Chapter 16

Automated Structure Determination from NMR Spectra

Elena Schmidt and Peter Güntert

Abstract

Three-dimensional structures of proteins in solution can be calculated on the basis of conformational restraints derived from NMR measurements. This chapter gives an overview of the computational methods for NMR protein structure analysis highlighting recent automated methods for the assignment of NMR spectra, the collection of conformational restraints, and the structure calculation.

Key words Protein structure, NMR structure determination, Automated assignment, Resonance assignment, NOESY assignment, Conformational restraints, Network anchoring, Constraint combination, Torsion angle dynamics, CYANA, FLYA

1 Introduction

Until some years ago NMR protein structure determination was a laborious undertaking that occupied a trained spectroscopist over several months for each new protein structure. It was then recognized that many of the time-consuming manual steps carried out by an expert during the process of spectral analysis could be accomplished by automated, computational approaches [1]. Today automated methods for NMR structure determination are playing an ever more prominent role and are superseding the conventional manual approaches to solving three-dimensional protein structures in solution. This chapter gives an introduction to automated NMR assignment and structure calculation methods. Parts of this chapter were first published in the doctoral thesis of Elena Schmidt [2] and in [3, 4].

In most cases, protein structure determination is performed by a standard sequence of steps that are illustrated in Fig. 1. In the following, this standard procedure [5–8] is described with emphasis on peak picking, chemical shift assignment, nuclear Overhauser effect (NOE) assignment, and structure calculation.

Raymond J. Owens (ed.), *Structural Proteomics: High-Throughput Methods*, Methods in Molecular Biology, vol. 1261, DOI 10.1007/978-1-4939-2230-7_16,

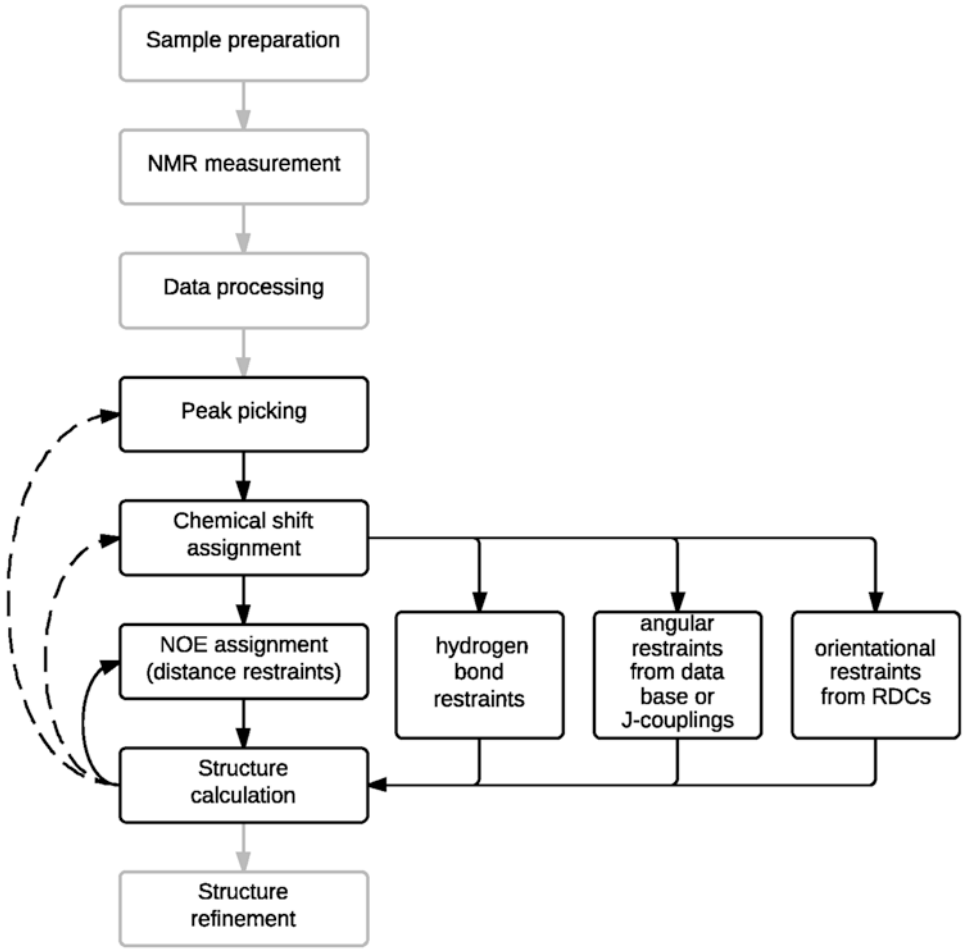

Fig. 1 Standard steps of NMR protein structure determination. The steps which are described in detail and can be applied iteratively are shown in *black boxes*. The utilization of structural information for the improvement of peak picking and chemical shift assignment is indicated with *dashed lines*

The first step in the process of protein structure determination is the preparation of the protein sample [9]. The protein to be studied is in most cases overexpressed in a bacterial system, which is usually grown on a $^{13}C/^{15}N$ isotopically enriched minimal medium. The protein is purified in order to obtain a sample of a few hundred microliters with a concentration in the (sub)millimolar range (>0.05 mM is sufficient using certain techniques).

The second step is the measurement of the atom signals with an NMR spectrometer, in which the protein sample is exposed to a strong magnetic field and a sequence of radiofrequency pulses. A set of different NMR experiments that differ from one another with respect to the pulse sequence are performed with the same sample. They result in experiment-specific signals that reveal the

covalent and spatial connectivities of the protein atoms. The third protein structure determination step is data processing. In order to obtain the NMR spectra, the measured time domain data is converted to frequency domain data using Fourier transformation and other techniques. In the fourth step "real" signals, which result from protein atoms, must be identified in the spectra and distinguished from noise and artifacts, which is referred to as peak picking.

The resulting peak lists are the basis for the next step, chemical shift assignment. The chemical shift values that are observed in the spectra are assigned to the corresponding protein atoms, since the relationship of the measured signals and the protein atoms is not known from the beginning.

The sixth step is NOE assignment. The cross peaks in NOESY spectra, which hold information about atom–atom distances in the 3D structure of the protein, are assigned to the respective atoms based on the chemical shift assignment. Distance restraints are deduced from the volumes of these peaks. In the seventh step the 3D structure is calculated based on distance restraints. The distance information can be complemented with angular restraints from chemical shifts or J-couplings, orientational restraints from RDCs, and hydrogen bond restraints. As soon as a preliminary structure is obtained, the structural information is used to improve the NOE assignment. This is done in several cycles.

Peak picking, chemical shift assignment, and NOE assignment are error-prone methods, especially when done with automated procedures. One possibility to reduce errors in automated procedures and to improve the structural quality is to apply these steps iteratively and to incorporate the structural information obtained from structure determination into peak picking and chemical shift assignment. This can be done by comparison of simulated spectra with real spectra in case of peak picking and structure-based chemical shift prediction and expected peak prediction in case of chemical shift assignment. Finally, the last structure determination step is to refine the structure using force fields adapted from molecular dynamics simulation packages.

2 Peak Picking

Peak picking is the procedure of extracting the positions of "real" peaks that result from molecule atoms from NMR spectra usually with several attributes like volume and shape. Resulting peak lists provide the basis for successful automated chemical shift and NOE assignment. Like chemical shift assignment, peak picking is a critical step in automated structure determination, since it is very prone to errors. Various automated programs exist for this task, but it is still common to pick peak lists manually or to refine them with manual intervention. Popular programs that can be used for peak

picking are CAPP [10], GIFA [11], AUTOPSY [12], ATNOS [13], SPARKY [14], NMRVIEW [15, 16], AURELIA [17], CcpNmr Analysis [18], and XEASY [19].

The challenge of peak picking is to identify all peaks even in overlapped regions and to distinguish between real peaks that are atom signals and noise or artifacts. The quality of the resulting peak lists has a large impact on chemical shift assignment, since real peaks that have not been included into the peak list can make it impossible to assign the respective atoms. On the other hand, additional peaks in the peak list may be confused with real peaks and can therefore lead to erroneous assignments.

Automated methods use several criteria to identify the set of real peaks in a spectrum. The most straightforward criteria are local maxima and the peak intensity. To identify also low-intensity peaks, the peak shape is considered. Simple shape attributes like line width are used, but also more advanced methods are applied to compare measured line shapes to ideal line shapes. Such procedures are implemented in AUTOPSY, CAPP, or ATNOS. Apart from just taking attributes of a specific peak into consideration, information about the experiment and, if available, about the assignment of the atoms and the protein structure can be included, e.g., in ATNOS. Peaks observed in other experiments or the symmetric properties of some spectra can help to distinguish between real peaks and artifacts, by providing the information whether or not a peak is expected at a specific position. This information can also be obtained from chemical shift assignments and the protein structure, which makes it possible to use peak picking, chemical shift assignment, and structure determination iteratively to refine a protein structure after the initial structure calculation.

3 Resonance Assignment

Every measured atom in a macromolecule has a specific chemical shift value, which depends on the chemical environment of this nucleus. The problem is that it is unknown from the start which atom leads to which chemical shift value. Revealing the relationship between atoms and chemical shifts is denoted as chemical shift assignment. Chemical shift assignment is necessary not only to evaluate the distance information in NOESY spectra for standard protein structure determination by NMR, but also in all cases in which atom-specific information has to be obtained from an NMR experiment. Examples are molecular interaction studies or alternative approaches for protein structure determination that are based on chemical shifts or RDCs, and investigations of protein dynamics.

To enable chemical shift assignment, several NMR experiments have to be performed that complement each other such that the connectivity of the atoms in a protein is represented. Based on the

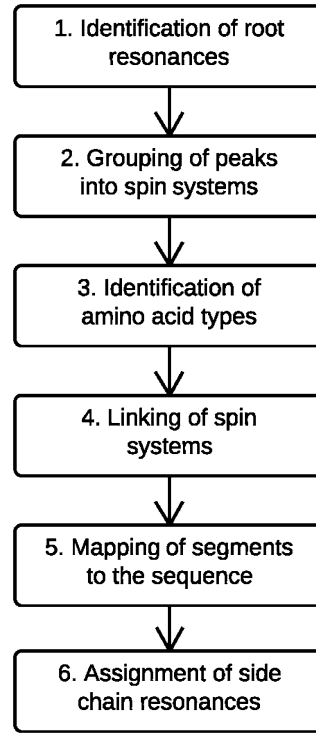

Fig. 2 Common steps of chemical shift assignment

knowledge about the covalent structure that can be deduced from the protein sequence, it is possible to establish the relationship between chemical shifts and atoms. Usually, a set of standard experiments are used to reveal the covalent atom connectivities.

Since the general strategy for chemical shift assignment has been described in the 1980s [20], there have been many attempts to establish an automated procedure for this process [1, 6, 21]. The published programs are based on various different optimization techniques including exhaustive search, best first approaches, genetic algorithms, and Monte Carlo methods. Some programs incorporate the whole chemical shift assignment process shown in Fig. 2, starting with peak lists or NMR spectra as input data and ending with the complete assignment of backbone and side chain atoms. Others are specialized in single steps of the process, usually sequence-specific assignment, in which the spin systems are assigned to the correct positions in the amino acid sequence of the protein. A selection of different programs for automated chemical shift assignment is listed in Table 1.

To date, none of these automatic approaches has been able to completely replace the procedure of manually assigning chemical shifts, since the completeness and reliability of the assignment results is often too low or the programs are only tested with

Table 1
Selected automated assignment programs

Program	Side chain	Spin systems	Input peak lists and notes
FLYA [39]	Yes	Yes	Any
GARANT [30, 31]	Yes	Yes	Standard 2D/3D through-bond, NOESY
PINE [24]	Yes	Yes	Standard 2D/3D through-bond, NOESY for RNA
AUTOASSIGN [22]	No	Yes	Standard 2D/3D through-bond
Moseley *et al.* [67]	No	Yes	3 3D/4D solid state
Li Sanctuary [68, 69]	Yes	Yes	Any general data set
ADAPT-NMR [25]	No	Yes	Integrated data collection
MARS [28]	No	No	Grouped chemical shifts
MATCH [26]	No	Yes	Designed for APSY [27]; possible with 3D
PASTA [70]	No	Yes	Any 2D/3D
PACES [71]	No	Assisted	Any triple resonance; user-assisted
MAPPER [72]	No	No	Shifts for amino acid stretches
DYNASSIGN [42]	Yes	Yes	Any
ASCAN [73]	Only	No	NOESY; backbone shifts
MONTE [74]	Yes	No	Grouped chemical shifts
TATAPRO [75]	No	Yes	Standard through-bond
Llinas et al. [76]	Yes	Yes	NOESY
Lukin et al. [77]	No	Yes	Standard through-bond
MCASSIGN2 [78]	Yes	Yes	2D/3D solid state; needs typing

The column "Side chain" gives information about what kind of assignment is done. "Yes" refers to assignment of backbone and side chains, "no" refers to assignment of the backbone, and "only" refers to assignment of the side chains. The column "Spin systems" gives information about whether the grouping of peaks into spin systems is done by the program. "yes" means that spin systems are generated automatically, "no" means that the peaks have to be grouped by the user, and "assisted" means that the grouping is done by the program but requires manual intervention

simulated data sets and fail when they are applied to real data of standard quality. The reason is that the human strategies that are used for manual assignment are difficult to convert into algorithms. Human experts follow the scheme for chemical shift assignment depicted in Fig. 2, but verify their decisions made in previous steps or complete intermediate results during the whole process, using

the information that they gained in the current step. It might be impossible, for example, to group some peaks correctly into two spin systems based solely on the information in the respective spectrum in cases in which the root resonances overlap. Later in the process these ambiguities can be resolved as soon as neighboring residues are sequence-specifically assigned.

Often, programs for automated assignment follow the standard steps for chemical shift assignment, but do not improve intermediate results that were obtained in previous steps. This can lead to erroneous assignments since not all information available was used. The same problem holds true if an automated procedure only performs single steps of the assignment process. The remaining steps have to be done manually by the user. Especially spin system identification, which is sometimes omitted by automated procedures, is time-consuming if done accurately and can in some cases only be completed during later steps of the manual assignment process. This is not possible if the results obtained in one step are not used to improve previous results. The situation is similar for peak picking. Peak lists are often updated during the assignment process, when chemical shift assignments can help to distinguish between noise or artifacts and real peaks. Automated assignment programs often use peak lists as input. Hence, the results produced by these programs strongly depend on the quality of these peak lists.

Another reason why programs for automated chemical shift assignment are not frequently used is that they are often restricted to a specific application. Many programs just allow for input from standard through-bond triple resonance spectra. In this case information from NOESY experiments, which is useful for side chain assignment, cannot be used and they cannot be applied whenever the measured data is different from the typical pattern.

In the next sections, a number of programs for automated chemical shift assignment are presented in more detail as examples for different optimization strategies and applications. AUTOASSIGN is the most popular program for automated chemical shift assignment of backbone atoms. PINE is able to perform all chemical shift assignment steps given standard peak lists obtained from through-bond spectra as input. ADAPT-NMR does not assign chemical shifts based on peak lists or NMR spectra, but directly controls the measurement of the data. MATCH is specialized on the assignment of high-dimensional APSY spectra. MARS can incorporate RDCs into the assignment process. The algorithm implemented in GARANT provided the bases for the FLYA automated assignment procedure, since it is not restricted to specific input spectra while providing complete chemical shift assignment and good results.

3.1 AUTOASSIGN

AUTOASSIGN [22] performs backbone assignment (HN, H^{α}, C, C^{α}, N, and C^{β}) based on a best-first approach using peak lists from the [^{15}N,^{1}H]-HSQC experiment and standard 3D triple-resonance

experiments that are commonly used for manual backbone assignment.

The program starts with filtering of peak lists based on the N-H dimensions that are common in all lists and aligning the lists according to the resonances of isolated peaks. Root resonances for the grouping of peaks into spin systems are obtained from all peaks in the [^{15}N,^{1}H]-HSQC spectrum, and completed by HNCO peaks for resonances that were not included in the [^{15}N,^{1}H]-HSQC.

In the next step the peaks from the other lists are assigned to these generic spin system roots if their N and H frequencies match those of the root peak. The peaks of a generic spin system are grouped into two lists of chemical shifts, the Cα and the C′ ladder, containing the chemical shifts of the respective amino acid itself and the previous one. According to the completeness, intensity, and degeneracy, the generic spin systems are categorized into distinct, overlapped, and weak. Based on the general chemical shift statistics of the BMRB and using Bayesian statistics for the C$^{\alpha}$ and C$^{\beta}$ chemical shifts, possible amino acid type lists are created for the C$^{\alpha}$ and C′ ladders of the generic spin systems. From these two lists a number of possible positions in the sequence of the respective dipeptide are deduced.

The following steps of linking different generic spin systems and assigning them to a position in the sequence are done using a constraint propagation method [23]. The principle is that these steps are first done with the category of distinct spin systems, followed by the overlapped and finally weak ones. The assignment of spin systems in the first category reduces the remaining assignment possibilities for all other spin systems and thereby simplifies assignment decisions for these spin systems. One cycle works as follows. All C$^{\alpha}$ and C′ ladders are compared and a list of possible nearest neighbor links between the spin systems is created. The list is first ranked by the number of matching frequencies, and then by a match value. Links between neighbors are established if they build a unique one-to-one match within the group of the same number of matching frequencies (best-first approach). In the next step these spin system neighbor pairs are assigned to the sequence if there is a unique one-to-one match.

3.2 PINE

PINE (Probabilistic Interaction Network of Evidence algorithm) [24] performs probabilistic chemical shift assignment of backbone and side chain atoms and determines the secondary structure of the respective protein. It uses the protein sequence and two- and three-dimensional peak lists obtained from through-bond experiments as input. The method can in principle be extended to other experiments. Prior information about atom assignments can be included in the calculation.

The first step is to build up a network connecting the measured chemical shifts to all labeling possibilities, i.e., the atom names in the protein. This network is built up as follows. In order to group measured peaks into spin systems, similarity scores between peaks are determined based on the distances of peaks in common dimensions.

Starting with peaks of the most sensitive experiments, normally [^{15}N,^{1}H]-HSQC or HNCO, spin systems are initialized and peaks with similarity scores greater than zero are added with a distance-dependent probability. Resulting spin systems cover peaks of spin system i and $i-1$. Connectivity scores between spin systems are calculated following the same scheme that was used for the calculation of similarity scores. According to the connectivity scores, the spin systems are grouped into triplet spin systems, since the usage of these triplet spin systems for assignment instead of single spin systems reduces the complexity of the network.

Amino acid typing is done by calculating a score for each combination of spin system and amino acid triplet that can be assigned to each other. The respective scoring of a single atom takes into account the BMRB chemical shift statistics and the secondary structure prediction with respect to the atom in the sequence. The score for a triplet is obtained by calculating the product of the respective atom scores in the triplet spin system.

The assignment is done in several iterations. The backbone assignment probabilities in the network are determined based on the amino acid scoring, connectivity experiments, backbone assignment in the previous iteration, and outlier detection. The topology and the probabilities of this network are changed during several iterations until a quasi-stationary state is reached, which means that topology and probabilities do not vary significantly. In each iteration an energy function is evaluated, and a belief propagation algorithm is applied to obtain an updated network and thereby probabilistic assignments as well as the probabilities for the secondary structure. Rather than a single assignment for each atom, several probability-weighted possibilities are obtained.

After the backbone assignment a separate network model for each amino acid is generated and the belief propagation algorithm is applied to obtain probabilistic side chain assignments.

3.3 ADAPT-NMR

ADAPT-NMR (Assignment-directed Data collection Algorithm utilizing a Probabilistic Toolkit in NMR) [25] provides fully automated backbone assignment and secondary structure prediction. It implements the concept of iterative chemical shift assignment and data measurement using a probabilistic network approach.

The algorithm starts with the generation of probabilistic spin systems on the basis of the [^{15}N,^{1}H]-HSQC spectrum. A probabilistic spin system has different attributes. Each attribute may have

different assignments at the same time. For example, attributes are the measured chemical shifts.

During the optimization cycle, evaluation of the probabilistic spin systems leads to the experiment type and the plane to be recorded next. The respective peaks are picked automatically and a probability based on peak and experiment characteristics is determined using several machine learning techniques. Subsequently, a pseudoenergy model is used to update probabilistic spin systems.

For high-probability peaks in 3D spectra that do not match the [^{15}N,^{1}H]-HSQC and overlapping regions additional spin systems can be introduced during the calculation. If the spin system quality is below a specified threshold, it is optimized in a new cycle of data collection and probabilistic network update. When the threshold is reached, the assignment step is done.

The core part of the assignment step originates from the PINE algorithm with some modifications, e.g., a fully probabilistic implementation. In the assignment step probabilities for chemical shift assignments, secondary structure states, and outlier chemical shift values are determined. As soon as an initial assignment is available, the assignment and the secondary structure are also considered for the data collection, i.e., selection of the next experiment and plane. At this point data is explicitly collected for spin systems that are weakly linked in order to maximize the information gain in the data collection step.

3.4 MATCH

MATCH (Mimetic Algorithm and Combinatorial Optimization Heuristics) [26] provides chemical shift assignment of the protein backbone using peak lists as input. The program was implemented for the use with APSY [27] spectra, but can also be used with standard triple experiments. It combines an evolutionary algorithm with a local optimization routine.

During the initialization process, the measured frequencies are grouped into spin systems and these are assembled into bigger fragments of a given maximum size. Therefore, connectivities between spin systems are identified based on a scoring function. Each spin system obtains the list of all possible fragments it is part of and finally, isolated spin systems are removed and control parameters for the calculation are set according to the degree of ambiguity of the data.

To start with, an initial population of assignment solutions is generated. Each assignment solution is obtained by randomly selecting a fragment of maximum length and mapping it to the position in the sequence with the highest sequence-specific score. The sequence-specific score evaluates the agreement of the measured chemical shifts with the chemical shifts statistics of the BMRB. The process is repeated first with the remaining fragments of the same length. Afterwards the length of the fragments to be mapped is decremented. Fragments including spin systems that are

already mapped are not considered in further selections. During the local optimization, pairs of fragments that could be mapped to the sequence but have not yet been assigned permanently are selected randomly and tested for compatibility of spin systems and matching adjacent spin systems and possibly they are interchanged. If the sequence-specific score exceeds a given threshold, a temporary assignment is created, which means that the assignment will not be changed by the local optimization anymore. Permanent assignments are established if a specific assignment can be found in more than a given fraction of the population.

In the global optimization routine the solutions of several individuals are combined into new assignment solutions. The individuals are ranked according to their sequence-specific score and the best-scored solutions are used to build up the next generation. The optimization ends when all atoms have permanently been assigned.

If the calculation does not converge, the whole process is repeated with modified control parameters. The optimization is repeated several times to obtain independent results. An assignment is output if it is present in at least 50 % of the independent runs.

3.5 MARS

The program MARS [28, 29] performs backbone chemical shift assignment using intra- and inter-residue chemical shifts, which have to be grouped into "pseudoresidues" by the user. The program can include RDCs into the assignment process if a 3D protein structure is available.

The first step of the algorithm is the generation of all possible sequential connections between pseudoresidues. Connections that do not agree with the experimental data are removed at later stages. Distances between the experimental chemical shifts of the pseudoresidues and chemical shift predictions for possible sequence assignments are calculated. These predictions are obtained based on the standard BMRB statistics, correcting for neighbor residue effects and secondary structure effects. According to these distances, the sequence positions are ranked and a pseudoenergy is determined.

The optimization starts with a random assignment of the peudoresidues to positions in the amino acid sequence. Starting from a random pseudoresidue, segments of five residues including this residue are assembled based on the connectivity information. The segments are mapped onto all possible positions of the sequence. The probability for an assignment is calculated using the pseudoenergy and the solutions are ranked, the minimum representing the best solution. If starting with the last pseudoresidue of the respective segments leads to the same assignment solution, the solution is considered reliable and a penalty for all other possibilities is included into the pseudoenergy function. The procedure is repeated with all pseudoresidues as starting points for the assembly. In subsequent steps the segment size is decreased. The procedure

is further repeated adding varying noise to the predicted chemical shifts. Consistent solutions are considered as reliable. If RDC measurements and a 3D protein structure are available for the respective protein, the agreement between the measured RDCs and RDCs calculated from the provided protein structure is included into the distance function.

3.6 GARANT

GARANT (General Algorithm for Resonance AssignmeNT) [30] is a program for automatic chemical shift assignment of backbone and side chain atoms. It can be used for chemical shift assignment solely based on a set of peak lists and the protein amino acid sequence, but can also assign chemical shifts based on a given 3D protein structure [31].

The general concept of GARANT is that the connectivities between the atoms of a protein that can be revealed with different NMR experiments are represented by a network of expected peaks and protein atoms. The mapping of this network to the network of measured peaks and their chemical shifts leads to an assignment of the atoms to the respective chemical shifts. The optimization problem of finding the mapping possibility that corresponds to the correct assignment solution is solved using an evolutionary algorithm in combination with a local optimization routine.

Both optimization routines use a scoring scheme based on the concept of mutual information. The different terms of the scoring function for the evolutionary optimization evaluate the assignment of a chemical shift to an atom, based on the agreement with general chemical shift statistics, the mapping of an expected peak to a measured peak, and the agreement of the chemical shifts of single signals and the respective atom chemical shift. Ambiguous peak mappings lead to a reduced score. The global score results from the sum over the terms resulting from all atom and peak mappings.

The evolutionary optimization routine is controlled by a simulated annealing temperature schedule. It uses a population of 50 assignment solutions by default. For the construction of a new generation, parent solutions with a high global score are favored. Mappings for a specific residue are adapted from parent solutions as far as possible. If no parent solution is available, those of residues with similar spin systems are considered. If no mapping solution could be determined that way, new mappings are generated.

The local optimization routine selects unmapped or ambiguously mapped expected peaks and evaluates the assignment of neighboring atoms based on a local score and reassigns these atoms, if necessary. The local optimization routine also uses the mutual information-based scoring scheme, except that for the scoring of an atom only contributions that are directly related to the atom are included.

An advantage of the GARANT algorithm compared to many other methods is that in principle it can solve the assignment

problem using every combination of spectra that contain sufficient information for the assignment. Even though the algorithm is generally applicable to any kind of common NMR spectrum, the specification of a given set of spectra is done within the program and can only be extended by changes of the C++ source code.

The fully automated chemical shift assignment program GARANT was introduced in 1996. In the respective publications [30, 31] the application of GARANT was demonstrated by assigning different proteins up to a sequence length of 165 amino acids, based on data sets that consisted solely of homonuclear 2D spectra as well as data sets consisting of 3D spectra. GARANT has been used in various other projects, and has also been adapted for calculations with APSY [27] spectra and applied to the assignment of 4–7D APSY spectra in different applications [32, 33].

To obtain full automation of chemical shift assignment starting from NMR spectra, GARANT was combined with the automated peak picking program AUTOPSY [12] and a program for calibration and filtering, PICS [34]. GARANT was combined with AUTOPSY, the structure calculation program CYANA, and the molecular dynamics simulation package OPALp to achieve fully automated structure determination, including peak picking, chemical shift assignment, NOE assignment, structure calculation, and energy refinement [35]. It has been shown that this strategy is in principle also applicable to sparse data [36, 37]. Even structure determination based solely on NOESY data was successful [38].

3.7 FLYA

The new FLYA automated chemical shift assignment procedure has been implemented and applied to several targets [4, 39]. As described above, various programs for automated chemical shift assignment have been developed before, but none of these approaches has become a standard procedure. The main reasons for this are the generally low accuracy of the assignment results and restrictions to specific applications.

The aforementioned GARANT program for automated chemical shift assignment is based on an optimization strategy that can in principle be applied to every kind of NMR spectrum and provides good assignment results compared to other programs. However, assignment calculations with GARANT take relatively long, i.e., several hours in some cases, the application to nonstandard data sets is not straightforward and the accuracy of the results leaves room for improvement. This situation led to the development of a new algorithm, FLYA, with the following objectives: (1) improving the accuracy of the chemical shift assignment, (2) improving the flexibility of the method in order to address a wider range of problems, (3) shortening the run time of the algorithm, and (4) incorporating automated resonance assignment into the CYANA software in order to simplify the application of the algorithm in conjunction with other CYANA modules, e.g., NOE distance restraint assignment and structure calculation by torsion angle dynamics.

The resulting implementation of the FLYA automated assignment algorithm [39] in the CYANA software package includes a modified expected peak generation procedure, a new scoring scheme, and various further improvements of the optimization algorithm over the earlier GARANT approach.

NMR resonance assignment is based on experiments that correlate nuclear spins such that they give rise to cross peaks in multidimensional spectra. Assignment experiments are chosen to complement each other in such a way that the connectivity of the atoms in a protein can be represented by a network of peaks that are expected to be observed. Mapping this network of expected peaks with unknown positions to the unassigned measured peaks with known positions provides an assignment of the frequencies to the spins [30, 31]. The FLYA algorithm for automated backbone and side chain resonance assignment uses this general approach to assign all kinds of NMR spectra. It is implemented in the software package CYANA [40, 41]. As input, FLYA uses exclusively the sequence of the protein and unassigned peak lists from any combination of multidimensional solution-state or solid-state NMR spectra.

All experimental data is used simultaneously in order to exploit optimally the redundancy present in the input peak lists and to avoid potential pitfalls of assignment strategies in which results obtained in a given step remain fixed input data for subsequent steps. Instead of prescribing a specific assignment strategy, the FLYA resonance assignment algorithm generates the peaks expected in a given spectrum by applying a set of rules for through-bond or through-space polarization transfer, and determines the resonance assignment by constructing an optimal mapping between the expected peaks, assigned by definition but having unknown positions, and the measured peaks, initially unassigned but with known positions in the spectrum [30, 31, 39, 42].

The rules for generating expected peaks have been implemented for many different solution-state and solid-state NMR experiments. Expected peaks for experiments like NOESY or DARR, which give signals between atoms that are close in space, are obtained using random structures of the respective proteins [39]. An expected peak is generated for each atom pair up to a given cutoff on the maximal distance between the two atoms in the ensemble of random structures. This will generate expected peaks only if the atoms are close together in the primary structure, e.g., for intra-residual and sequential distances. It corresponds to the generation of expected peaks for NOE-based experiments in solution NMR. Expected peaks for all other experiments are obtained based on the covalent connections between atoms. For each experiment the covalent bond patterns that hold this information are provided to the algorithm in the CYANA library file. It is straightforward to add new experiments or to modify the rules for existing experiments.

The best mapping of expected peaks to measured peaks is obtained using an evolutionary optimization routine that works with a population of individuals, each representing an assignment solution for the protein. This evolutionary optimization is complemented by local optimization. Solutions that are produced during the optimization are generated such that the search space of an expected peak for a mapping is defined by a chemical shift statistics (by default from the BMRB [43], or user defined), the deviations of the measured frequencies of measured peaks that are assigned to the same atom remain within a given tolerance, and an expected peak can be mapped to only one measured peak. The first generation of solutions is generated randomly, but subject to these conditions. In each generation a local optimization algorithm takes small parts of a mapping back and reassigns the expected peaks for a defined number of iterations, 15,000 is default. Afterwards the different solutions of one generation are recombined into a new generation. The individuals and the specific parts of an individual that contribute to a new individual are selected via a scoring function. The solution that maximizes this function is given as the final assignment at the end of the calculation.

The global score for complete assignment solutions evaluates four attributes of an assignment solution, the distribution of chemical shift values with respect to the given shift statistics, the alignment of peaks assigned to the same atom, the completeness of the assignment, and a penalty for chemical shift degeneracy. The global score G is defined by

$$G = \frac{\sum_{a \in A}\left[W_1(a) Q_1(a) + \sum_{n \in N_a'} W_2(a,n) Q_2(a,n) / b(n) \right]}{\sum_{a \in A_0}\left[W_1(a) + \sum_{n \in N_a} W_2(a,n) \right]}$$

The term A_0 denotes the set of all atoms for which expected peaks exist, $A \subseteq A_0$ the set of assigned atoms, N_a the set of expected peaks for atom a, and $N_a' \subseteq N_a$ the subset of expected peaks that are mapped to a measured peak. $b(n)$ refers to the ambiguity of the assignment and equals the number of expected peaks that are assigned to the same measured peak as expected peak n. Unassigned atoms and unmapped peaks contribute through the normalization by the denominator. Relative weights of the individual contributions are given by $w_1(a)$ and $w_2(a, n)$ and in [39] these were set to $w_1(a) = 4$ and $w_2(a, n) = 1$ for all calculations. The quality measure $Q_1(a)$ represents the agreement of the average chemical shift $\bar{\omega}(a)$ in the chemical shift list of atom a with the corresponding general chemical shift statistics. Similarly, $Q_2(a, n)$ measures the agreement between the chemical shift $\omega(a, n)$ of atom a obtained from the measured peak to which the expected peak n is mapped and the

average frequency of the atom in the assigned peaks of the corresponding spectrum [39]. The quality measures Q are designed such that a perfect match corresponds to $Q=1$, $Q<1$ in all other cases, a deviation that is considered "as bad as no assignment" yields $Q=0$, and an infinitely large deviation $Q=-\infty$. Consequently, the global score G is normalized such that $G=1$ for a (hypothetical) perfect assignment, and $G<1$ in all other cases.

The main difference to solution NMR lies in the rules for generating expected peaks, which have been implemented for many different solid-state NMR experiments (Table 1).

To improve and assess the accuracy of the assignment, m independent runs of the algorithm are performed with different random seeds. For each atom a consensus chemical shift is computed from the values obtained in the individual runs [34, 35, 39]. The consensus chemical shift $\tilde{\omega}(a)$ for an atom a is the value that maximizes the function

$$\mu(\omega)=\frac{1}{m}\sum_{j=1}^{m}\exp\left(-\frac{1}{2}\left(\frac{\omega-\omega_j(a)}{\epsilon(a)}\right)^2\right),$$

where $\omega_j(a)$ is the chemical shift value obtained for atom a in run j, and $\epsilon(a)$ is the chemical shift tolerance. The maximum value of this function, $\mu(\tilde{\omega}(a))$, is a measure of the self-consistency of the chemical shift values obtained in the individual runs of the algorithm, since it approximately equals the fraction of runs that yielded a chemical shift value within the tolerance $\epsilon(a)$ from the consensus value $\tilde{\omega}(a)$. This quantity can be calculated without knowledge of reference assignments. If all chemical shift values are identical, then $\mu(\tilde{\omega}(a))=1$. We consider assignments with $\mu(\tilde{\omega}(a))\geq 0.8$ as "strong" or self-consistent, all others as "weak." Weak assignments should be considered as tentative, although they are correct in many cases.

In the following, an example of a CYANA macro for a standard chemical shift assignment is shown, which refers to the published assignment calculations [39]. Executing this macro leads to a standard chemical shift assignment of the protein atoms and evaluation of the assignment results based on reference shifts. NOE assignment and structure calculation are not done.

```
1. assigncs_accH:=0.03
2. assigncs_accC:=0.4
3. assigncs_accN:=0.4
4. assignpeaks:=N15NOESY,C13NOESY,\
5. C13HSQC,N15HSQC,HCCHTOCSY,HCCHCOSY,HNCA,HNcaCO,HNCO,HNcoCA,\
6. CBCANH,CBCAcoNH,HBHAcoNH,CcoNH,HCcoNH
```

```
7. cyanalib
8. read seq protein.seq
9. flya shiftreference=protein.prot runs=20 assignpeaks=
   $assignpeaks
```

Lines 1–3: The tolerances are set for the assignment calculation and the comparison with the reference chemical shifts in the file "protein.prot," 0.03 ppm for hydrogen atoms and 0.4 ppm for carbon and nitrogen atoms.

Line 4: The peak list names of NOESY experiments are specified. Expected peaks for all experiments are generated according to the entries in the CYANA library file. In case of expected peaks for NOESY experiments distances are deduced from a random structure ("start.pdb"). Alternatively, the user can specify a structure.

Line 5/6: The peak list names of the through-bond experiments are specified.

Line 7: The CYANA standard library is read.

Line 8: The protein amino acid sequence is read from the file "protein.seq".

Line 9: The command flya, which is specified in the macro "flya.cya," is executed. The chemical shifts in file "protein.prot" obtained from manual assignment are used as reference for the automated assignment. The number of independent assignment runs is set to 20. The content of the variable "assignpeaks" is given as input for the parameter "assignpeaks" of the macro "flya.cya".

4 NOE Assignment

Structure determination with distance restraints obtained from NOESY spectra relies on the fact that ^{1}H nuclei which are separated by less than 5 Å in the protein lead to cross-peaks in the spectra according to the isolated two-spin approximation. The cross-peak volumes are proportional to the inverse sixth power of the distance between the nuclei. Since neighbor effects may weaken the observed signal, the presence of a cross-peak provides only an upper distance limit for the distance between the two hydrogen atoms [44]. In order to convert the peak volumes observed in the spectra to distance restraints, a spectrum-dependent calibration constant has to be determined.

During NOE assignment the observed cross-peaks in a NOESY spectrum are assigned to the corresponding atom pairs according to the resonance assignment, in order to generate a list of distance restraints, which provides the basis for structure calculation. An illustration of a protein structure with all distance restraints, which

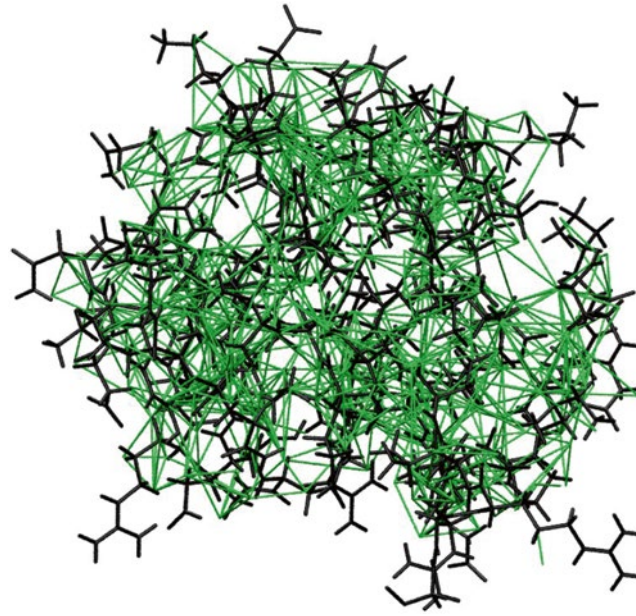

Fig. 3 Distance restraints that were obtained from ^{13}C and ^{15}N-resolved NOESY spectra and used as input for the structure calculation of the protein. The protein atoms are shown in *black*. Distance restraints in *green* connect corresponding hydrogen atoms

were obtained from NOESY spectra and used for the structure calculation, is shown in Fig. 3.

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 [7]. 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 semiautomatic approaches [45–47], 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 [48, 49], ARIA [50–53], AUTOSTRUCTURE [54], KNOWNOE [55], CANDID [56] and a similar algorithm implemented in CYANA [57], PASD [58], and a Bayesian approach [59]. Automated NOE assignment algorithms generally require a high degree of completeness of the backbone and side chain chemical shift assignments [60].

4.1 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 [41, 57]. This algorithm is a re-implementation of the former CANDID algorithm [56] 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. The precision of the structure determination normally improves with each subsequent cycle. Accordingly, the cutoff for acceptable distance restraint violations in the calculation of $P_{\text{structure}}$ is tightened from cycle to cycle. Third, each assignment possibility is evaluated for its network anchoring (see below), which is quantified by

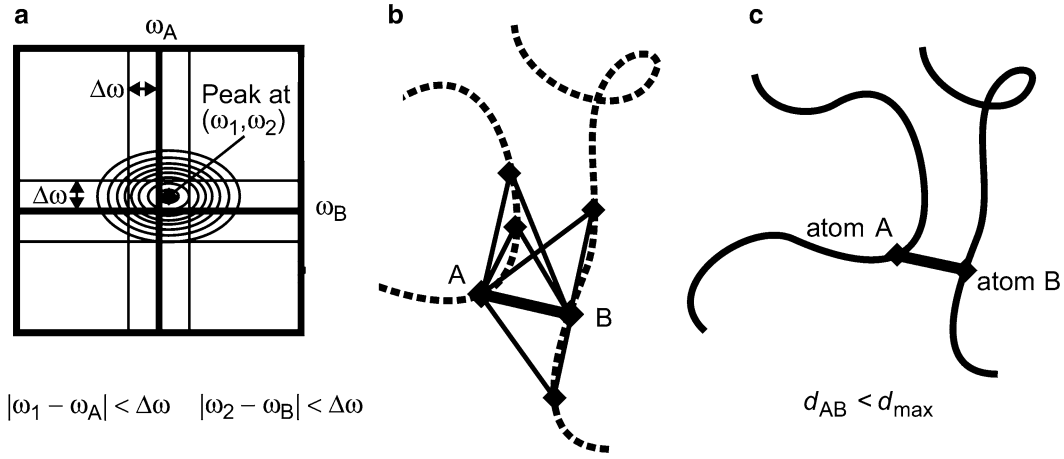

Fig. 4 Three conditions that must be fulfilled by a valid assignment of a NOESY cross peak to two protons A and B in the CYANA automated NOESY assignment algorithm: (**a**) Agreement between the proton chemical shifts ω_A and ω_B and the peak position (ω_1, ω_2) within a tolerance of $\Delta\omega$. (**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 probability P_{network}. Only assignment possibilities for which the product of the three probabilities is above a threshold,

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

are accepted (Fig. 4). 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.

4.2 Ambiguous Distance Restraints

Because of the limited accuracy of chemical shift values and peak positions many NOESY cross peaks cannot be attributed to a single unique spin pair but have an ambiguous NOE assignment comprising multiple spin pairs. Ambiguous distance restraints [61] 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 [53], 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, $d_{\text{eff}} \leq b$, 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.

4.3 Network Anchoring

Each 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 three-dimensional 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, $P_1, P_2, \ldots$, 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 $P_{\text{network}} = 1 - (1 - P_1)\cdot(1 - P_2)\cdots$, which is never smaller than the highest probability of an individual network anchoring contribution.

4.4 Constraint Combination

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" [56]. Ambiguous distance restraints are generated with combined assignments from different, in general unrelated, cross peaks (Fig. 5). 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.

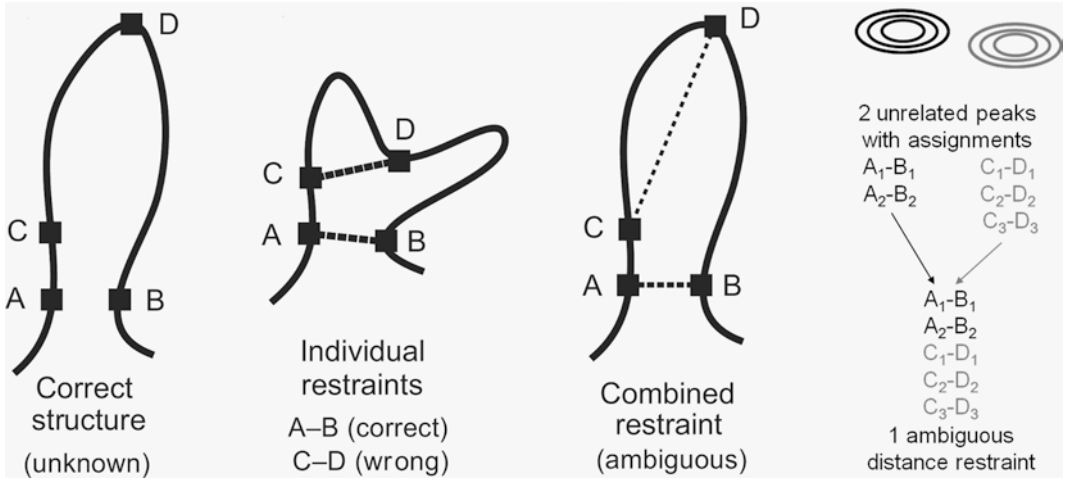

Fig. 5 Schematic illustration of the effect of constraint combination in the case of two distance restraints, a correct one connecting atoms A and B, and a wrong one between atoms C and D. A structure calculation that uses these two restraints as individual restraints that have to be satisfied simultaneously 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*

5 Structure Calculation

The three-dimensional protein structure is calculated using the list of distance restraints, which are obtained from NOESY spectra. Commonly used programs for structure calculation are CYANA [40] (formerly DYANA [41]/DIANA [62]), Xplor-NIH [63], and CNS [64] (formerly X-PLOR [65]), where CYANA is the program most widely used for NMR structure calculation [66].

The most efficient algorithm for the calculation of 3D protein structures from distance restraints, which is also implemented in CYANA, performs simulated annealing by molecular dynamics simulation in torsion angle space. The simulated annealing procedure minimizes a potential energy function, which takes distance restraints, angle restraints, and a repulsive potential into account. Atom distance information from NOESY spectra can be complemented, e.g., by angle restraints, orientational restraints, and hydrogen bond restraints.

A CYANA structure calculation with automated NOE assignment can be completed in less than 1 h for a 10–15 kDa protein, provided that the structure calculations can be performed in parallel, for instance on a Linux cluster system.

In the following, an example of a CYANA macro for a standard combined automated NOE assignment and structure calculation is shown. Executing this macro performs the automated NOE assignment and the structure calculation of a protein.

```
1. peaks := c13.peaks,n15.peaks,aro.peaks
2. prot := demo.prot
3. restraints := demo.aco
4. tolerance := 0.04, 0.03, 0.45
5. structures := 100,20
6. steps := 10000
7. randomseed := 434726
8. cyanalib
9. read seq protein.seq
10. noeassign peaks=$peaks prot=$prot autoaco
```

Line 1: The names of the input NOESY peak lists are specified.

Line 2: The names of the input chemical shift lists are specified. In this case, there is one chemical shift list that is used for all peak lists.

Line 3: The names of additional input restraint files, in this case a file with torsion angle restraints, are specified.

Line 4: Tolerances are set for the NOE assignment calculation, i.e., 0.04/0.03 ppm for hydrogen atoms in the indirect/direct dimensions, and 0.45 ppm for carbon and nitrogen atoms.

Line 5: The numbers of conformers that are calculated (100) and analyzed [20] are specified.

Line 6: The number of torsion angle dynamics steps in the structure calculation is specified.

Line 7: The random number generator seed for generating initial structures is specified.

Line 8: The CYANA standard library is read.

Line 9: The protein amino acid sequence is read from the file "protein.seq".

L*ine 10*: The command noeassign, which is specified in the macro "noeassign.cya" is executed with the given NOESY peak lists and chemical shift list(s) as input. The option "autoaco" specifies that weak torsion angle restraints for the Ramachandran plot and staggered side chain rotamers will be generated and used for the structure calculations.

References

1. Moseley HNB, Montelione GT (1999) Automated analysis of NMR assignments and structures for proteins. Curr Opin Struct Biol 9:635–642
2. Schmidt E (2014) Institute of biophysical chemistry. Goethe University, Frankfurt am Main
3. Güntert P (2011) In: Lian LY, Roberts GCK (eds). Protein NMR spectroscopy: principal techniques and applications. Wiley, Chichester, UK. pp 159–192
4. Schmidt E, Gath J, Habenstein B et al (2013) Automated solid-state NMR resonance assignment of protein microcrystals and amyloids. J Biomol NMR 56:243–254
5. Billeter M, Wagner G, Wüthrich K (2008) Solution NMR structure determination of proteins revisited. J Biomol NMR 42:155–158
6. Gronwald W, Kalbitzer HR (2004) Automated structure determination of proteins by NMR spectroscopy. Prog Nucl Magn Reson Spectrosc 44:33–96
7. Güntert P (2003) Automated NMR protein structure calculation. Prog Nucl Magn Reson Spectrosc 43:105–125
8. Malmodin D, Billeter M (2005) High-throughput analysis of protein NMR spectra. Prog Nucl Magn Reson Spectrosc 46:109–129
9. Muskett FW (2011) In: Lian LY, Roberts GCK (eds). Protein NMR spectroscopy: practical techniques and applications. Wiley, Chichester, UK. pp 5–21
10. 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
11. Pons JL, Malliavin TE, Delsuc MA (1996) Gifa V. 4: a complete package for NMR data set processing. J Biomol NMR 8:445–452
12. Koradi R, Billeter M, Engeli M et al (1998) Automated peak picking and peak integration in macromolecular NMR spectra using AUTOPSY. J Magn Reson 135:288–297
13. Herrmann T, Güntert P, Wüthrich K (2002) Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. J Biomol NMR 24:171–189
14. Goddard TD, Kneller DG (2001) Sparky 3. University of California, San Francisco
15. Johnson BA (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. Methods Mol Biol 278:313–352
16. Johnson BA, Blevins RA (1994) NMR view—a computer program for the visualization and analysis of NMR data. J Biomol NMR 4:603–614
17. Neidig KP, Geyer M, Gorler A et al (1995) Aurelia, a program for computer-aided analysis of multidimensional NMR spectra. J Biomol NMR 6:255–270
18. Vranken WF, Boucher W, Stevens TJ et al (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins 59:687–696

19. Bartels C, Xia TH, Billeter M et al (1995) The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. J Biomol NMR 6:1–10
20. Wüthrich K, Wider G, Wagner G, Braun W (1982) Sequential resonance assignments as a basis for determination of spatial protein structures by high-resolution proton nuclear magnetic resonance. J Mol Biol 155:311–319
21. Guerry P, Herrmann T (2011) Advances in automated NMR protein structure determination. Q Rev Biophys 44:257–309
22. Zimmerman DE, Kulikowski CA, Huang YP et al (1997) Automated analysis of protein NMR assignments using methods from artificial intelligence. J Mol Biol 269:592–610
23. Zimmerman D, Kulikowski C, Wang LZ et al (1994) Automated sequencing of amino acid spin systems in proteins using multidimensional HCC(CO)NH-TOCSY spectroscopy and constraint propagation methods from artificial intelligence. J Biomol NMR 4:241–256
24. Bahrami A, Assadi AH, Markley JL, Eghbalnia HR (2009) Probabilistic interaction network of evidence algorithm and its application to complete labeling of peak lists from protein NMR spectroscopy. PLoS Comput Biol 5:e1000307
25. Bahrami A, Tonelli M, Sahu SC, Singarapu KK et al (2012) Robust, integrated computational control of NMR experiments to achieve optimal assignment by ADAPT-NMR. PLoS One 7:e33173
26. Volk J, Herrmann T, Wüthrich K (2008) Automated sequence-specific protein NMR assignment using the memetic algorithm MATCH. J Biomol NMR 41:127–138
27. Hiller S, Fiorito F, Wüthrich K, Wider G (2005) Automated projection spectroscopy (APSY). Proc Natl Acad Sci U S A 102:10876–10881
28. Jung YS, Zweckstetter M (2004) Mars—robust automatic backbone assignment of proteins. J Biomol NMR 30:11–23
29. Jung YS, Zweckstetter M (2004) Backbone assignment of proteins with known structure using residual dipolar couplings. J Biomol NMR 30:25–35
30. 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
31. 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
32. Fiorito F, Hiller S, Wider G, Wüthrich K (2006) Automated resonance assignment of proteins: 6D APSY-NMR. J Biomol NMR 35:27–37
33. Hiller S, Wider G, Wüthrich K (2008) APSY-NMR with proteins: practical aspects and backbone assignment. J Biomol NMR 42: 179–195
34. Malmodin D, Papavoine CHM, Billeter M (2003) Fully automated sequence-specific resonance assignments of heteronuclear protein spectra. J Biomol NMR 27:69–79
35. López-Méndez B, Güntert P (2006) Automated protein structure determination from NMR spectra. J Am Chem Soc 128:13112–13122
36. Ikeya T, Takeda M, Yoshida H et al (2009) Automated NMR structure determination of stereo-array isotope labeled ubiquitin from minimal sets of spectra using the SAIL-FLYA system. J Biomol NMR 44:261–272
37. 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
38. Ikeya T, Jee J-G, Shigemitsu Y et al (2011) Exclusively NOESY-based automated NMR assignment and structure determination of proteins. J Biomol NMR 50:137–146
39. Schmidt E, Güntert P (2012) A new algorithm for reliable and general NMR resonance assignment. J Am Chem Soc 134:12817–12829
40. Güntert P (2009) Automated structure determination from NMR spectra. Eur Biophys J 38:129–143
41. 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
42. Schmucki R, Yokoyama S, Güntert P (2009) Automated assignment of NMR chemical shifts using peak-particle dynamics simulation with the DYNASSIGN algorithm. J Biomol NMR 43:97–109
43. Ulrich EL, Akutsu H, Doreleijers JF et al (2008) BioMagResBank. Nucleic Acids Res 36:D402–D408
44. Levitt MH (2008) Spin dynamics: basics of nuclear magnetic resonance. Wiley, New York
45. 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
46. 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

47. Meadow 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
48. 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
49. 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
50. 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
51. Linge JP, Habeck M, Rieping W, Nilges M (2003) ARIA: automated NOE assignment and NMR structure calculation. Bioinformatics 19:315–316
52. Rieping W, Habeck M, Bardiaux B (2007) ARIA2: automated NOE assignment and data integration in NMR structure calculation. Bioinformatics 23:381–382
53. 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
54. Huang YJ, Tejero R, Powers R, Montelione GT (2006) A topology-constrained distance network algorithm for protein structure determination from NOESY data. Proteins 62:587–603
55. Gronwald W, Moussa S, Elsner R et al (2002) Automated assignment of NOESY NMR spectra using a knowledge based method (KNOWNOE). J Biomol NMR 23:271–287
56. Herrmann T, Güntert P, Wüthrich K (2002) 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
57. Güntert P (2004) Automated NMR structure calculation with CYANA. Methods Mol Biol 278:353–378
58. Kuszewski J, Schwieters CD, Garrett DS et al (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
59. 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
60. 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
61. 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
62. Güntert P, Braun W, Wüthrich K (1991) Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. J Mol Biol 217:517–530
63. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. J Magn Reson 160:65–73
64. Brünger AT, Adams PD, Clore GM (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr Sect D Biol Crystallogr 54:905–921
65. Brünger AT (1992) X-PLOR, Version 3.1. A system for X-ray crystallography and NMR. Yale University Press, New Haven
66. Williamson MP, Craven CJ (2009) Automated protein structure calculation from NMR data. J Biomol NMR 43:131–143
67. Moseley HNB, Sperling LJ, Rienstra CM (2010) Automated protein resonance assignments of magic angle spinning solid-state NMR spectra of beta 1 immunoglobulin binding domain of protein G (GB1). J Biomol NMR 48:123–128
68. Li KB, Sanctuary BC (1997) Automated resonance assignment of proteins using heteronuclear 3D NMR. 2. Side chain and sequence-specific assignment. J Chem Inf Comput Sci 37:467–477
69. Li KB, Sanctuary BC (1997) 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
70. Leutner M, Gschwind RM, Liermann J et al (1998) Automated backbone assignment of labeled proteins using the threshold accepting algorithm. J Biomol NMR 11:31–43

71. Coggins BE, Zhou P (2003) PACES: Protein sequential assignment by computer-assisted exhaustive search. J Biomol NMR 26:93–111
72. 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
73. Fiorito F, Herrmann T, Damberger FF, Wüthrich K (2008) Automated amino acid side-chain NMR assignment of proteins using ^{13}C- and ^{15}N-resolved 3D [^{1}H,^{1}H]-NOESY. J Biomol NMR 42:23–33
74. Hitchens TK, Lukin JA, Zhan YP et al (2003) MONTE: an automated Monte Carlo based approach to nuclear magnetic resonance assignment of proteins. J Biomol NMR 25:1–9
75. 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
76. Grishaev A, Llinás M (2002) Protein structure elucidation from NMR proton densities. Proc Natl Acad Sci U S A 99:6713–6718
77. 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
78. Hu KN, Qiang W, Tycko R (2011) A general Monte Carlo/simulated annealing algorithm for resonance assignment in NMR of uniformly labeled biopolymers. J Biomol NMR 50:267–276