J Biomol NMR (2015) 62:453–471
DOI 10.1007/s10858-015-9924-9

ARTICLE

Combined automated NOE assignment and structure calculation with CYANA

Peter Güntert[1,2,3] · **Lena Buchner**[1]

Received: 17 February 2015 / Accepted: 17 March 2015 / Published online: 24 March 2015

Abstract The automated assignment of NOESY cross peaks has become a fundamental technique for NMR protein structure analysis. A widely used algorithm for this purpose is implemented in the program CYANA. It has been used for a large number of structure determinations of proteins in solution but was so far not described in full detail. In this paper we present a complete description of the CYANA implementation of automated NOESY assignment, which differs extensively from its predecessor CANDID by the use of a consistent probabilistic treatment, and we discuss its performance in the second round of the critical assessment of structure determination by NMR.

Keywords Automated assignment · NOESY · Distance restraints · Structure calculation · CYANA · CASD-NMR

Introduction

The structure determination of biological macromolecules by NMR in solution relies primarily on distance restraints derived from cross peaks in NOESY spectra. A large number of assigned NOESY cross peaks are necessary to compute an accurate three-dimensional (3D) structure because many of the NOEs are short-range with respect to the sequence and thus carry little information about the tertiary structure and because NOEs are generally interpreted as loose upper bounds in order to implicitly account for internal motions and spin diffusion. Alternatively, accurate distance measurements have become available with eNOEs (Vögeli et al. 2012). Obtaining a comprehensive set of distance restraints from NOESY spectra is in practice not straightforward. The large amount of data, as well as resonance and peak overlap, spectral artifacts and noise, and the absence of expected signals because of fast relaxation turn interactive NOESY cross peak assignment into a laborious and error-prone task, even if it is supported by semi-automated tools that propose and check assignment possibilities (Güntert et al. 1993; Kobayashi et al. 2007; Skinner et al. 2015). Therefore, the development of computer algorithms for automating this often most time-consuming step of a protein structure determination by NMR has been pursued intensely (Guerry and Herrmann 2011). Several algorithms have been developed for the automated analysis of NOESY spectra given the chemical shift assignments, e.g. NOAH (Mumenthaler and Braun 1995; Mumenthaler et al. 1997), ARIA (Nilges et al. 1997; Rieping et al. 2007), ASDP (Huang et al. 2006), KNOWNOE (Gronwald et al. 2002), CANDID (Herrmann et al. 2002a), PASD (Kuszewski et al. 2004), and AutoNOE-Rosetta (Zhang et al. 2014). Automated NOESY peak picking guided by intermediate structures has also been integrated into the method (Herrmann et al. 2002b).

The basic problem of NOESY assignment is the ambiguity of cross peak assignments if only the match between cross peak positions and the chemical shift values of candidate resonances is considered. It has been shown that the number of assignment possibilities based on chemical shift

✉ Peter Güntert
guentert@em.uni-frankfurt.de

[1] Center for Biomolecular Magnetic Resonance, Institute of Biophysical Chemistry, Goethe University Frankfurt am Main, Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany

[2] Laboratory of Physical Chemistry, ETH Zürich, Zurich, Switzerland

[3] Graduate School of Science, Tokyo Metropolitan University, Hachioji, Tokyo, Japan

matching increases exponentially with the uncertainty in the peak and resonance positions. As a consequence, there are in general not a sufficient number of unambiguously assigned distance restraints to obtain a structure (Mumenthaler et al. 1997). Ambiguous distance restraints make it possible to use also NOEs with multiple assignment possibilities in a structure calculation (Nilges 1995). Nevertheless, to minimize the information loss, additional criteria have to be applied to resolve these ambiguities as far as possible, such as using secondary structure information (Huang et al. 2006) or a preliminary structure that is refined iteratively in cycles of NOE assignment and structure calculation (Mumenthaler and Braun 1995). The CANDID automated NOESY assignment method (Herrmann et al. 2002a) introduced the concepts of network anchoring to reduce the initial ambiguity of NOE assignments and constraint combination to reduce the impact of erroneous restraints. In CYANA, the conditions applied by CANDID for valid NOE assignments have been reformulated in a probabilistic framework that is conceptually more consistent and better capable to handle situations of high chemical shift-based ambiguity of the NOE assignments. Its implementation in the software package CYANA will be described in detail in this paper.

Recently, the first round of the CASD-NMR critical assessment of structure determination by NMR initiative (Rosato et al. 2009) evaluated several NMR structure determination methods by blind testing. Using high-quality data sets of small proteins from a structural genomics project it was found that the NOESY-based methods included in the test yielded structures with an accuracy of 2 Å RMSD or better to the subsequently released reference structures (Rosato et al. 2012). In the second part of this paper, we report on the outcome of applying the combined automated NOESY assignment and structure calculation algorithm in CYANA in the second round of the CASD-NMR project.

Algorithm

To introduce the algorithm we first reproduce an overview (Buchner and Güntert 2015b), followed by a detailed description of the input, parameters, implementation and output of the algorithm. In this section the names of CYANA commands, variables, and files are written in italics.

Overview

The algorithm for automated NOE assignment is a re-implementation of principles of the former CANDID procedure (Herrmann et al. 2002a) on the basis of a probabilistic treatment of the NOE assignment process. The key features of the algorithm are network anchoring to reduce the initial ambiguity of NOESY peak assignments, ambiguous distance restraints to generate conformational restraints from NOESY cross peaks with multiple possible assignments, and constraint combination to minimize the impact of erroneous distance restraints on the structure. Automated NOE assignment and the structure calculation are combined in an iterative process that comprises, typically, 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 between the atomic chemical shifts and the peak position is expressed by a Gaussian probability, P_{shifts}. Second, the probability $P_{\text{structure}}$ for agreement with the preliminary structure from the preceding cycle (if available) is computed. Third, 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 restrained 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. Only assignment possibilities for which the product of the three probabilities is above a threshold, $P_{\text{tot}} = P_{\text{shifts}} P_{\text{structure}} P_{\text{network}} \geq P_{\text{min}}$, are accepted. Cross peaks with a single accepted assignment yield a conventional unambiguous distance restraint. Cross peaks with multiple accepted assignments result in an ambiguous distance restraint.

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. CYANA uses "constraint combination" (Herrmann et al. 2002a) to reduce structural distortions from erroneous distance restraints. Medium-range and

long-range distance restraints are incorporated into "combined distance restraints", which are ambiguous distance restraints with assignments taken from different, in general unrelated, cross peaks. A basic property of ambiguous distance restraints is 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. This 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 erroneous NOE assignments on the resulting structure at the expense of a temporary loss of information. It is applied to medium- and long-range distance restraints in, by default, the first two cycles of combined automated NOE assignment and structure calculation with CYANA.

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 assessments relative to the protein 3D structure from the preceding cycle. The precision of the structure determination normally improves with each subsequent 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 use of subsequent refinement and analysis programs that cannot handle ambiguous distance restraints.

Input data

Required input data consists of the protein sequence, one or several 2D, 3D, or 4D NOESY peak lists in XEASY (Bartels et al. 1995) or NMRView (Johnson and Blevins 1994) format, and one or several corresponding chemical shift lists in XEASY or BMRB (Ulrich et al. 2008) format containing the sequence-specific resonance assignments. As an alternative to NOESY, it is possible to use peak lists from solid-state NMR experiments that yield distance information, such as DARR, PDSD, PAIN, etc. (Schütz et al. 2015). Without loss of generality, we will speak about NOE assignment in the following.

The peak lists provide for every peak its position in the spectrum and its volume or intensity. The types of atoms in the 2–4 dimensions (columns) of a peak list must be specified in a format statement, either in the header of the peak list, or as a parameter when reading the peak list into CYANA. The format statement consists of the spectrum type, and an atom label for each spectral dimension. Supported NOESY and solid-state NMR spectrum types are listed in Table 1. For instance, the format statement "C13NOESY HC C H" would describe a peak list of a 3D ^{13}C-resolved NOESY spectrum that stores in the first dimension the chemical shift position of the hydrogen atom directly bound to the carbon atom whose shift is stored in the second dimension, and in the third dimension the chemical shift position of the hydrogen atom that is connected by the NOE to the hydrogen atom in the first dimension. This format can be included in the peak list header with a line "#SPECTRUM C13NOESY HC C H".

The peak positions in the input peak lists must match the chemical shift values in the corresponding chemical shift list(s) within user-defined chemical shift tolerance values (see below). It is possible to use one chemical shift list for

Table 1 Supported NOESY and solid-state NMR experiments

Experiment	Spectrum type	Atom labels
2D [^{1}H,^{1}H]-NOESY	NOESY	H1, H2
3D ^{15}N-resolved [^{1}H,^{1}H]-NOESY	N15NOESY	H, HN, N
3D ^{13}C-resolved [^{1}H,^{1}H]-NOESY	C13NOESY	H, HC, C
4D ^{15}N/^{15}N-resolved [^{1}H,^{1}H]-NOESY	NNNOESY	H1, H2, N2, N1
4D ^{13}C/^{13}C-resolved [^{1}H,^{1}H]-NOESY	CCNOESY	H1, H2, C2, C1
4D ^{15}N/^{13}C-resolved [^{1}H,^{1}H]-NOESY	NCNOESY	HC, HN, N, C
2D DARR	DARR	C1, C2
2D PDSD	PDSD	C1, C2
2D PAIN	PAIN	N, C
3D ^{15}N-resolved PAIN	PAIN3D	C, N, HN

Spectrum type and atom labels are used in the input to CYANA to specify the type and order of dimensions (columns) of an input peak list. The dimensionality of a peak list equals the number of atom labels. The distance-dependent transfer occurs between the atoms in matching the first and second label. If present, the third and fourth labels correspond to the atoms that are covalently bound to the atoms matching the second and first label, respectively

several or all peak lists, or a separate chemical shift list for each peak list. Correlations between peaks in different peak lists are evaluated only via their assignments. It is thus not required that the shift of a given atom is the same in different peak lists but there must be consistency between each peak list and its corresponding chemical shift list. In practice it is still advantageous to use a single chemical shift list for all peak lists, provided that they are aligned well with each other, in order to avoid the burden of maintaining multiple chemical shift lists. The Peakmatch procedure (Buchner et al. 2013) can be used to optimally align multiple peak lists with each other or with a chemical shift list.

Peak assignments are not required in the input peak lists. If nevertheless the input peak lists contain peak assignments, these are by default not used in the assignment process but the algorithm reports whether the automatically determined assignments are consistent with the input peak assignment. Optionally, the user may specify that all or selected input peak assignments are to be kept by the algorithm. Such fixed assignments will not be changed by the algorithm and are used for the network anchoring of other peaks (see below).

In addition to the required input files, it is possible to provide an initialization macro file, named *init.cya* in the current working directory (Fig. 1a). The *init.cya* file contains CYANA commands that are executed automatically when the program is started, for example commands to read a given residue library and protein sequence, or to set CYANA system variables. If present, the *init.cya* macro is also re-executed at the beginning of each NOESY assignment cycle. In the absence of an *init.cya* macro, the program is initialized by reading the standard CYANA residue library (*cyana.lib*) and the most recent sequence file (with extension.seq) in the current working directory.

Optionally, the input may also comprise other, already assigned conformational restraints of any type and format that can be read by CYANA. These must be stored in files named with their respective default file name extension, e.g. additional upper and lower distance bounds (.upl, .lol), torsion angle restraints (.aco), vicinal scalar coupling constants (.cco), residual dipolar couplings (.rdc), pseudocontact shifts (.pcs), etc. Such additional restraints are used together with the automatically assigned NOE distance restraints in the structure calculations. They do not directly enter the NOE assignment algorithm.

To facilitate the automated assignment of NOEs in the special case that a structure of the protein is already known, it is possible to provide an input structure file in PDB format (Berman et al. 2000) named cycle*n*.pdb for cycle n. In this case the calculation will be started with cycle $n + 1$ by making NOESY cross peak assignments based on the structure read from the file cycle*n*.pdb, which may, for instance, be a crystal structure of the protein (with hydrogen atoms attached), an NMR structure bundle, or a homology model.

Commands and parameters

The combined automated NOE assignment and structure calculation with CYANA is executed by the *noeassign* command, which is implemented as a script in the INCLAN command language (Güntert et al. 1992) that can be modified by the user, if necessary. In the following, we refer to INCLAN scripts as *macros*. The general *noeassign* command is normally called through a short calculation-specific macro that specifies all user-defined parameters (Fig. 1b), such that the *noeassign* macro does not have to be changed by the user. The *noeassign* command in turn invokes directly or indirectly various lower-level CYANA commands, which are listed in Table 2.

The *noeassign* command has two required parameters: The *peaks* parameter specifies the names of the input peak list file(s), and the *prot* parameter specifies the names of the chemical shift list file(s). If there are several chemical shift lists, they must be given in the order corresponding to the peak lists. Additional optional parameters are: *format*, to specify, in the same order as the peak lists, the formats (see above) of the peak lists (by default, this information is expected to be included in the headers of the peak list files); *cycles* = *m*–*n*, to specify the calculation cycles to be performed (default: cycles = 1–7); *combination* = *m*–*n*, to specify the calculation cycles in which constraint combination (see below) is applied (default: *combination* = 1–2); *keep*, to specify the name of a macro that contains statements to select peaks whose input assignment is to be kept by the algorithm (by default no input assignments are kept); *calculation*, to specify the name of a macro that performs the structure calculation (default: *calculation* = *structcalc*, where *structcalc* is a standard CYANA macro to calculate a group of conformers starting from random initial structures). In addition the *noeassign* command has a number of options that can be selected by adding the following keywords to the command line: *autoaco* for the automatic generation of temporary torsion angle restraints to favor the allowed regions of the Ramachandran plot and staggered rotamers for torsion angles between tetrahedrically coordinated atoms (see below), *multiple* to allow for multiple ambiguous assignments in the final distance restraint list, *stereoexpand* to replace in the final structure calculation assignments to not stereospecifically assigned diastereotopic atoms by the corresponding pseudoatoms, and *details* to produce more detailed output, as described below.

a *init.cya:*

```
cyanalib                                    # read CYANA residue library
read seq demo.seq                           # read protein sequence
rmsdrange := 5-75                           # residue range for RMSD calculations
```

b *CALC.cya:*

```
peaks      := n15.peaks,c13.peaks,aro.peaks     # names of peak lists
prot       := demo.prot                         # names of chemical shift list(s)
restraints := demo.aco,demo.rdc                 # additional restraints
tolerance  := 0.04,0.03,0.45                    # shift tolerances: H, H', C'/N'
structures := 100,20                            # number of initial, final conformers
steps      := 10000                             # number of dynamics steps
randomseed := 434726                            # random number generator seed

noeassign peaks=$peaks prot=$prot autoaco
```

Fig. 1 Examples of CYANA macro file for a combined automated assignment and structure calculation run. **a** Initialization macro, *init.cya*, that contains commands to read the standard CYANA residue library and the sequence of the protein. **b** *CALC.cya* macro that specifies all user-defined parameters for a combined automated assignment and structure calculation run. In this example, the calculation uses three input peak lists from a 3D ^{15}N-resolved NOESY (n15.peaks), a 3D aliphatic ^{13}C-resolved NOESY (c13.peaks), and a 3D aromatic ^{13}C-resolved NOESY (aro.-peak), a single chemical shift list that applies to all three NOESYs (demo.prot), torsion angle restraints (demo.aco), and residual dipolar coupling restraints (demo.rdc). Multiple files are specified by a comma-separated list of filenames without intervening blanks. The tolerance for chemical shift matching is set to 0.04 ppm for the "free" ^{1}H dimension in the NOESY peak lists, 0.03 ppm for the ^{15}N- or ^{13}C-bound ^{1}H dimension, and 0.45 ppm for the ^{15}N or ^{13}C dimensions. Structure calculations are started from 100 conformers with random torsion angle values, and the 20 conformers with lowest final target function value are analyzed. Each conformer is calculated with 10,000 torsion angle dynamics steps. The random number generator for setting random torsion angle values and initial velocities for torsion angle dynamics is initialized with a seed value of 434726. In the last line the *noeassign* command is called with the above-specified peak and chemical shift lists and the option *autoaco* for the automatic generation of temporary torsion angle restraints to favor the allowed regions of the Ramachandran plot and staggered rotamers for torsion angles between tetrahedrically coordinated atoms

Table 2 CYANA commands used in the context of automated NOE assignment

Command	Description
noeassign	High-level command for a complete structure determination with 7 cycles and a final structure calculation
assign[a]	Low-level command for NOE assignment, called by *noeassign*
calibration	High-level command for reading chemical shift lists and peak lists, and converting peak volumes/intensities into upper distance bounds
distance split[a]	Split ambiguous distance restraints into unambiguous ones
distance combine[a]	Apply constraint combination
peakcheck	Read and check consistency of peak and chemical shift lists
cisprocheck	Check for *cis*-Pro based on $^{13}C^{\beta}/^{13}C^{\gamma}$ chemical shifts
molecules symmetrize[a]	Add symmetry-related distance restraints for symmetric multimers
ramaaco	Add ϕ/ψ torsion angle restraints to favor Ramachandran plot
rotameraco	Add side-chain torsion angle restraints to favor staggered rotamers
structcalc	High-level command to run structure calculation
calc_all	Calculate a bundle of structures
structure swap[a]	Determine and apply structure-based stereospecific assignments
savestereo	Save stereospecific assignments
distance modify[a]	Symmetrize distance restraints with not stereospecifically assigned atoms; discard duplicate and meaningless restraints
distance stereoexpand[a]	Replace not stereospecifically assigned atoms by pseudoatoms
overview	Print overview table of target function values, violations, RMSDs, etc
ramaplot	Generate Procheck Ramachandran plot

[a] Commands implemented in the Fortran source code of CYANA. All other commands listed in this table are implemented as scripts (macros) in the INCLAN command language of CYANA, and can be modified by the user

Consistency checks

Before starting the NOE assignment, the *noeassign* command checks the input peak lists and chemical shift lists for possible inconsistencies. First, the completeness of the ^{1}H chemical shift assignments is evaluated. Carbon-bound and backbone amide hydrogens with missing chemical shift assignments are listed, and the percentage of assignment completeness for these nuclei is reported. A high degree of completeness is crucial for successful automated NOE assignment with CYANA (Buchner and Güntert 2015b; Herrmann et al. 2002a; Jee and Güntert 2003). If the input peak lists contain assigned peaks, then their position is compared with the chemical shifts of the atoms to which they are assigned, and deviations that exceed the defined chemical shift tolerances are listed. Similarly, deviations of the chemical shift values for the same atom in different chemical shift lists are reported, as well as chemical shifts that deviate by more than 4 standard deviations from their average value in the BMRB chemical shift statistics (Ulrich et al. 2008), which is stored in the CYANA residue library.

The consistency between the *cis*/*trans* proline declarations in the sequence file, where *cis* proline is declared as cPRO and *trans* proline as PRO, and the $^{13}C^{\beta}/^{13}C^{\gamma}$ chemical shift values in the (first) input chemical shift list are checked using the empirical correlation between *cis* and *trans* peptide bonds and the difference between the $^{13}C^{\beta}$ and $^{13}C^{\gamma}$ chemical shifts (Schubert et al. 2002). To this end, the relative probabilities for *cis* and *trans* proline are calculated as $P_{cis} = p_{cis}/(p_{cis} + p_{trans})$ and $P_{trans} = p_{trans}/(p_{cis} + p_{trans})$, where

$$p_k = \frac{1}{\sqrt{2\pi}\sigma_k}\exp\left(-\frac{1}{2}\left(\frac{\omega(^{13}C^{\beta}) - \omega(^{13}C^{\gamma}) - \mu_k}{\sigma_k}\right)^2\right)$$

for $k = cis, trans$. Here $\omega(^{13}C^{\beta})$ and $\omega(^{13}C^{\gamma})$ are the proline $^{13}C^{\beta}$ and $^{13}C^{\gamma}$ chemical shift values, $\mu_{cis} = 9.64$ ppm, $\sigma_{cis} = 1.27$ ppm, $\mu_{trans} = 4.51$ ppm, and $\sigma_{trans} = 1.17$ ppm (Schubert et al. 2002). Consistency or inconsistency with the sequence file is reported if the relative probability for one of the forms is greater than 90 %, i.e. if either $P_{cis} > 0.9$ or $P_{trans} > 0.9$.

Covalently restrained short distances

The network anchoring of NOE assignments (see below) uses a list of short, covalently restrained distances, which is, by default, generated internally by the algorithm. To this end, 100 conformers of the protein are minimized by the variable target function method (Braun and Go 1985; Güntert et al. 1991a) without experimental restraints, considering only the steric repulsion. The 20 conformers with the lowest target function values are selected and all maximal distances in the 20 conformers between assigned atoms that are intraresidual and shorter than 9 Å or sequential and shorter than 5 Å are stored in the file *cycle0.upl* for later use in network anchoring. To distinguish the entries in this file from normal distance restraints, their relative weighting factor is set to zero. Alternatively, the user can also provide the *cycle0.upl* file explicitly, which can be used, for instance, to input a priori structural information into the network anchoring.

Calibration of upper distance bounds

Cross peaks are read and calibrated by CYANA's *calibration* command that assumes a $V = C/u^6$ relationship between the peak volume V and the upper distance bound u. The calibration constant C for each peak list is determined automatically by default, or can be specified explicitly by setting the variable *calibration_constant* to a comma-separated list of the calibration constants, given in the same order as the peak lists in the *peaks* parameter of the *noeassign* command. The automatic determination of the calibration constant sets C such that for each peak list the median peak volume corresponds to a user-defined reference distance, d_{ref}, which has a default value of 4.0 Å. Larger values of d_{ref} result in higher upper distance bounds and can be appropriate with high-quality NOESY spectra that reveal weak NOEs for longer distances than usual. Using the median renders the method insensitive to the presence of (often artifactual) very strong and/or very weak peak volumes. To avoid too short/long upper distance bounds for peaks with very strong/weak volumes, the upper distance bounds are restricted to minimal/maximal values given by the variable *upl_values*, by default upl_values := 2.4, 5.5 Å, i.e. regardless of the peak volume, the distance bound will never be set smaller than 2.4 Å or larger than 5.5 Å. The variable *upl_values* may also be set to more than two comma-separated values. In this case, only these discrete values will be used for the upper distance bounds. This can be used, for instance, to implement weak/medium/strong classes of NOEs rather than continuous "calibration curves".

Note that the values of the upper distance bounds depend only on the peak volume or intensity, not on the (initially unknown) assignments. Except for the possible application of distance bound "elasticity" (see below), they remain invariable throughout all cycles. It is physically more correct to use peak volumes rather than intensities to derive upper distance bounds. Nevertheless, it is often more robust to use peak intensities, whose measurement is in practice less affected by overlap than the determination of peak volumes. Furthermore, it was shown that errors up to 150 % in peak volumes/intensities have in general no significant effect on the resulting structures (Buchner and

Güntert 2015b). NOEs that involve groups of atoms with degenerate chemical shifts, e.g. methyl groups, are calibrated using the same value of the calibration constant C as other NOEs in the same spectrum. The fact that they represent interactions from multiple protons is taken into account by automatically "expanding" such restraints into ambiguous distance restraints among all corresponding protons during the structure calculation.

Cycles

Typically seven cycles of automated NOE assignment followed by structure calculation are performed. At the start of each cycle the program is reinitialized, the chemical shift list(s) and original peak lists are loaded and calibrated, and the structure from the preceding cycle is read, except in the first cycle, cycle 1, which is performed without initial structure. The structure is the only information that is taken from the preceding cycle. In particular, the peak assignments and NOE distance restraints from the previous cycle are not used. The *assign* command (see below) is used to assign the NOESY peaks, except those that are to be kept as in the input peak lists according to the *keep* parameter. Assigned NOESY cross peaks yield distance restraints that are used, possibly in conjunction with other independently read restraints, to calculate a bundle of conformers by simulated annealing with the CYANA torsion angle dynamics algorithm. A successful *noeassign* run yields already in the first cycle a structure bundle with an RMSD radius of <3 Å (Buchner and Güntert 2015b; Herrmann et al. 2002a; Jee and Güntert 2003). (The RMSD radius of a structure bundle is defined as the average RMSD for the backbone atoms of the well-defined region(s) (Kirchner and Güntert 2011) between the individual conformers and the mean coordinates of the bundle.) This possibly imprecise structure is then refined in subsequent cycles. To this end various parameters of the algorithm are changed between cycles (Table 3). In particular, the maximal acceptable violation for an assignment is decreased from 1.5 Å in cycle 1 to 0.1 Å in the last cycle to reflect the increasing expected accuracy of the structures. In the first two cycles constraint combination (see below) is applied to alleviate the impact of erroneous assignments on the structure calculation.

Output data for each cycle n comprise four files: (1) cycle$n$.noa: Assignment details about every NOESY peak. (2) cycle$n$.upl: NOE upper distance bounds obtained in cycle n. (3) cycle$n$.pdb: Structure bundle obtained in cycle n. (4) cycle$n$.ovw: Overview table for the structure calculation in cycle n. Additional files are written in the last cycle, usually cycle 7, or optionally for all cycles if the *noeassign* command is called with the option *details*: (5) A-cycle7.peaks: Assigned peak lists, where A is the name of an input peak list. A separate file is written for each input peak list. These output peak lists normally contain peaks with ambiguous assignments. They can be read by CYANA but in general not by XEASY. (6) A-cycle7-ref.peaks: Copy of the input peak list A in which the XEASY color code (the integer number in the column following the columns with chemical shift values) of a peak is set to 1, if the assignments made by the program are compatible with the assignment (if present) in the input peak list, 2, if the two assignments are incompatible, 3, if the peak is assigned by the algorithm but was not assigned in the input peak list, or 4, if the peak is unassigned. Apart from the color code, the input peak list is left unchanged. The input assignment of a peak, if present, is only used for comparison, not for assigning the peak in the automated procedure. The A-cycle7-ref.peaks files are readable by XEASY and other programs and can be used to visualize the results of the automated NOE assignment in the spectra. If the option *details* is set, also XEASY assignment files named A-cycle$n$.assign are written, if the input peak lists were in XEASY format, and, in cycles with constraint combination (normally cycles 1–2), the NOE distance restraints before applying constraint combination are saved as cycle$n$-uncombined.upl. If the *autoaco* option is set, then the temporary torsion angle restraints to favor allowed regions of the Ramachandran plot

Table 3 Cycle-dependent parameters of automated NOE assignment

Parameter	Value in cycle							Description
	1	2	3	4	5	6	7	
Δd (Å)	–	1.5	0.9	0.6	0.3	0.1	0.1	Maximal acceptable violation
$P_{\min}$	0.1	0.2	0.2	0.2	0.2	0.2	0.2	Probability threshold for acceptable assignments
$Q_{\min}$	0.45	0	0	0	0	0	0	Quality threshold for peaks
c	0.5	1	1	1	1	1	1	Confidence for network-anchoring contributions
f	1	1	1.25	1.25	1.25	1.25	1.25	Bounds elasticity: maximal increment factor
Combination	Yes	Yes	No	No	No	No	No	Apply constraint combination
Split	No	No	No	No	No	No	Yes	Split ambiguous restraints into unambiguous ones

and staggered rotamer positions are saved in the file cycle.aco, which is used in all cycles.

Ambiguous distance restraints

Ambiguous distance restraints (Nilges 1995) provide a powerful concept for handling ambiguities in the NOESY cross peak assignments. This is particularly important at the outset of a structure determination because the large majority of NOEs cannot be assigned unambiguously from chemical shift information alone (Mumenthaler et al. 1997). It is thus in general not possible to calculate a well-defined structure only from the initially unambiguous NOEs. Ambiguous distance restraints allow using the information contained in NOEs with multiple assignment possibilities in an unbiased way. 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 u 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, $d_{\text{eff}} \leq u$, with the "effective" or "r^{-6}-summed" distance

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

where 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.

The *assign* command for automated NOE assignment

The CYANA *assign* command performs automated assignment of the NOESY cross peaks on the basis of the given chemical shifts, knowledge of covalently restrained short distances, and the selected 3D conformers, if available. The low-level *assign* command is called in each cycle by the high-level *noeassign* macro.

The overall assignment strategy is as follows: First all assignment possibilities of a peak are generated on the basis of the chemical shift values that match the peak position within the tolerance defined by the CYANA variable *tolerance*, and a probability P_{shifts} for the chemical shift-based assignment is computed. Second, the probability $P_{\text{structure}}$ for agreement with a bundle of conformers, in general the structure from the previous cycle, if present, is computed as the fraction of the conformers in which the corresponding distance is shorter than the upper distance bound plus an acceptable violation, and assignment possibilities for which the product $P_{\text{shifts}} \times P_{\text{structure}}$ of these two probabilities is below the required probability threshold P_{min} are discarded. In the absence of a structure, e.g. in cycle 1, $P_{\text{structure}} = 1$. 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 restrained distances. The network anchoring probability P_{network} that the distance corresponding to an assignment is shorter than the upper distance bound is computed given the assignments of the other peaks but independent from knowledge of the 3D structure. Only assignment possibilities k for which the total probability P_k given by the product $P_{\text{shifts}} \times P_{\text{structure}} \times P_{\text{network}}$ of the three probabilities is above the required probability threshold P_{min}, are accepted. In addition, if a peak has safe short-range assignments, i.e. assignments corresponding to a covalently restrained distance that is shorter than the upper distance bound for the peak, then only those are retained and all others discarded. This is analogous to the approach of preferring obvious short-range assignments during manual NOE assignment. Next, the overall quality $Q = 1 - \prod_k (1 - P_k)$ of the entire assignment of a peak is computed from the probabilities P_k of its individual accepted assignment possibilities. The overall quality of a peak assignment is always at least as large as the highest probability of an accepted assignment possibility. Peaks are kept assigned only if their quality exceeds the quality threshold Q_{min}.

Parameters of the *assign* command are given in Table 4. The *assign* command makes assignments, if possible, for all peaks except those that were selected to be kept with their input assignment (according to the *keep* parameter of the *noeassign* command) and provides a report including details on the assignment of each individual peak (Fig. 2) and a summary table. If peaks are assigned on input (and not selected to be kept with their input assignment), the *assign* command

Table 4 Parameters of the *assign* command for automated NOESY assignment

Parameter	Default value	Description
matchfactor	0.5	Weighting factor Γ for chemical shift-based NOE assignment
link	1	Non-NOE link(s) in 3D/4D peaks: covalent (1) or intraresidual (2)
violation	–1.0 Å	If positive, cutoff for acceptable distance restraint violations
alignfactor	0.5	Weighting factor δ for peak alignment in network anchoring
pathlength	3	Maximal path length in network anchoring
confidence	1.0	Weight for relative assignment probabilities in network anchoring
distance	10.0 Å	Distance cutoff for storing assignment possibilities
probability	0.2	$P_{\min}$, threshold for total probability of an assignment
quality	0.5	$Q_{\min}$, threshold for quality factor Q of a peak
elasticity	1.0–1.0	Minimal/maximal factor for upper bound adaptation by "elasticity"
changevol	False	Change volume in peak list if upper bound elasticity is applied?
prefer	∞	Max. residue range for which intramolecular assignments are preferred
interrange	0–∞	Minimal/maximal residue range for intermolecular assignments
unassigned	0.1 ppm	Shift uncertainty above which an atom is considered as "unassigned"
short	False	Restrict NOE assignments with unassigned atoms to short-range

reassigns them without considering the input assignment but includes in the report a comparison between the input assignment and the new assignment made by the *assign* command.

Chemical shift-based NOE assignment

For the chemical shift-based assignment of a peak at position (ω_1,…, ω_D) in a D-dimensional spectrum, all assignment possibilities to atoms (α_1,…, α_D) are generated that fulfill the conditions $|\omega_k - \omega(\alpha_k)| \leq \Delta\omega_k$ for $k = 1,\ldots, D$, where $\omega(\alpha_k)$ is the chemical shift of atom α_k in the corresponding chemical shift list, and $\Delta\omega_k$ the applicable chemical shift tolerance. In addition, for valid assignment possibilities the atoms α_k must be of the type of nucleus that is detected in spectral dimension k, and atoms in spectral dimensions linked by through-bond magnetization transfer must be covalently bound to each other. The applicable chemical shift tolerance $\Delta\omega_k$ is the maximum of (1) the general chemical shift tolerance for dimension k, given by the CYANA variable *tolerance*, (2) the peak list-specific shift tolerance for dimension k, given optionally in the header of the peak list by a line "#TOLERANCE $\Delta\omega_1$ … $\Delta\omega_D$" (set to zero if absent), and (3) the chemical shift error for atom α_k, specified in the chemical shift list.

For each assignment possibility the closeness of the chemical shift match is quantified by

$$\chi^2 = \sum_{k=1}^{D} \frac{(\omega_k - \omega(\alpha_k))^2}{\Delta\varpi_k^2 + \Delta\omega_k^2},$$

where $\Delta\varpi_k$ is the maximum of the general and the peak list-specific shift tolerance for dimension k. The probability for chemical shift matching is calculated as

$$P_{\text{shift}} = Q\left(\frac{D}{2}, \frac{\chi^2}{2\Gamma^2}\right)$$

and stored for each assignment possibility. Here, Q denotes the regularized gamma function (not to be confused with the peak quality factor Q elsewhere in this paper) (Press et al. 1986), and Γ is the value of the parameter *matchfactor* of the *assign* command, with default value 0.5. Assignments with $P_{\text{shift}} < P_{\min}$ are discarded.

If the atom-specific chemical shift tolerance for an atom involved in an NOE is larger than the value of the parameter *unassigned* of the *assign* command, the shift is considered as "unassigned". As a safeguard to avoid NOE assignments with an extremely high degree of ambiguity, NOE assignments between two "unassigned" atoms are omitted. In addition, if the option *short* of the *assign* command is set, NOE assignments that involve an "unassigned" atom are restricted to sequentially short-range ones.

Structure-based NOE assignment

If a structure bundle is available, e.g. from the previous cycle, a structure-based probability $P_{\text{structure}}$ is calculated for each assignment possibility by counting the number of conformers in the structure bundle where the distance is shorter than the upper bound plus an acceptable violation Δd. If n out of a total of N conformers fulfill the upper distance bound, then $P_{\text{structure}} = n/N$. The acceptable violation is given by the value of the parameter *violation* of the *assign* command, which is in turn set for each cycle by the *noeassign* command according to Table 3. In earlier cycles high acceptable violations are used to account for the generally lower accuracy of the early structures, whereas in this last cycle this parameter is decreased to almost zero.

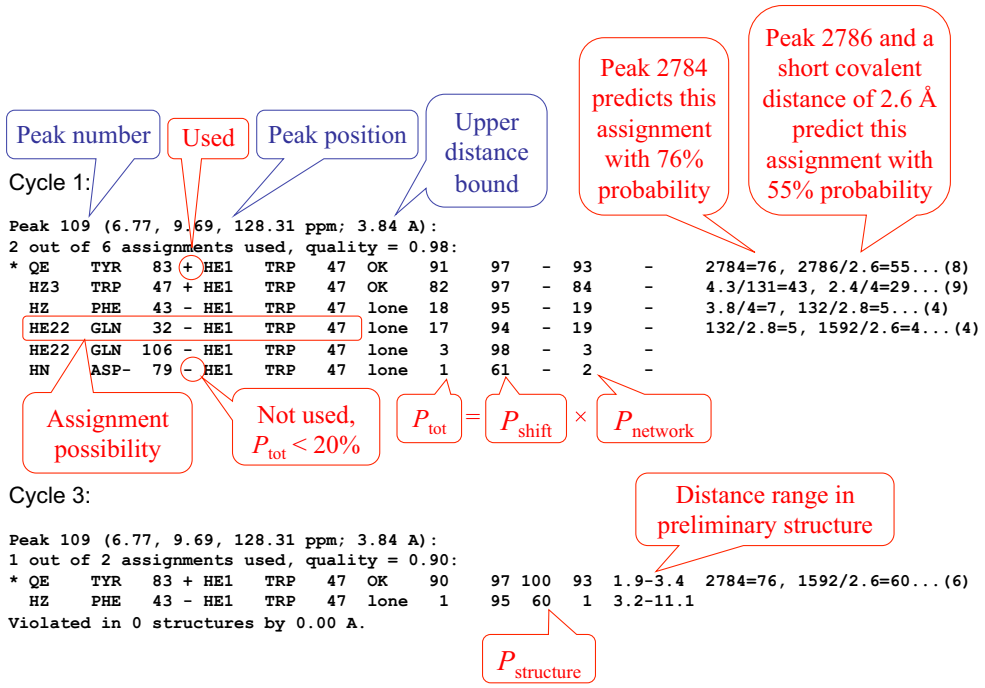

Fig. 2 Example assignment reports for a NOESY cross peak, generated by the CYANA command *assign* in cycle 1 and cycle 3 of combined automated NOESY assignment and structure calculation *Line 1:* Peak number, peak list, peak position, upper distance bound. *Line 2:* Number of used assignments, number of assignment possibilities, overall quality Q of the peak assignment. *Following lines:* Individual assignment possibilities. The following data is given from *left to right*: (1) Flag that indicates the input assignment, if present, by an *asterisk* if it is among the used assignments, or by an *exclamation mark* otherwise. (2) First atom, identified by its name, residue name, and residue number. (3) Flag to indicate whether the assignment possibility was used *plus sign* or not used *minus sign* in the distance restraint generated from this peak. (4) Second atom, identified by its name, residue name, and residue number. (5) Decision on the assignment possibility: "OK", good assignment with probability above the probability threshold P_{min}; "lone", network anchoring based probability too low ($P_{network} < P_{min}$). (Additional possibilities, not present in the figure, are: "far", structure-based probability too low ($P_{structure} < P_{min}$); "poor", individual probabilities above P_{min} but total probability too low. Note also that an assignment with decision "OK" may still not have been used (indicated by the flag '–' in column 3) because either the overall quality of the peak is too low ($Q < Q_{min}$), or because a sufficiently good short-range assignment is present. This case does not occur in the figure. (6) Total probability (in %) for the assignment possibility. (7) Probability (%) for match between peak position and chemical shifts (P_{shifts}). (8) Probability (%) for agreement with input structure bundle ($P_{structure}$), or a hyphen in *cycle 1*, where a structure is not available yet. (9) Probability (%) derived from network anchoring ($P_{network}$). (10) Minimal and maximal distance in the structure bundle, normally from the preceding cycle (Å), or a hyphen in cycle 1, where a structure is not available yet. (11) Most important individual contributions to the network anchoring based probability, ordered by decreasing size. For each contribution listed, the number after the equal sign is the probability (%) for the contribution identified in front of the equal sign, as follows (only the second and third possibilities appear in the example): a real number r indicates a covalently restrained distance shorter than r Å; an integer number indicates the peak number of a (symmetrically related) peak with the same assignment; an integer and a real number (i/r, or vice versa) indicate that the peak with number i connects the first atom to a third atom whose distance from the second atom is covalently restrained to be shorter than r Å, or vice versa; two integer numbers (i/j) indicate the numbers i and j of two peaks that relate the two atoms of the present assignment through a third atom; an integer preceded by a tilde ($\sim i$) indicates that the peak with number i connects two atoms that are covalently restrained to be <3.5 Å from the first and second atom of the present assignment possibility, respectively. For reasons of space, an *ellipsis* followed by the total number of contributions in parenthesis indicates that not all contributions with probability >1 % are printed. *Last line for cycle 3* (not present for cycle 1, where a structure is not available yet): Number of conformers in which the upper distance limit of the ambiguous distance restraint formed by the accepted assignments (marked by *plus sign* in the preceding lines) is violated by more than the maximally acceptable violation, and the average size of the violation

Network anchoring

Checking assignment possibilities against a structure is a straightforward and powerful method to eliminate incorrect assignments and to greatly reduce the peak assignment ambiguity. In the first cycle of a structure determination this is not possible because no structure is available yet. Network-anchoring of NOE assignments (Güntert 2009; Herrmann et al. 2002a) is a technique to partially replace the information that could be gained from a structure by exploiting the fact that correct NOE assignments must correspond to a self-consistent set of distance restraints. In the present algorithm network anchoring is implemented by estimating, on the basis of the assignment possibilities of the other NOESY cross peaks, a probability $P_{network}$ that the distance for a given assignment between two atoms α

and β is shorter than the corresponding upper distance bound u obtained from the peak volume, i.e. that a restraint with this assignment could be fulfilled in the context of all other NOE distance restraints. To this end, connections between the two atoms that result from other peaks or from the covalent structure are considered. From each such connection a probability $P_k = P(d_{\alpha\beta} \leq u)$ that the distance between the atoms α and β is shorter than the upper bound u is derived without reference to a 3D structure, and these probabilities are combined into the network-anchoring probability P_{network} for the assignment of the given peak to atoms α and β as

$$P_{\text{network}} = 1 - \prod_k (1 - P_k),$$

where the product extends over all indirect connections and includes an additional term with $k = 0$ for the a priori probability P_0 to fulfill the restraint in the absence of any information from other peaks. $P_0 = (u/R)^3$, where $R = (20N_{\text{a}}\ 3/4\pi)^{-1/3}$ with N_{a} equal to the total number of atoms, is used as a rough estimate of the radius of the protein in Å. Note, that an indirect connection with vanishing probability does not affect the network-anchoring probability P_{network}, which is always at least as high as the highest P_k, and, because of the a priori probability P_0, never zero. The following three types of connections are taken into account for computing the network-anchoring probability P_{network} for an assignment between atoms α and β:

1. A different peak with an assignment possibility to the same atoms α and β, or a covalently restrained distance between atoms α and β. In this case the probability $P(d_{\alpha\beta} \leq u)$ is estimated from the other peak as

$$P\left(d_{\alpha\beta} \leq u\right) = P'^{(\text{rel})} \pi_\alpha \pi_\beta \min\left(1, (u/u')^3\right),$$

where u' is the upper distance bound of the other peak or from the covalently restrained distance, $P'^{(\text{rel})}$ is the relative probability of the assignment possibility of the other peak among all assignment possibilities of that peak, π_α and π_β are probabilities for the matching of the positions of the two peaks with respect to atoms α and β, respectively. $P'^{(\text{rel})} = \pi_\alpha = \pi_\beta = 1$ for a covalently restrained distance. The term $(u/u')^3$ represents the ratio of the volumes of two spheres with radii u and u' to estimate the probability that the distance $d_{\alpha\beta}$ is shorter than u, given that it is shorter than u'. To calculate π_α, the match of the two peak positions with respect to atom α is quantified by $\chi^2 = (\omega - \omega')^2/(\Delta\omega^2 + \Delta\omega'^2)$, where ω and ω' are the chemical shift coordinates of atom α in the two peaks, and $\Delta\omega$ and $\Delta\omega'$ are the maximum of the general and the peak list-specific chemical shift tolerances (see above) in the dimensions corresponding to atom α in the two peaks. If atom α is covalently linked to a heavy atom that is detected in both peaks (in 3D or 4D NOESY spectra), an analogous term for the matching of this heavy atom is added to the above χ^2 value. The probability π_α is then computed as $\pi_\alpha = Q(m/2, \chi^2/2\delta^2)$, where Q denotes the regularized gamma function (see above), m is the number of shifts involved in the alignment (1 or 2), and δ is the value of the parameter *alignfactor* of the *assign* command, with default value 0.5. π_β is calculated in the same way.
2. Two different peaks with assignment possibilities to atoms α and γ and to atoms γ and β, where γ is an arbitrary third atom. A covalently restrained distance may replace one of the two peaks, but not both. In this case the probability $P(d_{\alpha\beta} \leq u)$ is estimated from the two other peaks as

$$P\left(d_{\alpha\beta} \leq u\right) = P_{\alpha\gamma}^{(\text{rel})} P_{\gamma\beta}^{(\text{rel})} \pi_\alpha \pi_\beta \pi_\gamma P\left(u_{\alpha\gamma}, u_{\gamma\beta}, u\right),$$

where $u_{\alpha\gamma}$ is the upper distance bound of the other peak assigned to atoms α and γ, $u_{\gamma\beta}$ is the upper distance bound of the other peak assigned to atoms γ and β, π_α, π_γ, π_β are defined as in the preceding paragraph, $P_{\alpha\gamma}^{(\text{rel})}$ and $P_{\gamma\beta}^{(\text{rel})}$ are the relative assignment probabilities of the two assignment possibilities, and $P(a, b, u)$ is an estimate of the probability that the distance $d_{\alpha\beta}$ is shorter than u, given that $d_{\alpha\gamma} \leq a$ and $d_{\gamma\beta} \leq b$, which is computed as

$$P(a,b,u) = \begin{cases} \dfrac{(a^2+4ab+b^2)(a-b)^4 - 9(a^2-b^2)^2u^2 + 16(a^3+b^3)u^3 + (u^2-9(a^2+b^2))u^4}{32a^3b^3} & \text{if } u < a+b \\ 1 & \text{otherwise} \end{cases}$$

To derive this expression, one considers the intersection volume of two spheres with radii u and b and a distance x between their centers,

$$V(u,b,x) = \begin{cases} \dfrac{\pi(b+u-x)^2(6bu-3(b^2+u^2)+2(b+u)x+x^2)}{12x} & \text{if } |b-u| \le x \le b+u \\ \dfrac{4\pi}{3}\min(b,u)^3 & \text{if } x \le |b-u| \\ 0 & \text{otherwise} \end{cases}$$

Integrating $V(u, b, x)$ over all positions $\boldsymbol{x}$ within a sphere of radius a provides the desired probability, i.e.

$$P(a,b,u) = \frac{1}{\frac{4\pi}{3}a^3\frac{4\pi}{3}b^3}\int_{|\boldsymbol{x}|\le a} V(u,b,|\boldsymbol{x}|)d^3x = \frac{4\pi}{\frac{4\pi}{3}a^3\frac{4\pi}{3}b^3}\int_0^a V(u,b,x)x^2dx,$$

which yields the above expression for $P(a, b, u)$.

3. A different peak with an assignment possibility to atoms α' and β', where α' and β' are two atoms with a covalently restrained distance $\le$3.5 Å from atoms α and β, respectively. Typical examples for this situation are two mutually supportive interstrand NOEs in a β-sheet. In this case the probability $P(d_{\alpha\beta} \le u)$ is estimated as

$$P\left(d_{\alpha\beta} \le u\right) = P_{\alpha'\beta'}P\left(u_{\alpha\alpha'} + u_{\beta\beta'}, u_{\alpha'\beta'}, u\right),$$

where $u_{\alpha\alpha'}$ and $u_{\beta\beta'}$ are the upper bounds for the covalently restrained distances between α and α' and β and β', respectively. Only such atoms α and β' for which $u_{\alpha\alpha'} \le 3.5$ Å and $u_{\beta\beta'} \le 3.5$ Å are considered.

In principle, it would be possible to consider further indirect contacts for network anchoring. However, this is not done in the present algorithm because of the limited additional information that could be gained, and because of the considerable memory and computation time requirements to search for them. In exceptional cases, e.g. if large chemical shift tolerances are applied to a large protein, it can become necessary to limit the search for network anchoring connectivities. This can be achieved by setting the *pathlength* parameter of the *assign* command to a value smaller than three. With *pathlength* = 1 only network anchoring connections of type (1) are considered, with *pathlength* = 2 only network anchoring connections of types (1) and (2) are considered.

The calculation of the network anchoring probability for a given assignment possibility uses the relative probabilities of assignments of other peaks (e.g. P', $P_{\alpha\gamma}$ etc. in the above formulas), which are not known at the outset since they depend in turn on other network anchoring probabilities. Therefore, the calculation of the network anchoring probabilities has to be performed iteratively. The relative assignment probability of the kth assignment possibility of a given peak is defined as

$$P_k^{(rel)} = \frac{cP_k}{\sum_j P_j},$$

where

$$P_k = P_{\text{shifts}} \times P_{\text{structure}} \times P_{\text{network}}$$

is the absolute assignment probability, c is the value of the parameter *confidence* of the *assign* command, and the summation extends over all assignments of the peak. The *confidence* parameter c is set to 0.5 in cycle 1 and 1.0 in all other cycles (Table 3) to take into account that the assignment possibilities are less reliably known in the first cycle, where no structure is available, than in later cycles. The iterative determination of the network anchoring probabilities is initialized by setting $P_{\text{network}} = 1$ for the calculation of the relative assignment probabilities, and continued by iterations, in which the relative assignment probabilities are updated with the network anchoring probabilities of the previous iteration until all changes of assignment probabilities are below 1 %, or a maximal number of 15 iterations has been reached. From iteration 6 onwards, the updated assignment probabilities are set to the arithmetic mean of the previous and the new assignment probability to accelerate the convergence.

Upper bound elasticity

In manual NOE assignment it is not uncommon that unreliably determined peak volumes or intensities are corrected in the course of multiple iterations of manual assignment and structure calculation. As long as peak volumes are decreased, and hence upper bounds increased, this corresponds to a more conservative interpretation of the NOE data, which is unproblematic. The principal reason for overestimated peak volumes is spectral overlap that can easily be detected by visual inspection of the spectrum. The automated NOE assignment algorithm works on peak lists and has therefore not the possibility to check for the validity of peak volumes directly in the spectra. Instead, to partially replace the manual checking and correction of peak volumes, the algorithm allows for upper distance

bound "elasticity" which increases upper bounds that are consistently but weakly violated. To this end, the fraction p of conformers that violate the upper distance bound u is calculated. If $p < 0.8$, the upper distance bound is increased in 4 steps, $i = 1,\ldots, 4$, to $u_i = u\ (1 + i/4\ (f - 1))$, where the maximal increment factor f is a parameter ($u_4 = u\ f$). If the fraction p_i of conformers that violate the upper distance bound b_i is greater than 0.8, the original upper bound u is replaced by b_i. If all $p_i < 0.8$, i.e. if the violation is too large to be overcome by a slight increase of bound, the original upper bound remains unchanged. In the *noeassign* command, elasticity is allowed from cycle 3 onwards with a maximal increment factor of $f = 1.25$, i.e. individual upper bounds may be increased automatically by maximally 25 % if this allows to fulfill the restraint in more than 80 % of the conformers.

Optionally, upper bounds can also be decreased if the actual distances in the input conformers are consistently shorter than the original upper distance bound. However, this option is not used with the *noeassign* command because it can potentially lead to biased structures. If an upper distance bound is changed, its modified value is indicated in the first line of the report on the assignment of the peak. The additional option *changevol* can be used to overwrite the peak volumes in the output peak lists by the corrected value corresponding to the modified upper distance bound.

Symmetric multimers

The *assign* command provides special features for symmetric multimers, which can be defined with the *molecules define* command. In this case, only assignments having the first atom in the first monomer are made. The corresponding symmetry-related distance restraints can be added afterwards with the *molecules symmetrize* command. Homomultimer assignments can be restricted to only intramolecular or only intermolecular ones by setting the (XEASY) color codes of the corresponding peaks to 8 or 9, respectively. Furthermore, intermolecular assignments between residues i and j are considered only if $|i - j|$ is within the range specified by the *interrange* parameter. Intermolecular assignments of a peak are also excluded if the peak has at least one intramolecular assignment between residues i and j with $|i - j|$ less than or equal to the value of the parameter *prefer*.

Constraint combination

Constraint combination (Herrmann et al. 2002a) is the most important technique to reduce the impact of erroneous restraints on the structure calculation. It is a generalization of the concept of ambiguous distance restraints in that temporarily ambiguous distance restraints are formed from the assignments of two, in general randomly selected, unrelated peaks. Such combined restraints will only lead to a distortion of the structure if all their assignments are erroneous. Therefore, combining a correct and an erroneous restraint will result in a correct combined restraint, and the effect of the erroneous restraint is suppressed at the expense of a temporary loss of information. For instance, if $p = 10$ % of all restraints are erroneous, then one expects only $p^2 = 1$ % of the combined restraints to be erroneous. Constraint combination is implemented in the *distance combine* command and by default applied in the first two cycles. In the later cycles, when erroneous assignments can be filtered out readily by comparison with the structure, the restraints from each NOESY cross peak are used individually to fully exploit their information content. In automated NOESY assignment with the *noeassign* command, "4-4 constraint combination" (Herrmann et al. 2002a) is applied to groups of 4 randomly selected individual restraints, which do not contain any intraresidual or sequential assignments. The list of individual restraints is sorted by the quality factor Q of the corresponding peaks. For each group of four individual restraints, (A, B, C, D), restraint A is selected randomly from the first, B from the second, C from the third, and D from the fourth quarter of the list. Four combined restraints are formed from the assignments of restraints A and C, A and D, B and C, and B and D, respectively. In this way the total number of restraints remains the same but their ambiguity is increased and their chance to be erroneous is decreased. The upper distance bound u of a combined restraint is obtained from the distance bounds u_1 and u_2 of the individual restraints as $u = (u_1^{-6} + u_2^{-6})^{-1/6}$.

Splitting of ambiguous distance restraints into unambiguous ones

In the last cycle (normally cycle 7) the remaining ambiguous distance restraints are converted into unambiguous distance restraints by the CYANA command *distance split*, which generates an unambiguous distance restraint from each assignment of an ambiguous distance restraint that contributes more than a given minimal amount of, by default, 25 % to the peak volume. The upper distance bounds u_j of these unambiguous restraints are obtained from the original distance bound u of the ambiguous distance restraint with $j = 1,\ldots, n$ assignments by $u_j = u\ (1 + \varepsilon\ (v_j^{-1/6} - 1))$, where ε is a parameter and v_j denotes the relative contribution of the jth assignment to the peak volume, given by $v_j = < d_j^{-6}/\Sigma_k\ d_k^{-6} >$. The distances d_j are measured in the structure from the preceding cycle, the summation runs over the n assignments, and the averaging is over all conformers of the structure bundle from the preceding cycle. The parameter ε determines

by how much the upper bounds of the unambiguous restraints are extended according to their peak volume contribution and has a default value of 0.6. Unambiguous restraints are only generated for assignments with a significant relative peak volume contributions of $v_j > 0.25$.

$$T_k = \begin{cases} T_{\text{high}} & \text{in steps } k = 1, \ldots, N/5 \\ T_{\text{med}} + (T_{\text{high}} - T_{\text{med}})(1 - s_k)^4 & \text{in steps } k = N/5 + 1, \ldots, N_1 \end{cases}$$

Structure calculation

The structure calculation is carried out using the CYANA torsion angle dynamics algorithm (Güntert et al. 1997) and the standard simulated annealing schedule implemented in the *anneal* macro that is applied to a given number of start conformers with random torsion angle values. The, by default, 20 conformers with the lowest final target function values are saved as a structure bundle. The number of random start conformers is typically 100. It may be increased for difficult systems with low convergence of the structure calculation.

The standard simulated annealing protocol in the program CYANA comprises $N \geq 1000$ torsion angle dynamics time steps. It starts from a conformation with all torsion angles treated as independent, uniformly distributed random variables and consists of five phases:

1. *Initial minimization.* A short conjugate gradient minimization is applied to reduce high energy interactions that could otherwise disturb the torsion angle dynamics algorithm: 100 conjugate gradient minimization steps are performed, including only distance restraints between atoms up to 3 residues apart along the sequence, followed by a further 100 minimization steps including all restraints. For efficiency, all hydrogen atoms are excluded from the check for steric overlap, the repulsive core radii of heavy atoms without covalently bound hydrogen atoms are decreased by 0.2 Å with respect to their standard values, and the radii of heavy atoms with covalently bound hydrogens are decreased by 0.05 Å. The weighting factors in the target function are set to 1 for user-defined upper and lower distance bounds, and to 0.5 for steric lower distance bounds.
2. *First simulated annealing stage with reduced heavy atom radii.* A torsion angle dynamics trajectory with $N_1 = (N - 200)/3$ time steps is generated. The first $N/5$ of these torsion angle dynamics steps are performed at a constant high reference temperature T_{high} of, typically, 10,000 K, followed by slow cooling according to a fourth-power law to an intermediate reference temperature $T_{\text{med}} = T_{\text{high}}/20$, i.e. the reference temperature is set to

   $$T_k = T_{\text{high}}$$

   in steps $k = 1, \ldots, N/5$, and

   $$T_k = T_{\text{med}} + (T_{\text{high}} - T_{\text{med}})(1 - s_k)^4$$

   in steps $k = N/5 + 1, \ldots, N_1$, where s_k is a linear function varying from 0 at $k = N/5$ to 1 at $k = N_1$. The time step is initialized to 2 fs. The list of van der Waals lower distance bounds is updated every 50 steps using a cutoff equal to twice the largest van der Waals radius plus 1 Å (=4.2 Å for proteins) for the van der Waals pair list generation throughout all torsion angle dynamics phases.
3. *Second simulated annealing stage with normal heavy atom radii and, later, normal hydrogen atom radii.* The repulsive core radii of all heavy atoms are reset to their standard values, 50 conjugate gradient minimization steps are performed, and the torsion angle dynamics trajectory is continued for $2N_1$ time steps starting with an initial time step that is half as long as the last preceding time step. The reference temperature is decreased according to a fourth-power law from the intermediate temperature T_{med} to zero reference temperature, i.e. the reference temperature is set to

   $$T_k = T_{\text{med}}(1 - s_k)^4$$

   in steps $k = 1, \ldots, 2N_1$, where $s_k = k/2N_1$. After half of these time steps, the hydrogen atoms are included, with their standard radii, in the steric overlap check, and 50 conjugate gradient minimization steps are performed before continuing the trajectory, starting with a time step that is half as long as the last preceding time step.
4. *Low temperature phase with increased weight for steric repulsion.* The weighting factor for steric restraints is increased to 2, and 50 conjugate gradient minimization steps are performed, followed by 200 torsion angle dynamics steps at zero reference temperature, starting with a time step that is half as long as the last preceding time step.
5. *Final minimization.* A final minimization with a maximum of, typically, 1000 conjugate gradient steps is applied. Normally, the minimization will stop well before 1000 steps have been executed.

In general, the only user-modified parameter of the simulated annealing schedule is the number N of torsion

angle dynamics steps, which is typically 10,000 and may be increased up to about 50,000 for "difficult" systems, for which structures with low target function value exist but are difficult to find by the algorithm. The "success rate" of the algorithm depends on the molecular system and the restraints used. In general, the success rate decreases with increasing molecular size, higher multimeric states, increasing ambiguity of restraint assignments, sparse restraint data sets with low (but not trivially low) information content, data sets with inconsistent long-range restraints, and the presence of residual dipolar coupling or pseudocontact shift restraints. In an automated NOESY assignment calculation, the success rate is normally lowest in cycle 1 because of the higher ambiguity of NOE cross peak assignments, and then improves slightly from cycle to cycle.

Final structure calculation

The final structure calculation uses the NOE distance restraints from the last NOE assignment cycle (normally cycle 7). The distance restraints are interpreted without automatic, on-the-fly swapping of stereospecific assignments in order to make them compatible with refinement and validation programs that can only handle conventional distance restraints. Stereospecific assignments that are consistent over all 20 conformers from the last cycle are fixed (Orts et al. 2013).

Distance restraints with not stereospecifically assigned diastereotopic pairs are modified to account for the absence of the stereospecific assignment by the command *distance modify*, which applies pseudoatom corrections and eliminates meaningless and duplicate distance restraints (Güntert et al. 1991a, b). Alternatively, if the *noeassign* command is called with the option *stereoexpand*, the command *distance stereoexpand* can be used, which replaces all assignments to not stereospecifically assigned diastereotopic atoms by the corresponding pseudoatoms, and simultaneously modifies the upper distance bound value u by a value u_Q that is suitable for $1/r^6$ summation over all diastereotopic atoms: $u_Q = (u^{-6} + (u + q)^{-6})^{-1/6}$, where q is half the distance between the two diastereotopic partners, if one diastereotopic atom is involved in the assignment (e.g. a restraint between H^{α} and $H^{\beta 2}$), or $u_Q = (u^{-6} + (u + q_1)^{-6} + (u + q_2)^{-6} + (u + q_1 + q_2)^{-6})^{-1/6}$ if two diastereotopic atoms are involved in the assignment (e.g. a restraint between two $H^{\beta 2}$s), and q_1 and q_2 are the half-distances between the diastereotopic partners within the two prochiral groups. Using the *stereoexpand* option maintains a one-to-one correspondence between peaks and distance restraints, which is not always the case with the *distance modify* command.

The final structure calculation produces the following output files: (1) final.upl: Final NOE upper distance bounds. (2) final.aco: Torsion angle restraints used in the final structure calculation (only if the *autoaco* option is set). (3) final.pdb: Final structure bundle. (4) final.ovw: Overview table for the final structure calculation. (5) finalstereo.cya: Stereospecific assignments determined on the basis of the NOE distance restraints (Orts et al. 2013). (6) *A*-final.prot: Copy of the input chemical shift list *A* in which the chemical shifts of stereospecifically assigned diastereotopic atoms are swapped, if necessary. (7) rama.ps: Ramachandran plot for the final structure.

Summary table

The *cyanatable* command can be used to generate a summary table of a complete structure calculation with automated NOESY assignment by the *noeassign* command. An example is shown in Fig. 3. The summary table is particularly useful to get a quick overview of the outcome of the structure determination, e.g. to detect possible problems such as, for example, a low number of peak assignments for a given peak list (indicating a possible systematic shift of the peak positions relative to the chemical shift values in chemical shift list), high target function values in the initial cycle (indicating severe inconsistencies in the data), high RMSDs in the initial cycle (data insufficient or too contradictive to generate a well-defined structure in the first cycle), a low number of (in particular long-range) restraints (indicating lack of data or lack of convergence of the structure calculation), etc.

Results and discussion

The algorithm described in this paper has been used in the CYANA software package for the calculation of a large number NMR protein structures. A considerable fraction of all NMR structures in the PDB have been solved using the present algorithm (Guerry and Herrmann 2011; Williamson and Craven 2009), including also large proteins (Kainosho et al. 2006) and structures determined by solid-state NMR (Schütz et al. 2015).

Results of the application of this algorithm in the second round of the CASD-NMR project (Rosato et al. 2009, 2012) have been presented in two recent publications (Buchner and Güntert 2015a; Schmidt and Güntert 2013). For a series of proteins, the chemical shift assignments and both unrefined and refined NOESY peak lists were made available by the North East Structural Genomics (NESG) consortium before the corresponding manually refined 3D structures were released by the PDB. The NOE assignment and structure calculation results (Schmidt and Güntert 2013) showed that with refined NOESY peak lists, the

a

Cycle	1	2	3	4	5	6	7	
Peaks:								
selected	6359	6359	6359	6359	6359	6359	6359	
in n15.peaks	1529	1529	1529	1529	1529	1529	1529	
in c13.peaks	4680	4680	4680	4680	4680	4680	4680	
in aro.peaks	150	150	150	150	150	150	150	
assigned	6268	6287	6259	6269	6252	6225	6217	
unassigned	91	72	100	90	107	134	**142**	
without assignment possibility	32	37	37	38	38	38	38	
with violation below 0.5 A	59	0	6	6	18	42	36	
with violation between 0.5 and 3.0 A	0	22	44	33	37	41	57	
with violation above 3.0 A	0	13	13	13	14	13	11	
in n15.peaks	13	11	23	18	25	34	39	
in c13.peaks	75	59	72	70	80	98	101	
in aro.peaks	3	2	5	2	2	2	2	
with diagonal assignment	608	608	608	608	608	608	608	
Cross peaks:								
with off-diagonal assignment	5660	5679	5651	5661	5644	5617	5609	
with unique assignment	2670	3448	3872	4074	4451	4624	4668	
with short-range assignment |i-j|<=1	3969	3950	3881	3844	3802	3769	3762	
with medium-range assignment 1<|i-j|<5	864	765	749	766	749	750	744	
with long-range assignment |i-j|>=5	827	964	1021	1051	1093	1098	1103	
Upper distance limits:								
total	3608	3257	3076	3010	2886	2799	2718	2835
short-range, |i-j|<=1	1975	1772	1592	1509	1416	1350	1217	1285
medium-range, 1<|i-j|<5	1246	1017	616	618	580	571	566	584
long-range, |i-j|>=5	387	468	868	883	890	878	935	**966**
Average assignments/constraint	4.00	2.35	1.45	1.39	1.28	1.22	1.00	1.00
Average target function value	29.89	30.92	63.40	10.92	9.11	4.47	5.71	2.64
RMSD (residues 5..75):								
Average backbone RMSD to mean	**0.52**	0.51	0.24	0.23	0.28	0.21	0.23	0.17
Average heavy atom RMSD to mean	1.02	1.00	0.70	0.72	0.71	0.66	0.67	0.54

b

Cycle	1	2	3	4	5	6	7	final
Peaks:								
selected	6359	6359	6359	6359	6359	6359	6359	
in n15.peaks	24.0%	24.0%	24.0%	24.0%	24.0%	24.0%	24.0%	
in c13.peaks	73.6%	73.6%	73.6%	73.6%	73.6%	73.6%	73.6%	
in aro.peaks	2.4%	2.4%	2.4%	2.4%	2.4%	2.4%	2.4%	
assigned	98.6%	98.9%	98.4%	98.6%	98.3%	97.9%	97.8%	
unassigned	1.4%	1.1%	1.6%	1.4%	1.7%	2.1%	**2.2%**	
without assignment possibility	0.5%	0.6%	0.6%	0.6%	0.6%	0.6%	0.6%	
with violation below 0.5 A	0.9%	0.0%	0.1%	0.1%	0.3%	0.7%	0.6%	
with violation between 0.5 and 3.0 A	0.0%	0.3%	0.7%	0.5%	0.6%	0.6%	0.9%	
with violation above 3.0 A	0.0%	0.2%	0.2%	0.2%	0.2%	0.2%	0.2%	
in n15.peaks	0.9%	0.7%	1.5%	1.2%	1.6%	2.2%	2.6%	
in c13.peaks	1.6%	1.3%	1.5%	1.5%	1.7%	2.1%	2.2%	
in aro.peaks	2.0%	1.3%	3.3%	1.3%	1.3%	1.3%	1.3%	
with diagonal assignment	9.6%	9.6%	9.6%	9.6%	9.6%	9.6%	9.6%	
Cross peaks:								
with off-diagonal assignment	89.0%	89.3%	88.9%	89.0%	88.8%	88.3%	88.2%	
with unique assignment	42.0%	54.2%	60.9%	64.1%	70.0%	72.7%	73.4%	
with short-range assignment |i-j|<=1	62.4%	62.1%	61.0%	60.4%	59.8%	59.3%	59.2%	
with medium-range assignment 1<|i-j|<5	13.6%	12.0%	11.8%	12.0%	11.8%	11.8%	11.7%	
with long-range assignment |i-j|>=5	13.0%	15.2%	16.1%	16.5%	17.2%	17.3%	17.3%	
Upper distance limits:								
total	3608	3257	3076	3010	2886	2799	2718	2835
short-range, |i-j|<=1	54.7%	54.4%	51.8%	50.1%	49.1%	48.2%	44.8%	45.3%
medium-range, 1<|i-j|<5	34.5%	31.2%	20.0%	20.5%	20.1%	20.4%	20.8%	20.6%
long-range, |i-j|>=5	10.7%	14.4%	28.2%	29.3%	30.8%	31.4%	34.4%	34.1%
Average assignments/constraint	4.00	2.35	1.45	1.39	1.28	1.22	1.00	1.00
Average target function value	29.89	30.92	63.40	10.92	9.11	4.47	5.71	2.64
RMSD (residues 5..75):								
Average backbone RMSD to mean	**0.52**	0.51	0.24	0.23	0.28	0.21	0.23	0.17
Average heavy atom RMSD to mean	1.02	1.00	0.70	0.72	0.71	0.66	0.67	0.54

◀**Fig. 3** Summary table of a structure calculation with automated NOESY assignment. The summary was produced with CYANA command *cyanatable* and shows results obtained with the refined NOESY peak list for the CASD-NMR protein OR135. **a** Table with absolute values, produced with the command *cyanatable −l*. **b** Table with percentage values, produced with the command *cyanatable −lp*. The same data is shown in both tables. The tables comprise one column for each of the cycles 1–7 and for the final structure calculation. Entries that are of particular importance to assess the outcome of the structure determination are indicated in *bold*, i.e. the number of unassigned peaks, the number of long-range distance restraints, and the RMSD of the structure bundle from *cycle 1*

resulting structure coincided with the partially manually determined reference structure within 0.37–1.59 Å backbone RMSD for the structured regions. Using the raw peak lists containing 4–68 % artifact peaks, accurate structures with 0.61–1.64 Å backbone RMSD from the reference structure were obtained for six out of eight proteins. For two proteins larger RMSDs to the reference structure of 3.38 and 6.73 Å were obtained (Schmidt and Güntert 2013). For both proteins, the low reliability of the structure calculation with automated NOE assignment was obvious from the high percentage of unassigned NOESY peaks (55 and 56 % for HR5460A and StT322, respectively), and high RMSDs (5.6/5.8 Å) and target function values (6500/2500 $Å^2$) in the first cycle.

A large-scale investigation of the "convergence radius" of the present algorithm has been conducted and is presented in an accompanying paper (Buchner and Güntert 2015b). To this end, the algorithm was applied to a large number of data sets. In addition to the original experimental data sets, various modified data sets were generated that mimic typical limitations of NMR spectra and their analysis, e.g. missing peaks, additional artifact peaks, imprecise peak picking and peak volume determination, and missing or erroneous resonance assignments. The results (Buchner and Güntert 2015b) showed that the algorithm is robust with respect to imperfections of the NOESY peak lists but susceptible against more than 10 % missing or erroneous resonance assignments (Buchner and Güntert 2015b).

Assessing the reliability of NMR protein structures is a particularly important issue when automated NOESY assignment is used, because traditional criteria that have been applied with manually assigned NOESY cross peaks, such as low final CYANA target function values and small restraint violations (Güntert 1998), or low RMSD values for the final structure bundle, are no longer informative with an algorithm that automatically discards incompatible NOE distance restraints. Therefore, other criteria, such as the RMSD radius of the structure bundle in cycle 1, the RMSD drift, defined as the RMSD between the mean coordinates of the structure bundle in cycle 1 and the final structure bundle, and the number of discarded NOESY cross peaks have to be used (Buchner and Güntert 2015b; Herrmann et al. 2002a; Jee and Güntert 2003). Especially dangerous are false positives, i.e. cases, where the evaluation parameters meet the required criteria but the final structure is nevertheless misfolded. A criterion that combines the RMSD radius R of the structure bundle in cycle 1 and the RMSD drift D into a new quantity $C = ((3R/2)^2 + D^2)^{1/2}$, is particularly successful to exclude false positives: Only 0.01 % of all structure calculations with $C < 3.0$ Å resulted in final structures with an RMSD of more than 3 Å to the reference structure (Buchner and Güntert 2015b). The quantity C is highly correlated with the accuracy of the final structure as measured by the RMSD to the reference structure. Over a wide range of accuracies, the RMSD bias exceeds the corresponding C value only very rarely (Buchner and Güntert 2015b).

As an extension of the present algorithm, consensus structure bundles (Buchner and Güntert 2015a) provide an even better means to reliably estimate the accuracy of a structure obtained with automated NOESY assignment by using consensus restraints derived from multiple independent runs of the present algorithm with different random number generator seed values. Another extension of the algorithm, the REGMEAN procedure (Gottstein et al. 2012), yields a single structure representation of the final structure bundle that is as close as possible to its mean coordinates while maintaining perfect covalent geometry and on average equal steric quality and an equally good fit to the NMR data as the individual conformers of the bundle.

Conclusions

In this paper we have given a detailed presentation of the automated NOE assignment algorithm in the software package CYANA that has, despite of its widespread use, never been fully described. The algorithm is based on a consistent probabilistic model of the NOE assignment process and can replace the large majority of manual NOE assignment procedures. Large-scale evaluations of the algorithm in an accompanying paper (Buchner and Güntert 2015b), where the effect of data imperfections, i.e. incomplete or erroneous chemical shift assignments, missing NOESY cross peaks, inaccurate peak positions, inaccurate peak intensities, lower dimensionality NOESY spectra, and higher tolerances for the matching of chemical shifts and peak positions were simulated, have established the range of applicability of the algorithm as well as criteria to assess the expected accuracy of NMR protein structures determined with the algorithm. The results show that the algorithm is remarkably robust with regard to imperfections of the NOESY peak lists and the size of chemical shift

tolerances but susceptible to lacking or erroneous resonance assignments, in particular for nuclei that are involved in many NOESY cross peaks (Buchner and Güntert 2015b). Imperfections within the chemical shift assignment can cause severe problems during NOE assignment and structure calculation. Already 10 % of missing or erroneous chemical shifts can result in inaccurate structures with RMSD bias values above 3 Å. In some cases of high-quality data and many 3D NOESY peaks, higher percentages of missing or erroneous chemical shifts can be tolerated. Less severe problems arise from missing peaks, errors in peak positions and volumes as well as lower resolution simulated by using higher assignment tolerances. Furthermore, it was shown that in general data imperfections cannot be overcome by longer simulated annealing during the structure calculations. The convergence of the initial structure calculation cycle and the RMSD drift between the first and the last cycle can be combined in a weighted average and used as an indication for the reliability of a structure calculation result.

Automated NOESY assignment can be combined with automated sequence-specific resonance assignment with the Garant (Bartels et al. 1997) or FLYA (Schmidt and Güntert 2012) algorithms in order to perform a complete NMR structure determination without manual interventions (López-Méndez and Güntert 2006). In favorable cases, this can even be achieved using exclusively experimental data from NOESY spectra (Ikeya et al. 2011; Schmidt and Güntert 2013).

Acknowledgments We gratefully acknowledge financial support by the Lichtenberg program of the Volkswagen Foundation and a Grant-in-Aid for Scientific Research of the Japan Society for the Promotion of Science (JSPS).

References

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

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

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28:235–242

Braun W, Go N (1985) Calculation of protein conformations by proton–proton distance constraints—a new efficient algorithm. J Mol Biol 186:611–626

Buchner L, Güntert P (2015a) Increased reliability of NMR protein structures by consensus structure bundles. Structure 23:425–434

Buchner L, Güntert P (2015b) Systematic evaluation of combined automated NOE assignment and structure calculation with CYANA. J Biomol NMR. doi:10.1007/s10858-015-9921-z

Buchner L, Schmidt E, Güntert P (2013) Peakmatch: a simple and robust method for peak list matching. J Biomol NMR 55:267–277

Gottstein D, Kirchner DK, Güntert P (2012) Simultaneous single-structure and bundle representation of protein NMR structures in torsion angle space. J Biomol NMR 52:351–364

Gronwald W, Moussa S, Elsner R, Jung A, Ganslmeier B, Trenner J, Kremer W, Neidig KP, Kalbitzer HR (2002) Automated assignment of NOESY NMR spectra using a knowledge based method (KNOWNOE). J Biomol NMR 23:271–287

Guerry P, Herrmann T (2011) Advances in automated NMR protein structure determination. Q Rev Biophys 44:257–309

Güntert P (1998) Structure calculation of biological macromolecules from NMR data. Q Rev Biophys 31:145–237

Güntert P (2009) Automated structure determination from NMR spectra. Eur Biophys J 38:129–143

Güntert P, Braun W, Wüthrich K (1991a) 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

Güntert P, Qian YQ, Otting G, Müller M, Gehring W, Wüthrich K (1991b) Structure determination of the Antp(C39S) homeodomain from nuclear magnetic resonance data in solution using a novel strategy for the structure calculation with the programs DIANA, CALIBA, HABAS and GLOMSA. J Mol Biol 217:531–540

Güntert P, Dötsch V, Wider G, Wüthrich K (1992) Processing of multidimensional NMR data with the new software PROSA. J Biomol NMR 2:619–629

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

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

Herrmann T, Güntert P, Wüthrich K (2002a) 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

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

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

Ikeya T, Jee J-G, Shigemitsu Y, Hamatsu J, Mishima M, Ito Y, Kainosho M, Güntert P (2011) Exclusively NOESY-based automated NMR assignment and structure determination of proteins. J Biomol NMR 50:137–146

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

Johnson BA, Blevins RA (1994) NMR view—a computer program for the visualization and analysis of NMR data. J Biomol NMR 4:603–614

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

Kirchner DK, Güntert P (2011) Objective identification of residue ranges for the superposition of protein structures. BMC Bioinform 12:170

Kobayashi N, Iwahara J, Koshiba S, Tomizawa T, Tochio N, Güntert P, Kigawa T, Yokoyama S (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

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

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

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

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

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

Nilges M, Macias MJ, O'Donoghue 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

Orts J, Vögeli B, Riek R, Güntert P (2013) Stereospecific assignments in proteins using exact NOEs. J Biomol NMR 57:211–218

Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1986) Numerical recipes. The art of scientific computing. Cambridge University Press, Cambridge

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

Rosato A, Bagaria A, Baker D, Bardiaux B, Cavalli A, Doreleijers JF, Giachetti A, Guerry P, Güntert P, Herrmann T, Huang YJ, Jonker HRA, Mao B, Malliavin TE, Montelione GT, Nilges M, Raman S, van der Schot G, Vranken WF, Vuister GW, Bonvin AMJJ (2009) CASD-NMR: critical assessment of automated structure determination by NMR. Nat Methods 6:625–626

Rosato A, Aramini JM, Arrowsmith C, Bagaria A, Baker D, Cavalli A, Doreleijers JF, Eletsky A, Giachetti A, Guerry P, Gutmanas A, Güntert P, He YF, Herrmann T, Huang YPJ, Jaravine V, Jonker HRA, Kennedy MA, Lange OF, Liu GH, Malliavin TE, Mani R, Mao BC, Montelione GT, Nilges M, Rossi P, van der Schot G, Schwalbe H, Szyperski TA, Vendruscolo M, Vernon R, Vranken WF, de Vries S, Vuister GW, Wu B, Yang YH, Bonvin AMJJ (2012) Blind testing of routine, fully automated determination of protein structures from NMR data. Structure 20:227–236

Schmidt E, Güntert P (2012) A new algorithm for reliable and general NMR resonance assignment. J Am Chem Soc 134:12817–12829

Schmidt E, Güntert P (2013) Reliability of exclusively NOESY-based automated resonance assignment and structure determination of proteins. J Biomol NMR 57:193–204

Schubert M, Labudde D, Oschkinat H, Schmieder P (2002) A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on ^{13}C chemical shift statistics. J Biomol NMR 24:149–154

Schütz AK, Vagt T, Huber M, Ovchinnikova OY, Cadalbert R, Wall J, Güntert P, Böckmann A, Glockshuber R, Meier BH (2015) Atomic-resolution three-dimensional structure of amyloid beta fibrils bearing the Osaka mutation. Angew Chem Int Edit 54:331–335

Skinner SP, Goult BT, Fogh RH, Boucher W, Stevens TJ, Laue ED, Vuister GW (2015) Structure calculation, refinement and validation using CcpNmr analysis. Acta Crystallogr D 71:154–161

Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Wenger RK, Yao HY, Markley JL (2008) BioMagResBank. Nucleic Acids Res 36:D402–D408

Vögeli B, Kazemi S, Güntert P, Riek R (2012) Spatial elucidation of motion in proteins by ensemble-based structure calculation using exact NOEs. Nat Struct Mol Biol 19:1053–1057

Williamson MP, Craven CJ (2009) Automated protein structure calculation from NMR data. J Biomol NMR 43:131–143

Zhang Z, Porter J, Tripsianes K, Lange OF (2014) Robust and highly accurate automatic NOESY assignment and structure determination with Rosetta. J Biomol NMR 59:135–145