into the structure calculation considerably hampered the performance of early automatic NOESY assignment algorithms. When using ambiguous distance restraints, every NOESY cross peak is treated as the superposition of the signals from each of its possible assignments by applying relative weights proportional to the inverse sixth power of the corresponding interatomic distances. A NOESY cross peak with a unique assignment possibility gives rise to an upper bound b on the distance $d(\alpha,\beta)$ between two hydrogen atoms, α and β. A NOESY cross peak with $n > 1$ assignment possibilities can be interpreted as the superposition of n degenerate signals and interpreted as an ambiguous distance restraint, $d_{\text{eff}} < b$, with the "effective" or "r^{-6}-summed" distance

$$d_{\text{eff}} = \left(\sum_{k=1}^{n} d_k^{-6} \right)^{-1/6}.$$

Each of the distances $d_k = d(\alpha_k, \beta_k)$ in the sum corresponds to one assignment possibility to a pair of hydrogen atoms, α_k and β_k. The effective distance d_{eff} is always shorter than any of the individual distances d_k. Thus, an ambiguous distance restraint will be fulfilled by the correct structure provided that the correct assignment is included among its assignment possibilities, regardless of the possible

presence of other, incorrect assignment possibilities. Ambiguous distance restraints make it possible to interpret NOESY cross peaks as correct conformational restraints also if a unique assignment cannot be determined at the outset of a structure determination. Including multiple assignment possibilities, some but not all of which may later turn out to be incorrect, does not result in a distorted structure but only in a decrease of the information content of the ambiguous distance restraints.

Combined automated NOE assignment and structure calculation with CYANA

A widely used algorithm for the automated interpretation of NOESY spectra is implemented in the NMR structure calculation program CYANA (Güntert 2004; Güntert et al. 1997). This algorithm is a re-implementation of the former CANDID algorithm (Herrmann et al. 2002b) on the basis of a probabilistic treatment of the NOE assignment, combined in an iterative process that comprises seven cycles of automated NOE assignment and structure calculation, followed by a final structure calculation using only unambiguously assigned distance restraints. Between subsequent cycles, information is transferred exclusively through the intermediary 3D structures. The molecular structure obtained in a given cycle is used to guide the NOE assignments in the following cycle. Otherwise, the same input data are used for all cycles, that is, the amino acid sequence of the protein, one or several chemical shift lists from the sequence-specific resonance assignment, and one or several lists containing the positions and volumes of cross peaks in 2D, 3D or 4D NOESY spectra. The input may further include previously assigned NOE upper distance bounds or other previously assigned conformational restraints for the structure calculation.

In each cycle, first all assignment possibilities of a peak are generated on the basis of the chemical shift values that match the peak position within given tolerance values, and the quality of the fit is expressed by a Gaussian probability, P_{shifts}. Second, in all but the first cycle the probability $P_{\text{structure}}$ for agreement with the preliminary structure from the preceding cycle, represented by a bundle of conformers, is computed as the fraction of the conformers in which the corresponding distance is shorter than the upper distance bound plus the acceptable distance restraint violation cutoff. Assignment possibilities for which the product of these two probabilities is below the required probability threshold are discarded. Third, each remaining assignment possibility is evaluated for its network anchoring, i.e. its embedding in the network formed by the assignment possibilities of all the other peaks and the covalently restricted short-range distances. The network anchoring probability P_{network} that the distance corresponding to an assignment

possibility is shorter than the upper distance bound plus the acceptable violation is computed given the assignments of the other peaks but independent from knowledge of the 3D structure. Contributions to the network anchoring probability for a given, "current" assignment possibility result from other peaks with the same assignment, from pairs of peaks that connect indirectly the two atoms of the current assignment possibility via a third atom, and from peaks that connect an atom in the vicinity of the first atom of the current assignment with an atom in the vicinity of the second atom of the current assignment. Short-range distances that are constrained by the covalent geometry take, for network anchoring, the same role as an unambiguously assigned NOE. Individual contributions to the network anchoring of the current assignment possibility are expressed as probabilities, P_1, P_2, …, that the distance corresponding to the current assignment possibility satisfies the upper distance bound. The network anchoring probability is obtained from the individual probabilities as $P_{\text{network}} = 1 - (1 - P_1)(1 - P_2)\cdots$, which is never smaller than the highest probability of an individual network anchoring contribution. Only assignment possibilities for which the product of the three probabilities is above a threshold,

$$P_{\text{tot}} = P_{\text{shifts}} \cdot P_{\text{structure}} \cdot P_{\text{network}} \geq P_{\text{min}}$$

are accepted (Fig. 3). Cross peaks with a single accepted assignment yield a conventional unambiguous distance restraint. Otherwise, an ambiguous distance restraint is generated that embodies multiple accepted assignments.

In practice, spurious distance restraints may arise from the misinterpretation of noise and spectral artifacts, in particular at the outset of a structure determination, before 3D structure-based filtering of the restraint assignments can be applied. The key technique used in CYANA to reduce structural distortions from erroneous distance restraints is "constraint combination" (Herrmann et al. 2002b). Ambiguous distance restraints are generated with combined assignments from different, in general unrelated, cross peaks (Fig. 4). The basic property of ambiguous distance restraints that the restraint will be fulfilled by the correct structure whenever at least one of its assignments is correct, regardless of the presence of additional, erroneous assignments, then implies that such combined restraints have a lower probability of being erroneous than the corresponding original restraints, provided that the fraction of erroneous original restraints is smaller than 50%. Constraint combination aims at minimizing the impact of such imperfections on the resulting structure at the expense of a temporary loss of information. It is applied to medium- and long-range distance restraints in the first two cycles of combined automated NOE assignment and structure calculation with CYANA.
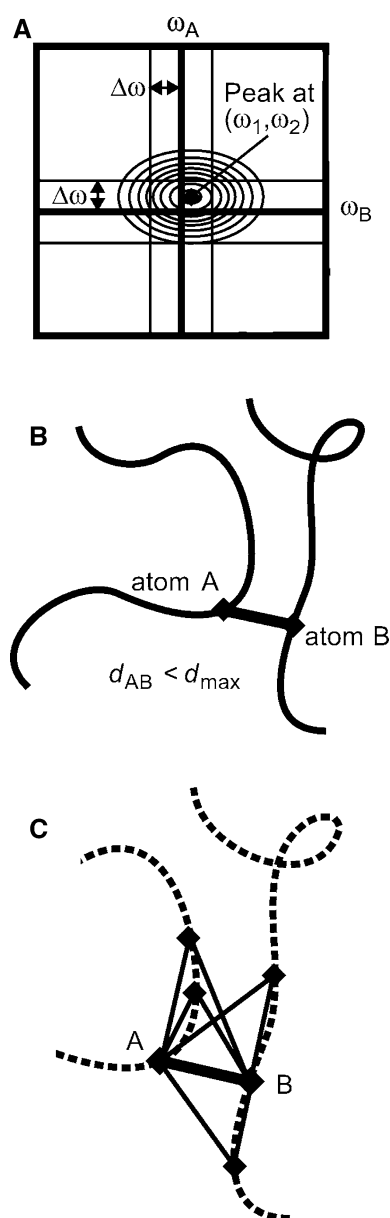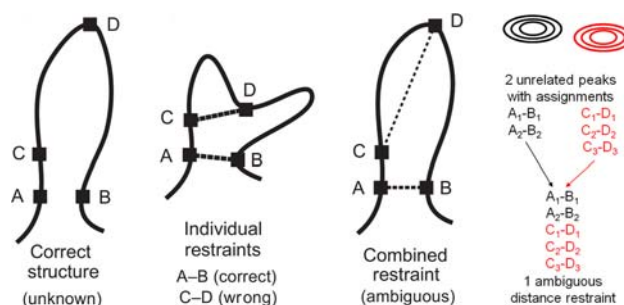
Fig. 4 Schematic illustration of the effect of constraint combination in the case of two distance restraints, a correct one connecting atoms A and B, and a wrong one between atoms C and D. A structure calculation that uses these two restraints as individual restraints that have to be satisfied simultaneously will, instead of finding the correct structure (shown, schematically, in the *first panel*), result in a distorted conformation (*second panel*), whereas a combined restraint that will be fulfilled already if one of the two distances is sufficiently short leads to an almost undistorted solution (*third panel*). The formation of a combined restraint from the assignments of two peaks is shown in the *right panel*

assessments relative to the protein 3D structure from the preceding cycle. The precision of the structure determination normally improves with each subsequent cycle. Accordingly, the cutoff for acceptable distance restraint violations in the calculation of $P_{structure}$ is tightened from cycle to cycle. In the final cycle, an additional filtering step ensures that all NOEs have either unique assignments to a single pair of hydrogen atoms, or are eliminated from the input for the structure calculation. This facilitates the subsequent use of refinement and analysis programs that cannot handle ambiguous distance restraints.

A CYANA structure calculation with automated NOE assignment can be completed in less than one hour for a 10–15 kDa protein, provided that the structure calculations can be performed in parallel, for instance on a Linux cluster system.

Non-classical approaches

Also non-classical approaches that do not rely on sequence-specific resonance assignments and methods using residual dipolar couplings or chemical shifts in conjunction with molecular modeling to determine the backbone structure without the need for side-chain assignments have been proposed.

Assignment-free methods

It is a truth almost universally acknowledged, that a spectroscopist in possession of a good spectrum, must be in want of sequence-specific resonance assignments. However, the chemical shift assignment by itself has no

Fig. 3 Three conditions that must be fulfilled by a valid assignment of a NOESY cross peak to two protons A and B in the CYANA automated NOESY assignment algorithm: **a** agreement between the proton chemical shifts ω_A and ω_B and the peak position (ω_1, ω_2) within a tolerance of $\Delta\omega$. **b** Spatial proximity in a (preliminary) structure. **c** Network anchoring. The NOE between protons A and B must be part of a network of other NOEs or covalently restricted distances that connect the protons A and B indirectly through other protons

The distance restraints are then included in the input for the structure calculation with simulated annealing by the fast CYANA torsion angle dynamics algorithm (Güntert et al. 1997). The structure calculations typically comprise seven cycles. The second and subsequent cycles differ from the first cycle by the use of additional selection criteria for cross peaks and NOE assignments that are based on

biological relevance. It is required only as an intermediate step in the interpretation of the NMR spectra. Consequently, attempts have been made to devise a strategy for NMR protein structure determination that circumvents the chemical shift assignment step. Assignment-free NMR structure calculation methods exploit the fact that NOESY spectra provide distance information even in the absence of chemical shift assignments. This proton-proton distance information is used to calculate a spatial proton distribution. Since there is no association with the covalent structure at this point, the protons of the protein are treated as a cloud of unconnected particles. Provided that the emerging proton distribution is sufficiently clear, a model can then be built into the proton density in a manner analogous to X-ray crystallography where a structural model is placed into the electron density.

This general idea was first tested with simulated NOEs between backbone amide protons of lysozyme (Malliavin et al. 1992), and independently with synthetic NOE data for BPTI (Oshiro and Kuntz 1993). A more thorough treatment using simulated 4D NOESY data for two small proteins with 32 and 58 residues (Kraulis 1994) yielded average 3D real-space ^{1}H spin structures with less than 2 Å RMSD from the previously known structures, and sequence-specific assignments for more than 95% of the spins. Nevertheless, the algorithm has not become a routine tool for NMR structure determination, presumably because the requirements on the quality of the input data are still formidable from the experimental point of view, and because the algorithm had no facilities to deal with overlap among ^{1}H-X chemical shift pairs. In another approach it was proposed to fit structure and chemical shift data directly to NMR spectra rather than peak lists by simultaneously optimizing four variables per atom, three Cartesian coordinates and the chemical shift value (Atkinson and Saudek 1997). The determination of protein structures by NMR without chemical shift assignment is not restricted to NOESY spectra, but can incorporate data from "through-bond" experiments in the form of distances between unassigned and unconnected atoms (Atkinson and Saudek 2002). For instance, a ^{15}N–^{1}H HSQC peak yields a distance equal to the N–H bond length between the two corresponding atoms, and the HNCA spectrum yields, for each N–H pair, four distances to the two adjacent C^α atoms.

The most recent approach to NMR structure determination without chemical shift assignment is the CLOUDS protocol (Grishaev and Llinás 2002a, b) that demonstrated the feasibility of assignment-free structure determination using experimental rather than simulated data. A gas of unassigned, unconnected hydrogen atoms is condensed into a structured proton distribution (cloud) via a molecular dynamics simulated annealing scheme in which the internuclear distances and van der Waals repulsive terms are the

only active restraints. Proton densities are generated by combining a large number of such clouds, each computed from a different trajectory. The primary structure is threaded through the unassigned proton density by a Bayesian approach, for which the probabilities of sequential connectivity hypotheses are inferred from likelihoods of H^N–H^N, H^N–H^α, and H^α–H^α interatomic distances as well as ^{1}H NMR chemical shifts, both derived from public databases. Side chains are placed by a similar procedure.

As for all NMR spectrum analysis, resonance overlap presents a major difficulty also in applying assignment-free strategies. At present, a de novo protein structure determination by the assignment-free approach has not been reported yet, and it remains to be seen whether the assignment-free approach will be able to provide the reliability and the structural quality of the conventional method.

Residual dipolar couplings-based methods

Methods using residual dipolar couplings to determine the backbone structure without the need for side-chain assignments have been developed (Prestegard et al. 2005). In the first approach (Delaglio et al. 2000) the Protein Data Bank is searched for fragments of seven contiguous amino acid residues that fit the measured residual dipolar couplings. From consensus values of the torsion angles for the non-terminal residues of these fragments, an initial structure is built from overlapping fragments by "molecular fragment replacement" (MFR). Errors in the MFR-derived backbone torsion angles accumulate when building the initial model because the long-range information contained in the residual dipolar couplings is not yet used. However, this global orientational information can be reintroduced when using these rough models as starting structures in a subsequent refinement procedure based on a simple iterative gradient approach that adjusts ϕ/ψ to minimize the difference between measured and best-fitted dipolar couplings and between measured chemical shifts and those predicted by the model. It was demonstrated that the 3D structure of large protein backbone segments, and in favorable cases an entire small protein, can be calculated exclusively from dipolar couplings and chemical shifts (Delaglio et al. 2000). This and similar approaches (Rohl and Baker 2002) require assignments of the backbone chemical shifts as input.

In a further step, automated algorithms were developed that simultaneously perform the assignment and the determination of low resolution backbone structures on the basis of unassigned chemical shifts and residual dipolar couplings (Jung et al. 2004; Meiler and Baker 2003). The latter method relies on the de novo protein structure prediction algorithm ROSETTA (Simons et al. 1997) and a

Monte Carlo search for chemical shift assignments that produce the best fit of the experimental NMR data to a candidate 3D structure.

Chemical shift-based structure determination

Chemical shifts are the NMR parameter than can be measured most easily and accurately, and they are highly sensitive to their local environment. They are widely used to monitor conformational changes or ligand binding, and can yield information about specific features of protein conformations, notably dihedral angles (Cornilescu et al. 1999) and secondary structure (Wishart and Sykes 1994). However, the complex relationship between chemical shifts and 3D structure has impeded their direct use for tertiary structure determination. Recently, however, two approaches to 3D protein structure determination have been developed that use exclusively chemical shifts as experimental input data (Cavalli et al. 2007; Shen et al. 2008). Both methods do not rely on the quantum mechanical calculation of chemical shifts from first principles but exploit the availability of an ever growing data base of 3D protein structures (Berman et al. 2000) and corresponding chemical shifts (Seavey et al. 1991) to collect molecular fragment conformations from known protein structures that match the experimentally determined secondary chemical shifts of the protein under study. A secondary chemical shift is the deviation of a chemical shift from the residue-type dependent random coil chemical shift value of the corresponding atom. This separates the conformation dependence of the chemical shift from its residue-type dependence, which is a prerequisite for the sequence independent identification of molecular fragments with similar conformation. The molecular fragment conformations are found by extending the data base search method of the program TALOS (Cornilescu et al. 1999) to contiguous segments of several residues (Cavalli et al. 2007; Shen and Bax 2007). The fragment conformations are then assembled into a 3D structure of the entire protein using molecular modeling approaches.

The CHESHIRE algorithm was the first program to generate near-atomic resolution structures from chemical shifts (Cavalli et al. 2007). It first uses the $^1H^\alpha$, ^{15}N, $^{13}C^\alpha$ and $^{13}C^\beta$ secondary chemical shifts to predict the secondary structure of the protein and the backbone torsion angles, followed by the identification of three- and nine-residue segments on the basis of the secondary chemical shifts, the predicted secondary structure and the predicted backbone dihedral angles. Low resolution structures in which the side chains are represented by a single C^β atom are calculated by a Monte Carlo algorithm using the CHARMM force field (Brooks et al. 1983) complemented with terms for secondary structure packing and cooperative

hydrogen bonding. The previously determined three- and nine-residue fragments guide Monte Carlo moves. All atom conformers are generated. Finally, the 500 best scoring all atom conformers are refined by a Monte Carlo protocol during which an additional energy term is active that describes the correlation between experimental and predicted chemical shifts. The CHESHIRE algorithm yielded the structures of 11 proteins of 46–123 residues with an accuracy of 2 Å or better for the backbone RMSD.

The CS-ROSETTA method is based on the same concept (Shen et al. 2008). It combines the well established ROSETTA structure prediction program (Bradley et al. 2005) with a recently enhanced empirical relation between structure and chemical shifts (Shen and Bax 2007), which allows selection of database fragments that better match the structure of the unknown protein. Generating new protein structures by CS-ROSETTA involves two separate stages. First, polypeptide fragments are selected from a protein structural database, based on the combined use of $^{13}C^\alpha$, $^{13}C^\beta$, $^{13}C'$, ^{15}N, $^1H^\alpha$ and $^1H^N$ chemical shifts and the amino acid sequence pattern. In the second stage, these fragments are used for de novo structure generation, using the standard ROSETTA Monte Carlo assembly and relaxation methods. The method was calibrated using 16 proteins of known structure, and then successfully tested for nine proteins with 65–147 residues under study in a structural genomics project. For these, the CS-ROSETTA algorithm yielded full-atom models with 0.6–2.1 Å RMSD for the backbone atoms relative to the independently determined NMR structures.

Both methods require as experimental input the chemical shift assignments for the backbone and $^{13}C^\beta$ atoms. These shifts are generally available at an early stage of the traditional NMR structure determination process, before the collection and analysis of structural restraints. Side-chain chemical shift assignments beyond C^β, which are considerably harder to obtain than those for the backbone, are not necessary.

It must be noted that in contrast to the NOE-based conventional approach for which a well established theory exists that relates each piece of NMR data (the NOESY peak volume) to a corresponding conformational restraint, chemical shift-based structure determination is an empirical approach that exploits other, previously determined protein structures for solving the current structure of interest by assuming that the entire sequence of the protein can be covered by overlapping fragments that have a similar conformation in the current protein as corresponding stretches in already existing structures. There are no experimentally derived long-range conformational restraints. This implies that the correct tertiary structure has to be found—or may be missed—by the underlying molecular modeling algorithm. In practice, convergence

rapidly decreases with increasing protein size, and the CS-ROSETTA approach starts to fail for proteins larger than 130 residues (Shen et al. 2008). Convergence is also adversely affected by the presence of long, disordered loops.

The FLYA algorithm

Fully automated structure determination of proteins in solution (FLYA) yields, without human intervention, 3D protein structures starting from a set of multidimensional NMR spectra (López-Méndez and Güntert 2006). As in the classical manual approach, structures are determined by a set of experimental NOE distance restraints without reference to already existing structures or empirical molecular modeling information. In addition to the 3D structure of the protein, FLYA yields backbone and side-chain chemical shift assignments, and cross peak assignments for all spectra.

The FLYA algorithm (Fig. 5) uses as input data only the protein sequence and multidimensional NMR spectra. Any combination of commonly used heteronuclear and homonuclear 2D, 3D and 4D NMR spectra can be used as input for the FLYA algorithm, provided that it affords sufficient information for the assignment of the backbone and side-chain chemical shifts and for the collection of conformational restraints. Peaks are identified in the multidimensional NMR spectra using the automated peak picking algorithm of NMRView (Johnson 2004), or AUTOPSY (Koradi et al. 1998). Peak integrals for NOESY cross peaks are determined simultaneously. Since no manual corrections are applied, the resulting raw peak lists may contain, in addition to the entries representing true signals, a significant number of artifacts (see Figs. 2, 4 of López-Méndez and Güntert 2006). The following steps of the fully automated structure determination algorithm can tolerate the presence of such artifacts, as long as the majority of the true peaks have been identified.

Based on the peak positions and, in the case of NOESY spectra, peak volumes peak lists are prepared by CYANA (Güntert 2003; Güntert et al. 1997). Depending on the spectra, the preparation may include unfolding aliased signals, systematic correction of chemical shift referencing, and removal of peaks near the diagonal or water lines. The peak lists resulting from this step remain invariable throughout the rest of the procedure. An ensemble of initial chemical shift assignments is obtained by multiple runs of a modified version of the GARANT algorithm (Bartels et al. 1996, 1997) with different seed values for the random number generator (Malmodin et al. 2003). The original GARANT algorithm was modified for new spectrum types and for the treatment of NOESY spectra when 3D structures are available. In analogy to NMR structure
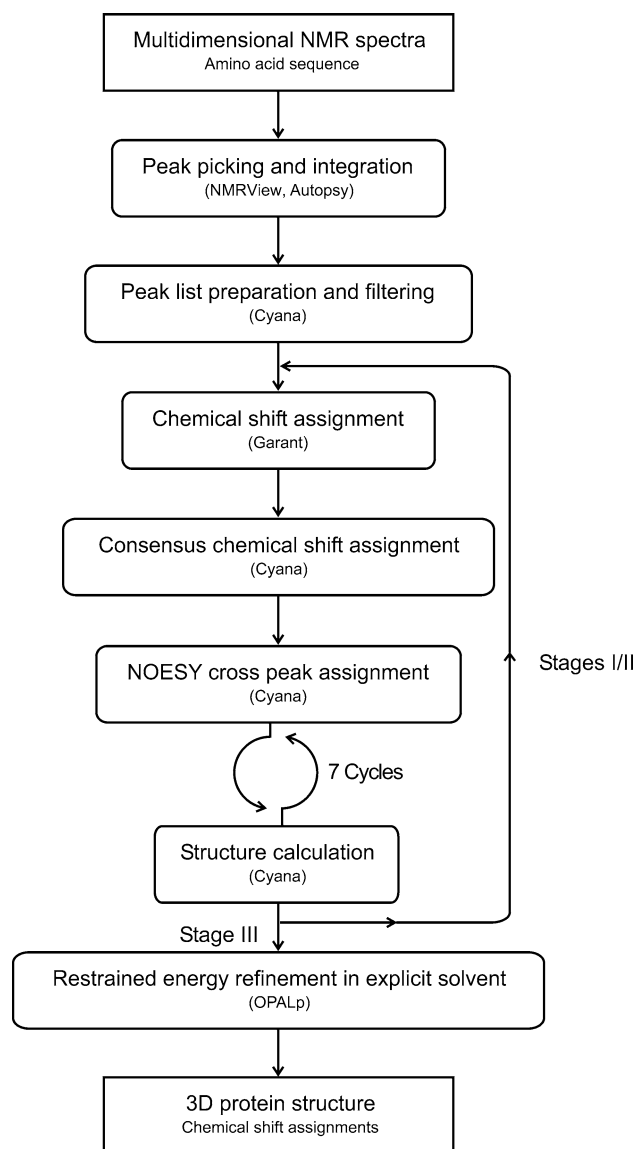

Fig. 5 Flowchart of the fully automated structure determination algorithm FLYA

calculation in which not a single structure but an ensemble of conformers is calculated using identical input data but different randomized start conformers, the initial chemical shift assignment produces an ensemble rather than a single chemical shift value for each ^1H, ^{13}C and ^{15}N nucleus. The peak position tolerance is typically set to 0.03 ppm for the ^1H dimensions and to 0.4 ppm for the ^{13}C and ^{15}N dimensions. These initial chemical shift assignments are consolidated by CYANA into a single consensus chemical shift list. The most highly populated chemical shift value in the ensemble is computed for each ^1H, ^{13}C and ^{15}N spin and selected as the consensus chemical shift value that will be used for the subsequent automated assignment of NOESY peaks. The consensus chemical shift for a given nucleus is the value ω that maximizes the function

$\mu(\omega) = \sum_j \exp\left(-(\omega - \omega_j)^2/2\Delta\omega^2\right)$, where the sum runs over all chemical shift values ω_j for the given nucleus in the ensemble of initial chemical shift assignments, and $\Delta\omega$ denotes the aforementioned chemical shift tolerance. NOESY cross peaks are assigned automatically (Herrmann et al. 2002b) on the basis of the consensus chemical shift assignments and the same peak lists and chemical shift tolerance values used already for the chemical shift assignment. The automated NOE assignment algorithm of the program CYANA is used. The overall probability for the correctness of possible NOE assignments is calculated as the product of three probabilities that reflect the agreement between the chemical shift values and the peak position, the consistency with a preliminary 3D structure (Güntert et al. 1993), and network anchoring (Herrmann et al. 2002b), i.e. the extent of embedding in the network formed by other NOEs. Restraints with multiple possible assignments are represented by ambiguous distance restraints (Nilges 1995). Seven cycles of combined automated NOE assignment and structure calculation by simulated annealing in torsion angle space and a final structure calculation using only unambiguously assigned distance restraints are performed. Constraint combination (Herrmann et al. 2002b) is applied in the first two cycles to all NOE distance restraints spanning at least three residues in order to minimize distortions of the structures by erroneous distance restraints that may result from spurious entries in the peak lists and/or incorrect chemical shift assignments.

A complete FLYA calculation comprises three stages. In the first stage, the chemical shifts and protein structures are generated de novo (stage I). In the next stages (stages II and III), the structures generated by the preceding stage are used as additional input for the determination of chemical shift assignments. Stages II and III are particularly important for aromatics residues and other resonances whose assignment rely on through-space NOESY information. At the end of the third stage, the 20 final CYANA conformers with the lowest target function values are subjected to restrained energy minimizations in explicit solvent against the AMBER force field (Cornell et al. 1995) using the program OPALp (Koradi et al. 2000; Luginbühl et al. 1996). The complete procedure is driven by the NMR structure calculation program CYANA, which is also used for parallelization of all time-consuming steps. The performance of the FLYA algorithm can be monitored at different steps of the procedure by quality measures that can be computed without referring to external reference assignments or structures (López-Méndez and Güntert 2006).

Structure calculations with the FLYA algorithm yielded 3D structures of three 12–16 kDa proteins that coincided closely with the conventionally determined structures (Fig. 6). Deviations were below 0.95 Å for the backbone atom positions, excluding the flexible chain termini, and 96–97% of all backbone and side-chain chemical shifts in the structured regions were assigned to the correct residues. The purely computational FLYA method is thus suitable to substitute all manual spectra analysis and overcomes a major efficiency limitation of the NMR method for protein structure determination.
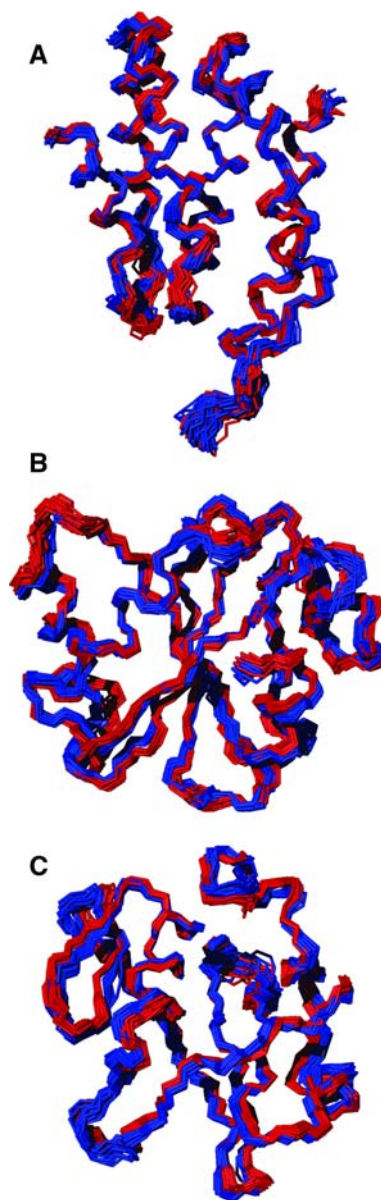

Fig. 6 Structures obtained by fully automated structure determination with the FLYA algorithm (*blue*) superimposed on the corresponding NMR structures determined by conventional methods (*red*). **a** ENTH domain At3g16270(9–135) from *Arabidopsis thaliana* (López-Méndez et al. 2004). **b** Rhodanese homology domain At4g01050(175–295) from *Arabidopsis thaliana* (Pantoja-Uceda et al. 2005). **c** Src homology domain 2 (SH2) from the human feline sarcoma oncogene Fes (Scott et al. 2005)

Various extensions of the basic FLYA algorithm can be envisaged. It is straightforward to further improve the results by interactive improvements of the peak lists, corrections of erroneous chemical shift assignments, and additional conformational restraints for torsion angles, hydrogen bonds, residual dipolar couplings, etc. For large or difficult proteins semiautomatic approaches are possible in which parts of the assignments are provided or confirmed by the user. NMR data processing could be incorporated in FLYA in order to start the procedure from the raw time-domain data from the NMR spectrometer. Alternative peak picking algorithms can be used. Improved performance can in principle be expected from recently developed "projected" NMR experiments (Atreya and Szyperski 2005; Freeman and Kupče 2003) that can yield data corresponding to that from higher-dimensional spectra combined with high-accuracy frequency information, thereby resulting in reduced assignment ambiguity. The currently static peak lists may be replaced by dynamic peak lists that will be updated continuously on the basis of intermediate results (Herrmann et al. 2002a) during a FLYA calculation. An optimized resonance assignment algorithm can reduce the computation time and make more sophisticated use of intermediate 3D structures. Additional refinement techniques can improve the structures with respect to common quality measures (Linge et al. 2003b; Nederveen et al. 2005). The number of input spectra can be reduced for well-behaved proteins.

The latter idea is of particular interest because a considerable amount of NMR measurement time was necessary to record the 13–14 input 3D spectra that were used as input for the aforementioned FLYA structure determinations. The influence of reduced sets of experimental spectra on the quality of NMR structures obtained with FLYA was investigated for the 12 kDa Src homology domain 2 from the human feline sarcoma oncogene Fes (Fes SH2) (Scott et al. 2006). FLYA calculations were performed for 5 reduced data sets selected from the complete set of 13 3D spectra of the earlier conventional structure determination (Scott et al. 2005). The reduced data sets utilized only CBCA(CO)NH and CBCANH for the backbone assignments and either all, some or none of the five original side-chain assignment spectra. In four of the five cases tested, the 3D structures deviated by less than 1.3 Å backbone RMSD from the conventionally determined Fes SH2 reference structure, showing that the FLYA algorithm is remarkably stable and accurate when used with reduced sets of input spectra.

Stereo-array isotope labeling (SAIL) (Kainosho et al. 2006) has been combined with the fully automated NMR structure determination algorithm FLYA (Takeda et al. 2007). SAIL provides a complete stereo and regiospecific pattern of stable isotopes, which yields much sharper resonance lines and reduced signal overlap without loss of information. Automated signal identification can be achieved with higher reliability for the fewer, sharper and more intense peaks of SAIL proteins. The danger of making erroneous assignments decreases with the number of nuclei and peaks to assign, and less spin diffusion allows NOEs to be interpreted more quantitatively. As a result of the superior quality of the SAIL NMR spectra, reliable fully automated analysis of the NMR spectra and structure calculation are possible using fewer input spectra than with conventional uniformly $^{13}C/^{15}N$-labeled proteins. FLYA calculations with SAIL ubiquitin using a single "through-bond" 3D spectrum in addition to the ^{13}C-edited and ^{15}N-edited NOESY spectra for the restraint collection yielded structures with an accuracy of 0.83–1.15 Å for the backbone RMSD to the conventionally determined solution structure (Ikeya et al. 2008), showing the feasibility of fully automated NMR structure analysis from a minimal set of spectra.

Conclusions

Fully automated NMR structure determination of proteins up to 140 amino acid residues is possible now, provided that good quality input spectra are available. Purely computational methods for NMR structure analysis can cope with the amount of overlap and artifacts present in typical experimental NMR spectra. Their combination with optimal stable isotope labeling can enable automated NMR structure determination of proteins with a molecular weight above 20 kDa, for which the large number of chemical shifts and peaks renders the traditional manual analysis method particularly cumbersome and error-prone. For the future, we expect fully automated NMR protein structure determination to replace most manual and semi-automatic approaches and to produce structures of the same quality as by manual spectrum analysis.

References

Altieri AS, Byrd RA (2004) Automation of NMR structure determination of proteins. Curr Opin Struct Biol 14:547–553. doi:10.1016/j.sbi.2004.09.003

Andrec M, Levy RM (2002) Protein sequential resonance assignments by combinatorial enumeration using $^{13}C\alpha$ chemical shifts and their (i, i − 1) sequential connectivities. J Biomol NMR 23:263–270. doi:10.1023/A:1020236105735

Antz C, Neidig KP, Kalbitzer HR (1995) A general Bayesian method for an automated signal class recognition in 2D NMR spectra combined with a multivariate discriminant analysis. J Biomol NMR 5:287–296. doi:10.1007/BF00211755

Atkinson RA, Saudek V (1997) Direct fitting of structure and chemical shift to NMR spectra. J Chem Soc Faraday Trans 93:3319–3323. doi:10.1039/a702834b

Atkinson RA, Saudek V (2002) The direct determination of protein structure by NMR without assignment. FEBS Lett 510:1–4

Atreya HS, Szyperski T (2005) Rapid NMR data collection. Methods Enzymol 394:78–108. doi:10.1016/S0076-6879(05)94004-4

Atreya HS, Sahu SC, Chary KVR, Govil G (2000) A tracked approach for automated NMR assignments in proteins (TATAPRO). J Biomol NMR 17:125–136. doi:10.1023/A:1008315111278

Atreya HS, Chary KVR, Govil G (2002) Automated NMR assignments of proteins for high throughput structure determination: TATAPRO II. Curr Sci 83:1372–1376

Bailey-Kellogg C, Widge A, Kelley JJ, Berardi MJ, Bushweller JH, Donald BR (2000) The NOESY JIGSAW: automated protein secondary structure and main-chain assignment from sparse, unassigned NMR data. J Comput Biol 7:537–558. doi:10.1089/106652700750050934

Bailey-Kellogg C, Chainraj S, Pandurangan G (2005) A random graph approach to NMR sequential assignment. J Comput Biol 12:569–583. doi:10.1089/cmb.2005.12.569

Baran MC, Huang YJ, Moseley HNB, Montelione GT (2004) Automated analysis of protein NMR assignments and structures. Chem Rev 104:3541–3555. doi:10.1021/cr030408p

Bartels C, Xia TH, Billeter M, Güntert P, Wüthrich K (1995) The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. J Biomol NMR 6:1–10. doi:10.1007/BF00417486

Bartels C, Billeter M, Güntert P, Wüthrich K (1996) Automated sequence-specific NMR assignment of homologous proteins using the program GARANT. J Biomol NMR 7:207–213. doi:10.1007/BF00202037

Bartels C, Güntert P, Billeter M, Wüthrich K (1997) GARANT: a general algorithm for resonance assignment of multidimensional nuclear magnetic resonance spectra. J Comput Chem 18:139–149. doi:10.1002/(SICI)1096-987X(19970115)18:1<139::AID-JCC13>3.0.CO;2-H

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H et al (2000) The protein data bank. Nucleic Acids Res 28:235–242. doi:10.1093/nar/28.1.235

Bernstein R, Cieslar C, Ross A, Oschkinat H, Freund J, Holak TA (1993) Computer-assisted assignment of multidimensional NMR spectra of proteins—application to 3D NOESY-HMQC and TOCSY-HMQC Spectra. J Biomol NMR 3:245–251. doi:10.1007/BF00178267

Bhavesh NS, Panchal SC, Hosur RV (2001) An efficient high-throughput resonance assignment procedure for structural genomics and protein folding research by NMR. Biochemistry 40:14727–14735. doi:10.1021/bi015683p

Bradley P, Misura KM, Baker D (2005) Toward high-resolution de novo structure prediction for small proteins. Science 309:1868–1871. doi:10.1126/science.1113801

Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 4:187–217. doi:10.1002/jcc.540040211

Buchler NEG, Zuiderweg ERP, Wang H, Goldstein RA (1997) Protein heteronuclear NMR assignments using mean-field simulated annealing. J Magn Reson 125:34–42. doi:10.1006/jmre.1997.1106

Carrara EA, Pagliari F, Nicolini C (1993) Neural networks for the peak picking of nuclear magnetic resonance spectra. Neural Netw 6:1023–1032

Cavalli A, Salvatella X, Dobson CM, Vendruscolo M (2007) Protein structure determination from NMR chemical shifts. Proc Natl Acad Sci USA 104:9615–9620. doi:10.1073/pnas.0610313104

Chatterjee A, Bhavesh NS, Panchal SC, Hosur RV (2002) A novel protocol based on HN(C)N for rapid resonance assignment in (15N, 13C) labeled proteins: implications to structural genomics. Biochem Biophys Res Commun 293:427–432. doi:10.1016/S0006-291X(02)00240-1

Chen ZZ, Lin GH, Rizzi R, Wen JJ, Xu D, Xu Y et al (2005) More reliable protein NMR peak assignment via improved 2-interval scheduling. J Comput Biol 12:129–146. doi:10.1089/cmb.2005.12.129

Choy WY, Sanctuary BC, Zhu G (1997) Using neural network predicted secondary structure information in automatic protein NMR assignment. J Chem Inf Comput Sci 37:1086–1094. doi:10.1021/ci970012c

Coggins BE, Zhou P (2003) PACES: protein sequential assignment by computer-assisted exhaustive search. J Biomol NMR 26:93–111. doi:10.1023/A:1023589029301

Corne SA, Johnson AP, Fisher J (1992) An artificial neural network for classifying cross peaks in two-dimensional NMR spectra. J Magn Reson 100:256–266

Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM et al (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 117:5179–5197. doi:10.1021/ja00124a002

Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR 13:289–302. doi:10.1023/A:1008392405740

Croft D, Kemmink J, Neidig KP, Oschkinat H (1997) Tools for the automated assignment of high-resolution three-dimensional protein NMR spectra based on pattern recognition techniques. J Biomol NMR 10:207–219. doi:10.1023/A:1018329420659

Dancea F, Günther U (2005) Automated protein NMR structure determination using wavelet de-noised NOESY spectra. J Biomol NMR 33:139–152. doi:10.1007/s10858-005-3093-1

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on Unix pipes. J Biomol NMR 6:277–293. doi:10.1007/BF00197809

Delaglio F, Kontaxis G, Bax A (2000) Protein structure determination using molecular fragment replacement and NMR dipolar couplings. J Am Chem Soc 122:2142–2143. doi:10.1021/ja993603n

Duggan BM, Legge GB, Dyson HJ, Wright PE (2001) SANE (Structure assisted NOE evaluation): an automated model-based approach for NOE assignment. J Biomol NMR 19:321–329. doi:10.1023/A:1011227824104

Eccles C, Güntert P, Billeter M, Wüthrich K (1991) Efficient analysis of protein 2D NMR spectra using the software package EASY. J Biomol NMR 1:111–130. doi:10.1007/BF01877224

Eghbalnia HR, Bahrami A, Wang LY, Assadi A, Markley JL (2005) Probabilistic identification of spin systems and their assignments including coil-helix inference as output (PISTACHIO). J Biomol NMR 32:219–233. doi:10.1007/s10858-005-7944-6

Freeman R, Kupče E (2003) New methods for fast multidimensional NMR. J Biomol NMR 27:101–113. doi:10.1023/A:1024960302926

Friedrichs MS, Mueller L, Wittekind M (1994) An automated procedure for the assignment of protein ^{1}HN, ^{15}N, ^{13}C$^\alpha$, ^{1}H$^\alpha$, ^{13}C$^\beta$ and ^{1}H$^\beta$ resonances. J Biomol NMR 4:703–726. doi:10.1007/BF00404279

Garrett DS, Powers R, Gronenborn AM, Clore GM (1991) A common sense approach to peak picking two-, three- and four-dimensional spectra using automatic computer analysis of contour diagrams. J Magn Reson 95:214–220

Goddard TD, Kneller DG (2001) Sparky 3. University of California, San Francisco

Grishaev A, Llinás M (2002a) CLOUDS, a protocol for deriving a molecular proton density via NMR. Proc Natl Acad Sci USA 99:6707–6712. doi:10.1073/pnas.082114199

Grishaev A, Llinás M (2002b) Protein structure elucidation from NMR proton densities. Proc Natl Acad Sci USA 99:6713–6718. doi:10.1073/pnas.042114399

Gronwald W, Kalbitzer HR (2004) Automated structure determination of proteins by NMR spectroscopy. Prog Nucl Magn Reson Spectrosc 44:33–96. doi:10.1016/j.pnmrs.2003.12.002

Gronwald W, Willard L, Jellard T, Boyko RE, Rajarathnam K, Wishart DS et al (1998) CAMRA: chemical shift based computer aided protein NMR assignments. J Biomol NMR 12:395–405. doi:10.1023/A:1008321629308

Gronwald W, Moussa S, Elsner R, Jung A, Ganslmeier B, Trenner J et al (2002) Automated assignment of NOESY NMR spectra using a knowledge based method (KNOWNOE). J Biomol NMR 23:271–287. doi:10.1023/A:1020279503261

Güntert P (2003) Automated NMR protein structure calculation. Prog Nucl Magn Reson Spectrosc 43:105–125. doi:10.1016/S0079-6565(03)00021-9

Güntert P (2004) Automated NMR structure calculation with CYANA. Methods Mol Biol 278:353–378

Güntert P, Berndt KD, Wüthrich K (1993) The program ASNO for computer-supported collection of NOE upper distance constraints as input for protein structure determination. J Biomol NMR 3:601–606. doi:10.1007/BF00174613

Güntert P, Mumenthaler C, Wüthrich K (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol 273:283–298. doi:10.1006/jmbi.1997.1284

Güntert P, Salzmann M, Braun D, Wüthrich K (2000) Sequence-specific NMR assignment of proteins by global fragment mapping with the program MAPPER. J Biomol NMR 18:129–137. doi:10.1023/A:1008318805889

Habeck M, Rieping W, Linge JP, Nilges M (2004) NOE assignment with ARIA 2.0: the nuts and bolts. Methods Mol Biol 278:379–402

Hare BJ, Prestegard JH (1994) Application of neural networks to automated assignment of NMR spectra of proteins. J Biomol NMR 4:35–46. doi:10.1007/BF00178334

Herrmann T, Güntert P, Wüthrich K (2002a) Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. J Biomol NMR 24:171–189. doi:10.1023/A:1021614115432

Herrmann T, Güntert P, Wüthrich K (2002b) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol 319:209–227. doi:10.1016/S0022-2836(02)00241-3

Hitchens TK, Lukin JA, Zhan YP, McCallum SA, Rule GS (2003) MONTE: an automated Monte Carlo based approach to nuclear magnetic resonance assignment of proteins. J Biomol NMR 25:1–9. doi:10.1023/A:1021975923026

Huang YPJ, Moseley HNB, Baran MC, Arrowsmith C, Powers R, Tejero R et al (2005) An integrated platform for automated analysis of protein NMR structures. Methods Enzymol 394:111–141. doi:10.1016/S0076-6879(05)94005-6

Huang YJ, Tejero R, Powers R, Montelione GT (2006) A topology-constrained distance network algorithm for protein structure determination from NOESY data. Proteins. Struct Funct Bioinform 62:587–603. doi:10.1002/prot.20820

Hung LH, Samudrala R (2006) An automated assignment-free Bayesian approach for accurately identifying proton contacts from NOESY data. J Biomol NMR 36:189–198. doi:10.1007/s10858-006-9082-1

Ikeya T, Yoshida H, Terauchi T, Kainosho M, Güntert P (2008) Automated NMR structure determination of stereo-array isotope labeled ubiquitin from minimal sets of spectra using the SAIL-FLYA system. J Biomol NMR (in press)

Jee J, Güntert P (2003) Influence of the completeness of chemical shift assignments on NMR structures obtained with automated NOE assignment. J Struct Funct Genomics 4:179–189. doi:10.1023/A:1026122726574

Johnson BA (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. Methods Mol Biol 278:313–352

Johnson BA, Blevins RA (1994) NMR View: a computer program for the visualization and analysis of NMR data. J Biomol NMR 4:603–614. doi:10.1007/BF00404272

Jung YS, Sharma M, Zweckstetter M (2004) Simultaneous assignment and structure determination of protein backbones by using NMR dipolar couplings. Angew Chem Int Ed 43:3479–3481. doi:10.1002/anie.200353588

Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Ono AM, Güntert P (2006) Optimal isotope labelling for NMR protein structure determinations. Nature 440:52–57. doi:10.1038/nature04525

Kamisetty H, Bailey-Kellogg C, Pandurangan G (2006) An efficient randomized algorithm for contact-based NMR backbone resonance assignment. Bioinformatics 22:172–180. doi:10.1093/bioinformatics/bti786

Keller RLJ (2004) Optimizing the process of nuclear magnetic resonance spectrum analysis and computer aided resonance assignment, PhD thesis. Institute of Molecular Biology and Biophysics. ETH, Zürich

Kjaer M, Andersen KV, Poulsen FM (1994) Automated and semiautomated analysis of homonuclear and heteronuclear multidimensional nuclear magnetic resonance spectra of proteins—the program PRONTO. Methods Enzymol 239:288–307. doi:10.1016/S0076-6879(94)39010-X

Kleywegt GJ, Boelens R, Kaptein R (1990) A versatile approach toward the partially automatic recognition of cross peaks in 2D ^{1}H NMR spectra. J Magn Reson 88:601–608

Kobayashi N, Iwahara J, Koshiba S, Tomizawa T, Tochio N, Güntert P et al (2007) KUJIRA, a package of integrated modules for systematic and interactive analysis of NMR data directed to high-throughput NMR structure studies. J Biomol NMR 39:31–52. doi:10.1007/s10858-007-9175-5

Koradi R, Billeter M, Engeli M, Güntert P, Wüthrich K (1998) Automated peak picking and peak integration in macromolecular NMR spectra using AUTOPSY. J Magn Reson 135:288–297. doi:10.1006/jmre.1998.1570

Koradi R, Billeter M, Güntert P (2000) Point-centered domain decomposition for parallel molecular dynamics simulation. Comput Phys Commun 124:139–147. doi:10.1016/S0010-4655(99)00436-1

Kraulis PJ (1989) ANSIG: a program for the assignment of protein ^{1}H 2D NMR spectra by interactive graphics. J Magn Reson 84:627–633

Kraulis PJ (1994) Protein three-dimensional structure determination and sequence-specific assignment of 13C-separated and 15N-separated NOE Data - a novel real-space ab-initio approach. J Mol Biol 243:696–718

Kumar A, Ernst RR, Wüthrich K (1980) A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton–proton cross-relaxation networks in biological macromolecules. Biochem Biophys Res Commun 95:1–6. doi:10.1016/0006-291X(80)90695-6

Kuszewski J, Schwieters CD, Garrett DS, Byrd RA, Tjandra N, Clore GM (2004) Completely automated, highly error-tolerant macromolecular structure determination from multidimensional nuclear overhauser enhancement spectra and chemical shift assignments. J Am Chem Soc 126:6258–6273. doi:10.1021/ja049786h

Leutner M, Gschwind RM, Liermann J, Schwarz C, Gemmecker G, Kessler H (1998) Automated backbone assignment of labeled proteins using the threshold accepting algorithm. J Biomol NMR 11:31–43. doi:10.1023/A:1008298226961

Li KB, Sanctuary BC (1997a) Automated resonance assignment of proteins using heteronuclear 3D NMR. 1. Backbone spin systems extraction and creation of polypeptides. J Chem Inf Comput Sci 37:359–366. doi:10.1021/ci960045c

Li KB, Sanctuary BC (1997b) Automated resonance assignment of proteins using heteronuclear 3D NMR. 2. Side chain and sequence-specific assignment. J Chem Inf Comput Sci 37:467–477. doi:10.1021/ci960372k

Lin HN, Wu KP, Chang JM, Sung TY, Hsu WL (2005) GANA: a genetic algorithm for NMR backbone resonance assignment. Nucleic Acids Res 33:4593–4601. doi:10.1093/nar/gki768

Linge JP, Habeck M, Rieping W, Nilges M (2003a) ARIA: automated NOE assignment and NMR structure calculation. Bioinformatics 19:315–316. doi:10.1093/bioinformatics/19.2.315

Linge JP, Williams MA, Spronk CAEM, Bonvin AMJJ, Nilges M (2003b) Refinement of protein structures in explicit solvent. Proteins. Struct Funct Bioinformatics 50:496–506. doi:10.1002/prot.10299

López-Méndez B, Güntert P (2006) Automated protein structure determination from NMR spectra. J Am Chem Soc 128:13112–13122. doi:10.1021/ja061136l

López-Méndez B, Pantoja-Uceda D, Tomizawa T, Koshiba S, Kigawa T, Shirouzu M et al (2004) Letter to the Editor: NMR assignment of the hypothetical ENTH-VHS domain At3g16270 from *Arabidopsis thaliana*. J Biomol NMR 29:205–206

Luginbühl P, Güntert P, Billeter M, Wüthrich K (1996) The new program OPAL for molecular dynamics simulations and energy refinements of biological macromolecules. J Biomol NMR 8:136–146. doi:10.1007/BF00211160

Lukin JA, Gove AP, Talukdar SN, Ho C (1997) Automated probabilistic method for assigning backbone resonances of (C-13, N-15)-labeled proteins. J Biomol NMR 9:151–166. doi:10.1023/A:1018602220061

Macura S, Ernst RR (1980) Elucidation of cross relaxation in liquids by 2D NMR spectroscopy. Mol Phys 41:95–117. doi:10.1080/00268978000102601

Malliavin TE, Rouh A, Delsuc MA, Lallemand JY (1992) Approche directe de la détermination de structures moléculaires à partir de l'effet Overhauser nucléaire. Comptes rendus de l'Academie des Sciences Serie II 315:653–659

Malmodin D, Papavoine CHM, Billeter M (2003) Fully automated sequence-specific resonance assignments of heteronuclear protein spectra. J Biomol NMR 27:69–79. doi:10.1023/A:1024765212223

Masse JE, Keller R (2005) AutoLink: automated sequential resonance assignment of biopolymers from NMR data by relative-hypothesis-prioritization-based simulated logic. J Magn Reson 174:133–151. doi:10.1016/j.jmr.2005.01.017

Masse JE, Keller R, Pervushin K (2006) SideLink: automated side-chain assignment of biopolymers from NMR data by relative-hypothesis-prioritization-based simulated logic. J Magn Reson 181:45–67. doi:10.1016/j.jmr.2006.03.012

Meadows RP, Olejniczak ET, Fesik SW (1994) A computer-based protocol for semiautomated assignments and 3D structure determination of proteins. J Biomol NMR 4:79–96. doi:10.1007/BF00178337

Meier BU, Bodenhausen G, Ernst RR (1984) Pattern recognition in two-dimensional NMR spectra. J Magn Reson 60:161–163

Meiler J, Baker D (2003) Rapid protein fold determination using unassigned NMR data. Proc Natl Acad Sci USA 100:15404–15409. doi:10.1073/pnas.2434121100

Moseley HNB, Montelione GT (1999) Automated analysis of NMR assignments and structures for proteins. Curr Opin Struct Biol 9:635–642. doi:10.1016/S0959-440X(99)00019-6

Moseley HNB, Monleon D, Montelione GT (2001) Automatic determination of protein backbone resonance assignments from triple resonance nuclear magnetic resonance data. Nucl Magn Reson Biol Macromol B 339:91–108

Moseley HNB, Riaz N, Aramini JM, Szyperski T, Montelione GT (2004) A generalized approach to automated NMR peak list editing: application to reduced dimensionality triple resonance spectra. J Magn Reson 170:263–277. doi:10.1016/j.jmr.2004.06.015

Mumenthaler C, Braun W (1995) Automated assignment of simulated and experimental NOESY spectra of proteins by feedback filtering and self-correcting distance geometry. J Mol Biol 254:465–480. doi:10.1006/jmbi.1995.0631

Mumenthaler C, Güntert P, Braun W, Wüthrich K (1997) Automated combined assignment of NOESY spectra and three-dimensional protein structure determination. J Biomol NMR 10:351–362. doi:10.1023/A:1018383106236

Nederveen AJ, Doreleijers JF, Vranken W, Miller Z, Spronk CAEM, Nabuurs SB et al (2005) RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. Proteins. Struct Funct Bioinformatics 59:662–672. doi:10.1002/prot.20408

Neidig KP, Saffrich R, Lorenz M, Kalbitzer HR (1990) Cluster analysis and multiplet pattern recognition in two-dimensional NMR spectra. J Magn Reson 89:543–552

Neidig KP, Geyer M, Gorler A, Antz C, Saffrich R, Beneicke W et al (1995) Aurelia, a program for computer-aided analysis of multidimensional NMR spectra. J Biomol NMR 6:255–270. doi:10.1007/BF00197807

Neuhaus D, Williamson MP (1989) The nuclear Overhauser effect in structural and conformational analysis. VCH, Weinheim

Nilges M (1995) Calculation of protein structures with ambiguous distance restraints: automated assignment of ambiguous NOE crosspeaks and disulfide connectivities. J Mol Biol 245:645–660. doi:10.1006/jmbi.1994.0053

Nilges M, Macias MJ, ODonoghue SI, Oschkinat H (1997) Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from beta-spectrin. J Mol Biol 269:408–422. doi:10.1006/jmbi.1997.1044

Olson JB, Markley JL (1994) Evaluation of an algorithm for the automated sequential assignment of protein backbone resonances: a demonstration of the connectivity tracing assignment tools (CONTRAST) software package. J Biomol NMR 4:385–410. doi:10.1007/BF00179348

Oshiro CM, Kuntz ID (1993) Application of distance geometry to the proton assignment problem. Biopolymers 33:107–115. doi:10.1002/bip.360330110

Pantoja-Uceda D, López-Méndez B, Koshiba S, Inoue M, Kigawa T, Terada T et al (2005) Solution structure of the rhodanese homology domain At4g01050(175–295) from *Arabidopsis thaliana*. Protein Sci 14:224–230. doi:10.1110/ps.041138705

Pfändler P, Bodenhausen G, Meier BU, Ernst RR (1985) Toward automated assignment of nuclear magnetic resonance spectra: pattern recognition in two-dimensional correlation spectra. Anal Chem 57:2510–2516. doi:10.1021/ac00290a018

Prestegard JH, Mayer KL, Valafar H, Benison GC (2005) Determination of protein backbone structures from residual dipolar couplings. Methods Enzymol 394:175–209. doi:10.1016/S0076-6879(05)94007-X

Pristovšek P, Rüterjans H, Jerala R (2002) Semiautomatic sequence-specific assignment of proteins based on the tertiary structure: the program st2nmr. J Comput Chem 23:335–340. doi:10.1002/jcc.10011

Rieping W, Habeck M, Bardiaux B, Bernard A, Malliavin TE, Nilges M (2007) ARIA2: automated NOE assignment and data integration in NMR structure calculation. Bioinformatics 23:381–382. doi:10.1093/bioinformatics/btl589

Rohl CA, Baker D (2002) De novo determination of protein backbone structure from residual dipolar couplings using Rosetta. J Am Chem Soc 124:2723–2729. doi:10.1021/ja016880e

Rouh A, Louisjoseph A, Lallemand JY (1994) Bayesian signal extraction from noisy FT NMR spectra. J Biomol NMR 4:505–518. doi:10.1007/BF00156617

Scott A, Pantoja-Uceda D, Koshiba S, Inoue M, Kigawa T, Terada T et al (2005) Solution structure of the Src homology 2 domain from the human feline sarcoma oncogene Fes. J Biomol NMR 31:357–361. doi:10.1007/s10858-005-0946-6

Scott A, López-Méndez B, Güntert P (2006) Fully automated structure determinations of the Fes SH2 domain using different sets of NMR spectra. Magn Reson Chem 44:S83–S88. doi:10.1002/mrc.1813

Seavey BR, Farr EA, Westler WM, Markley JL (1991) A relational database for sequence-specific protein NMR data. J Biomol NMR 1:217–236. doi:10.1007/BF01875516

Shen Y, Bax A (2007) Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. J Biomol NMR 38:289–302. doi:10.1007/s10858-007-9166-6

Shen Y, Lange O, Delaglio F, Rossi P, Aramini JM, Liu G et al (2008) Consistent blind protein structure generation from NMR chemical shift data. Proc Natl Acad Sci USA 105:4685–4690. doi:10.1073/pnas.0800256105

Simons KT, Kooperberg C, Huang E, Baker D (1997) Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. J Mol Biol 268:209–225. doi:10.1006/jmbi.1997.0959

Solomon I (1955) Relaxation processes in a system of two spins. Phys Rev 99:559–565. doi:10.1103/PhysRev.99.559

Takeda M, Ikeya T, Güntert P, Kainosho M (2007) Automated structure determination of proteins with the SAIL-FLYA NMR method. Nat Protocols 2:2896–2902. doi:10.1038/nprot.2007.423

Vitek O, Bailey-Kellogg C, Craig B, Kuliniewicz P, Vitek J (2005) Reconsidering complete search algorithms for protein backbone NMR assignment. Bioinformatics 21:230–236. doi:10.1093/bioinformatics/bti1138

Vitek O, Bailey-Kellogg C, Craig B, Vitek J (2006) Inferential backbone assignment for sparse data. J Biomol NMR 35:187–208. doi:10.1007/s10858-006-9027-8

Volk J, Herrmann T, Wüthrich K (2008) Automated sequence-specific protein NMR assignment using the memetic algorithm MATCH. J Biomol NMR 41:127–138. doi:10.1007/s10858-008-9243-5

Wang JY, Wang TZ, Zuiderweg ERP, Crippen GM (2005) CASA: an efficient automated assignment of protein mainchain NMR data using an ordered tree search algorithm. J Biomol NMR 33:261–279. doi:10.1007/s10858-005-4079-8

Wishart DS, Sykes BD (1994) The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. J Biomol NMR 4:171–180. doi:10.1007/BF00175245

Wu KP, Chang JM, Chen JB, Chang CF, Wu WJ, Huang TH et al (2006) RIBRA: an error-tolerant algorithm for the NMR backbone assignment problem. J Comput Biol 13:229–244. doi:10.1089/cmb.2006.13.229

Xu J, Straus SK, Sanctuary BC, Trimble L (1993) Automation of protein 2D proton NMR assignment by means of fuzzy mathematics and graph theory. J Chem Inf Comput Sci 33:668–682. doi:10.1021/ci00015a004

Xu J, Straus SK, Sanctuary BC, Trimble L (1994) Use of fuzzy mathematics for complete automated assignment of peptide ^{1}H 2D NMR spectra. J Magn Reson B 103:53–58. doi:10.1006/jmrb.1994.1006

Xu Y, Xu D, Kim D, Olman V, Razumovskaya J, Jiang T (2002) Automated assignment of backbone NMR peaks using constrained bipartite matching. Comput Sci Eng 4:50–62

Xu YZ, Wang XX, Yang J, Vaynberg J, Qin J (2006) PASA: a program for automated protein NMR backbone signal assignment by pattern-filtering approach. J Biomol NMR 34:41–56. doi:10.1007/s10858-005-5358-0

Zimmerman DE, Kulikowski CA, Huang YP, Feng WQ, Tashiro M, Shimotakahara S et al (1997) Automated analysis of protein NMR assignments using methods from artificial intelligence. J Mol Biol 269:592–610. doi:10.1006/jmbi.1997.1052