J Biomol NMR (2008) 42:23–33
DOI 10.1007/s10858-008-9259-x

ARTICLE

Automated amino acid side-chain NMR assignment of proteins using ^{13}C- and ^{15}N-resolved 3D [^{1}H,^{1}H]-NOESY

Francesco Fiorito · Torsten Herrmann · Fred F. Damberger · Kurt Wüthrich

Received: 8 May 2008 / Accepted: 15 July 2008 / Published online: 16 August 2008

Abstract ASCAN is a new algorithm for automatic sequence-specific NMR assignment of amino acid side-chains in proteins, which uses as input the primary structure of the protein, chemical shift lists of $^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and possibly $^{1}H^{\alpha}$ from the previous polypeptide backbone assignment, and one or several 3D ^{13}C- or ^{15}N-resolved [^{1}H,^{1}H]-NOESY spectra. ASCAN has also been laid out for the use of TOCSY-type data sets as supplementary input. The program assigns new resonances based on comparison of the NMR signals expected from the chemical structure with the experimentally observed NOESY peak patterns. The core parts of the algorithm are a procedure for generating expected peak positions, which is based on variable combinations of assigned and unassigned resonances that arise for the different amino acid types during the assignment procedure, and a corresponding set of acceptance criteria for assignments based on the NMR experiments used. Expected patterns of NOESY cross peaks involving unassigned resonances are generated using the list of previously assigned resonances, and tentative chemical shift values for the unassigned signals taken from the BMRB statistics for globular proteins. Use of this approach with the 101-amino acid residue protein FimD(25–125) resulted in 84% of the hydrogen atoms and their covalently bound heavy atoms being assigned with a correctness rate of 90%. Use of these side-chain assignments as input for automated NOE assignment and structure calculation with the ATNOS/CANDID/DYANA program suite yielded structure bundles of comparable quality, in terms of precision and accuracy of the atomic coordinates, as those of a reference structure determined with interactive assignment procedures. A rationale for the high quality of the ASCAN-based structure determination results from an analysis of the distribution of the assigned side chains, which revealed near-complete assignments in the core of the protein, with most of the incompletely assigned residues located at or near the protein surface.

Keywords Amino acid side-chain NMR assignment · ASCAN · Nuclear Overhauser effect · NOESY · Automation of protein structure determination

F. Fiorito · T. Herrmann · F. F. Damberger · K. Wüthrich
Institut für Molekularbiologie und Biophysik, ETH Zürich, CH-8093 Zurich, Switzerland

Present Address:
T. Herrmann
Université de Lyon, CNRS/ENS Lyon/UCB Lyon-1, Centre Européen de RMN à Très Hauts Champs de Lyon, 5 Rue de la Doua, 69100 Villeurbanne, France

K. Wüthrich (✉)
Department of Molecular Biology and Skaggs Institute of Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA
e-mail: kw@mol.biol.ethz.ch

Introduction

This paper describes a new computational protocol, ASCAN (automated side-chain resonance assignment), which operates on the 3D heteronuclear-resolved [^{1}H,^{1}H]-NOESY data sets that are subsequently used to collect the input of NOE-distance constraints for the structure calculation. Thereby ASCAN makes use of the chemical shift lists for the previously assigned backbone atoms, and the knowledge on the covalent polypeptide structure. Considering the dependence of the magnetization transfers in NOESY experiments on the inverse sixth power of the

^{1}H–^{1}H distances, the program performs a search for an optimal match between expected and experimentally observed NOE peaks. To this end, ASCAN projects expected peak patterns onto the experimental data and uses a novel scoring scheme for identifying new side-chain resonance assignments based on the closeness of fit between expected and experimental peak patterns. To make inevitable imperfections of experimental NMR data traceable, the chemical shifts of the previously assigned backbone and C^{β} atoms are used to guide both the peak-picking of the NOESY spectra and the search of new side-chain resonance assignments.

In biomolecular nuclear magnetic resonance (NMR) spectroscopy, automated resonance assignment promises to provide efficient, objective and reliable handling of the large data sets comprising typically thousands of resonance frequencies (Altieri and Byrd 2004; Baran et al. 2004; Güntert 2003; Moseley and Montelione 1999; Moseley et al. 2004). In practice, however, spectral artifacts such as noise bands, absence of signals due to local fast relaxation and accidental resonance overlap impose substantial obstacles for automated resonance assignment routines. Consequently, nearly all structure determinations so far used interactive algorithms that operate on listings of peak positions and volumes ("peak lists") rather than on the raw NMR spectra (Atreya et al. 2000; Buchler et al. 1997; Coggins and Zhou 2003; Gronwald and Kalbitzer 2004; Gronwald et al. 1998; Hyberts and Wagner 2003; Slupsky et al. 2003; Zimmerman et al. 1997). Interactive identification of NMR signals in the spectra of proteins ("peak picking") is a laborious task that depends often on largely arbitrary decisions by experienced spectroscopists, which would need to be reenacted by software capable of generating the peak lists automatically. Typically, robust automated spectral analysis has therefore been limited to spectral regions with scarcity of both signal overlap and artifacts, and manual re-inspection of the results is generally advised (Garrett et al. 1991; Koradi et al. 1998; Kraulis 1989; Orekhov et al. 2001). More recently, algorithms for automated peak-picking have been shown to benefit considerably in their performance if the spectral data can be supplemented with additional information, such as chemical shift lists of atoms that are correlated with as yet unidentified signals in the spectra to be analyzed (Herrmann et al. 2002b), or information on expected peak patterns derived from the magnetization pathways in the NMR experiments used (Moseley et al. 2004). These "constrained" peak picking algorithms mimic the *modus operandi* of an experienced spectroscopist, who analyses new signals in the context of previously assigned resonances. Such an approach, based on nearly complete sequence-specific resonance assignments, has been applied extensively to NOE assignments in 2D and heteronuclear-resolved 3D [^{1}H,^{1}H]-NOESY spectra (Herrmann et al. 2002a, b). Here, we adapt the same principles to assign side-chain resonances based on previous extensive backbone assignments. In its general philosophy, ASCAN differs from most previous procedures for automated resonance assignment of backbone and/or side-chain atoms in that it operates on the raw NMR data rather than on interactively generated peak lists, which are often extensively pre-processed to lead to satisfactory results (Atreya et al. 2000; Baran et al. 2004; Bartels et al. 1997; Coggins and Zhou 2003; Eghbalnia et al. 2005; Hyberts and Wagner 2003; Malmodin et al. 2003; Moseley and Montelione 1999; Moseley et al. 2001, 2004; Slupsky et al. 2003; Zimmerman et al. 1997).

The ASCAN strategy

The program ASCAN has been designed for the automation of sequence-specific amino acid side-chain assignments in proteins for which extensive previous assignments of the polypeptide backbone atoms $^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$ and possibly $^{1}H^{\alpha}$, and for $^{13}C^{\beta}$ are available (Fig. 1, [1]). Starting from the backbone chemical shift lists, ASCAN generates expected NOE-signal patterns (Fig. 1, [2])[1] that are subsequently compared with signal patterns observed in the experimental NOESY spectra. Thereby, in each cycle of the iterative ASCAN protocol, the information on expected signal patterns is updated based on the new assignments obtained in the preceding cycle. Experimental peak patterns to be compared with these predicted peak patterns are generated by a routine for automated peak picking of the three-dimensional (3D) heteronuclear-resolved [^{1}H,^{1}H]-NOESY spectra, which makes use of the chemical shifts of all previously assigned atoms (Fig. 1, [3]). This is combined with another routine, which identifies chemical shifts of those so far unassigned atoms that contribute to three-dimensional NOE correlation peaks for which two chemical shifts had previously been assigned (Fig. 1, [4]). This information is again updated in each iteration cycle, making use of the new assignments obtained in the previous cycle. The next step consists of an evaluation of the closeness of fit between expected peak patterns and the experimental data (Fig. 1, [5]), which in turn generates the input for a set of acceptance criteria for new resonance assignments (Fig. 1, [6]).

[1] ASCAN is also laid out to operate with TOCSY data sets, which provide similar information to that exploited in the NOESY spectra and could be used as supplementary input. Whereas the main text describes exclusively the use of ASCAN with NOESY, its use with TOCSY-type spectra is further discussed in Appendix 2, which also provides a description of the expected TOCSY peak patterns.

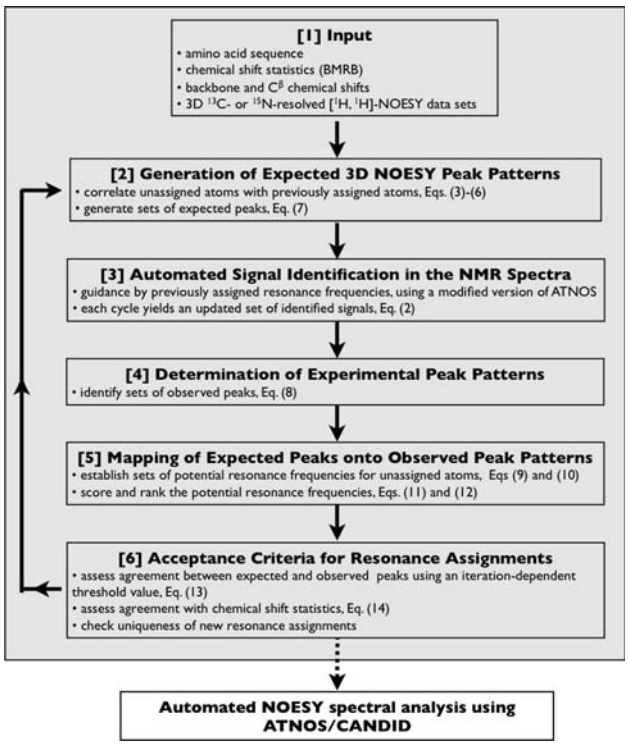

Fig. 1 Flow chart of automated ASCAN side-chain resonance assignment

[1] Input and basic notations

The input consisting of the amino acid sequence, the experimental chemical shifts of $^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and possibly $^{1}H^{\alpha}$ from the previous backbone resonance assignment, the chemical shift statistics for globular proteins from the BMRB, and one or several 3D heteronuclear-resolved [^{1}H,^{1}H]-NOESY data sets (Fig. 1, [1]) is used to generate a platform of data and of the correlations between them.

P is the set of NMR-active atoms, which is derived from the primary structure of the protein. A is the set of K assigned atoms, which is updated in each ASCAN cycle, with $A \subseteq P$.The corresponding chemical shifts are given by Eq. 1,

$$\Omega \equiv \{\Omega_{a_i} | a_i \in A; i = 1 \ldots K\}. \quad (1)$$

S is the set of L identified cross peaks in the 3D heteronuclear-resolved [^{1}H,^{1}H]-NOESY spectra,

$$S \equiv \left\{ \overrightarrow{\omega_\gamma} \equiv (\omega_{1,\gamma}, \omega_{2,\gamma}, \omega_{3,\gamma}) | \gamma = 1, \ldots, L \right\}, \quad (2)$$

which is updated in each ASCAN iteration cycle.

The magnetization transfer pathways in the NMR experiments used define the correlations between an unassigned hydrogen atom, $u \in P\backslash A$,with a set of assigned atom pairs,

$$D(u) \equiv \{(h_i, p_i) | h_i, p_i \in A\}. \quad (3)$$

In Eq. 3, $D(u)$ is the set of pairs of assigned atoms which are coupled with u, where each element of $D(u)$ is composed of an assigned heavy atom, h_i, and an assigned hydrogen atom, p_i. If in a given NOESY experiment an unassigned hydrogen atom, $u \in P\backslash A$, is expected to show a correlation to an assigned hydrogen atom, $h(u) \in A$, and the heavy atom covalently bound to the hydrogen atomp_i is also assigned, with $h(p_i) \in A$, then

$$(h(p_i), p_i) \in D(u). \quad (4)$$

In the case that the heavy atom covalently bound to u is assigned, with $h(u) \in A$, then

$$(h(u), p_i) \in D(u), \quad (5)$$

and in this situation it is not required that $h(p_i) \in A$.

The correlations of an unassigned atom, $u \in P\backslash A$, to the set of assigned atom pairs, $D(u)$, via dipolar ("through-space") couplings giving rise to ^{1}H–^{1}H nuclear Overhauser effects (NOE), can be evaluated from the fact that the fixed bond lengths, bond angles and chiralities of the covalent protein structure impose upper limits on the intraresidual and sequential ^{1}H–^{1}H distances (Güntert et al. 1991; Wüthrich 1986; Wüthrich et al. 1983). In ASCAN, these conformation-independent upper distance limits, $d_{up_i}^{cov}$, are computed analytically for all pairs of hydrogen atoms that are separated by one or two torsion angles. All the correlated pairs of hydrogen atoms then satisfy the Eq. 6,

$$d_{up_i}^{cov} \leq d_{max}. \quad (6)$$

The user-defined parameter d_{max} (Table 1) is set to a sufficiently small value to ensure that the corresponding pairs of hydrogen atoms lead to observable signals in the NOESY spectra used. All the correlated hydrogen atoms, p_i, identified using Eq. 6 are used to build the elements of $D(u)$ (Eq. 3) according to Eqs. 4 and/or 5.

[2] Generation of expected peak patterns

The expected pattern of M correlation peaks for a given unassigned hydrogen atom, $P_i(u)$, in the input spectra is determined based on Eq. 6:

$$E(u) \equiv \left\{ \overrightarrow{e(u)_i} \equiv \left(\Omega_u, \Omega_{h_i}, \Omega_{p_i}\right) | (h_i, p_i) \in D(u); i = 1 \ldots M \right\}. \quad (7)$$

In Eq. 7, $\overrightarrow{e(u)_i}$ is one of the M signals in the expected peak pattern. It is seen that two out of the three frequency coordinates of an expected peak are restrained by the resonance frequencies of previously assigned atoms, with $\Omega_{h_i}, \Omega_{p_i} \in \Omega$.Technical details of the generation of $E(u)$ are provided in Appendix 1.

[3] Automated signal identification in the NMR spectra

The identification of all signals of a N-dimensional spectrum is performed using the ATNOS algorithm (Herrmann et al.

Table 1 Parameters used in the standard protocol for automated side-chain resonance assignment with ASCAN

Symbol	Equation or parameter	Value
d_{max}	Eq. 6	5.5 Å
$\Delta\omega^p$ resp. $\Delta\vec{\omega}$	Eqs. 8, 9, 11, 12	0.025 ppm
$\Delta\omega^h$ resp. $\Delta\vec{\omega}$	Eqs. 8, 9, 11, 12	0.4 ppm
SN_{min}	Minimal signal-to-noise threshold used for signal identification	4
Iteration-dependent parameters		
C^m (Eq. 13)[a]	SN	T
First assignment phase[b]		
1.0–0.45	1.0	0.5
1.0–0.3	0.5	1.0
1.0–0.3	0.5	1.25
Second assignment phase		
1.0–0.45	1.0	0.5
1.0–0.45	0.5	0.5
1.0–0.45	1.0	1.0
1.0–0.3	1.0	0.5
1.0–0.3	0.5	1.0
1.0–0.3	0.5	1.25

For the definitions of symbols, see the text

[a] The parameter C^m is reduced in steps of 0.05 from 1.0 to the finishing values indicated, provided that new assignments were obtained in the preceding ASCAN cycle (see text)

[b] During the first assignment phase, only $^1H^\alpha$ and $^1H^\beta$ atoms are assigned, for which the resonance frequencies of the covalently bound heavy atom are known from the previous assignment of the polypeptide backbone (see text)

2002b). A key feature of ATNOS is the incorporation of chemical shift information into the process of signal identification ("peak picking"). To this end, ATNOS overlays a N-dimensional grid spanned by the resonance frequencies of the assigned atoms onto the N-dimensional heteronuclear-resolved [^{1}H,^{1}H]-NOESY spectrum. It then performs a search for local extrema only in close proximity to the grid points defined as intersections of N chemical shifts. In order to obtain information on unassigned resonance frequencies, the N-dimensional chemical shift grid used for the "standard" ATNOS spectral analysis (Herrmann et al. 2002b) is substituted by a (N-1)-dimensional chemical shift grid defined by the N-1 previously assigned atoms. In the case of a 3D heteronuclear-resolved [^{1}H,^{1}H]-NOESY data set, the 2D grid is thus spanned by the resonance frequencies of a previously assigned proton and a previously assigned heavy atom, and one then searches for the second proton frequency that provides the third coordinate of the peak in the 3D frequency space (see the definition of expected peak pattern). When compared with random peak picking, where all local extrema present in a spectrum would be taken as potential signals, this approach efficiently reduces the number of artifacts among the picked peaks.

[4] Determination of experimental peak patterns

Experimental N-dimensional cross peaks involving connectivities with unassigned atoms, u, as identified by the modified ATNOS algorithm (see [3] above), are defined as

$$O(u) \equiv \left\{ \overrightarrow{\omega(u)}_k \equiv \left(\omega_{1,k}, \omega_{2,k}, \omega_{3,k}\right) \middle| \left|\omega_{2,k} - \Omega_{h_k}\right| < \Delta\omega^h \wedge \left|\omega_{3,k} - \Omega_{p_k}\right| < \Delta\omega^p; k = 1 \ldots H \right\} \subset S, \tag{8}$$

where $\omega_{1,k}$ is the frequency of the unassigned atom, $(h_k, p_k) \in D(u)$,$\Delta\omega^h$ and $\Delta\omega^p$ are tolerance windows used for the chemical shifts of heavy atoms and hydrogen atoms, respectively (Table 1), and H is the number of cross peaks observed for the atom u. By definition (see above), two out of the three frequency coordinates of an experimental peak correlated with u are restrained by the resonance frequencies of previously assigned atoms, with $\Omega_{h_k}, \Omega_{p_k} \in \Omega$.

[5] Mapping of expected peaks onto experimental peak patterns

The function for the mapping of expected onto experimentally observed cross peaks, Q, which relates the sets of expected peaks, $E(u)$, (Eq. 7), to the experimentally observed peaks, $O(u) \subset S$ (Eq. 8), is defined by Eq. 9,

$$Q : (E(u), O(u)) \xrightarrow{\Delta\vec{\omega}} R(u), \tag{9}$$

with the tolerance window $\Delta\vec{\omega} = (\Delta\omega^p, \Delta\omega^h, \Delta\omega^p)$ (Table 1). Q yields a set of potential resonance frequencies for an unassigned ^{1}H atom, u,

$$R(u) \equiv \left\{ \Omega_{u_k} \equiv \omega_{1,k} | k = 1, \ldots, N \right\}, \tag{10}$$

The scoring function,

$$F : R(u) \xrightarrow{\Delta\vec{\omega}} [0, 1], \tag{11}$$

yields a scoring value for each potential resonance frequency, $\Omega_{u_k} \in R(u)$,which is calculated by Eq. 12,

$$F(\Omega_{u_k}) = \frac{1}{|M|} \sum_{i=1}^{M} C(\overrightarrow{e(u)}_i, \Delta\vec{\omega}), \tag{12}$$

where the sum runs over all the elements of the set of M expected peaks, $\overrightarrow{e(u)}_i \in E(u)$ (Eq. 7), and $C\left(\overrightarrow{e(u)}_i, \Delta\vec{\omega}\right)$ is a Gaussian weighting factor that has the value 1.0 for perfect agreement between expected and observed peak positions (Herrmann et al. 2002a, b) and of 0.0 if an

expected peak is not observed. For a given potential resonance assignment of the atom u, the scoring function relating the sets of expected and experimentally observed cross peak patterns (Eq. 12) thus covers the value range 0.0–1.0, and has a value of 1.0 for a perfect fit.

[6] Acceptance criteria for resonance assignments

The best matches found between expected and experimentally observed peaks yield potential new resonance assignments, whereby the frequency coordinates of the experimental cross peaks define the resonance frequencies of the newly assigned atoms. Due to inevitable imperfection of experimental NMR data recorded with macromolecules, one will typically predict some peaks for which no corresponding observed peak can be found. This is accounted for by an iteration-dependent threshold value, C^m, for the required minimal agreement between the two sets $E(u)$ (Eq. 7) and $O(u)$ (Eq. 8) in each iteration, m, of ASCAN. A previously unassigned atom, u, is thus newly assigned to a potential resonance frequency, $\Omega_{u_k} \in R(u)$,if the following three conditions are met:

(i) The scoring function value associated with a potential resonance assignment, $F(\Omega_{u_k})$ (Eq. 12), must satisfy Eq. 13,

$$F(\Omega_{u_k}) \geq C^m, \tag{13}$$

where C^m is an iteration-dependent threshold value (Table 1), and m denotes the current iteration cycle.

(ii) The potential resonance frequency, $\Omega_{u_k} \in R(u)$,must be within two standard deviations, $\sigma(u)$, of the predicted statistical mean value for the unassigned atom, $\omega(u)$, so that

$$|\omega(u) - \Omega_{u_k}| \leq 2\sigma(u), \tag{14}$$

where $\omega(u)$ and $\sigma(u)$ are taken from a statistical analysis of the chemical shifts in proteins (Seavey et al. 1991).

(iii) Only a single resonance frequency, $\Omega_{u_k} \in R(u)$, satisfies the conditions of Eqs. 13 and 14.

A standard protocol for ASCAN side-chain resonance assignments with globular proteins

Based on the experience gathered when working with a variety of different proteins, we propose an ASCAN protocol for automated side-chain resonance assignment that is composed of two assignment phases, each of which includes a succession of iterations that differ in the parameters used. The main iteration-dependent threshold, C^m (Eq. 13), is reduced from a starting value of 1.0 to the final values in steps of 0.05 (Table 1). Thereby the C^m value is decreased only if no new side-chain assignments were made in the preceding cycle. This parameter control ensures that new resonance assignments are always based on best match between observed and experimental peak pattern, by exploiting the steadily growing set of known resonance assignments.

During a first assignment phase, only ^{1}H atoms are assigned for which the resonance frequency of the covalently bound heavy atom was present in the input. In practice this means that during this first phase the assignment of the $^1H^{\alpha}$ and $^1H^{\beta}$ atoms is essentially completed, and that in a second phase the more peripheral amino acid side-chain atoms are being assigned. This second phase deals with the situation that the resonance frequencies for a given ^{1}H atom and its covalently bound heavy atom are both unknown. In this case, a two-step procedure is employed, whereby initially a set of potential resonance frequencies for the unassigned proton is obtained, and subsequently a set of potential resonance assignments for the covalently bound heavy atom is identified (a detailed description of this two-step routine is provided in Appendix 1). The entire assignment procedure makes use of fixed as well as iteration-dependent parameters (Table 1). In addition to C^m (Eq. 13), we introduced the two iteration-dependent parameters T and SN. T is a scaling factor for the user-defined tolerance window, $\Delta\vec{\omega}$ (Eqs. 8, 9, 11, 12), which is used to evaluate the closeness of the match between expected and experimentally observed peaks. Its purpose is to downscale the tolerance window to smaller values in early iterations, so that only atoms are assigned for which the agreement between experimental peak positions and chemical shift values of previously assigned atoms is extremely good. In later iterations the tolerance window is scaled to larger values, so as to allow the tentative assignment of atoms with less close agreement between experiment and the chemical shift values of the previously assigned atoms. The second newly introduced parameter, SN, provides for an iteration-dependent signal-to-noise threshold for the selection of potential cross peaks. Its purpose is to scale down the minimally required signal-to-noise threshold in the final iterations, so that signals with reduced intensity due to line broadening, solvent suppression, or inefficient magnetization transfer can be used for resonance assignment.

The development of the present ASCAN protocol was started with the protein FimD(25–125) (Nishiyama et al. 2005), using as input the nearly complete list of backbone assignments (BMRB accession code 6779), a 3D ^{15}N-resolved [^{1}H,^{1}H]-NOESY spectrum, and two 3D ^{13}C-resolved [^{1}H,^{1}H]-NOESY spectra with the carrier frequency in the aliphatic or aromatic region, respectively. The computation time for the resulting input data set (see

Fig. 1, [1]) was about 30 min on a single Intel Core™ Duo processor. The chemical shift lists obtained contain resonance frequencies for 83% of the ^{13}C- and ^{15}N-bound ^{1}H atoms, whereas the extent of the previous interactive resonance assignment was 91%. The agreement between ASCAN-based and interactive resonance assignments is better than 90%.

The sequence of FimD(25–125) contains five prolines, including the sequentially adjacent prolines 29 and 30. Since no $^{1}H^{N}$ shifts of prolines are contained in the sequence-specific backbone assignments in the ASCAN input (Fig. 1, [1]), the determination of resonance assignments for Pro side-chains atoms relies exclusively on the sequential NOE-connectivities between proline H^{α} and hydrogen atoms of the sequentially following residue (Wüthrich 1986). In FimD(25–125), the Pro residue 29 could therefore not be assigned. Since ASCAN requires that one hydrogen atom per residue is included in the input resonance assignments (Fig. 1, [1]), the assignment of prolines will always be limited.

The Fig. 2 affords a survey of the resonance assignments obtained for the different atom types and residue types in FimD(25–125). It shows that ASCAN assignments are limited for long side-chains, such as arginine and lysine (Fig. 2b), and that hydrophobic side-chains are overall assigned more completely than hydrophilic side-chains (Fig. 2b). This is also reflected in the survey of the assignment completeness for different atom types (Fig. 2a), which also shows that $^{1}H^{\alpha}$-atoms are more completely assigned than H^{β}-atoms, which in turn are more completely assigned than H^{γ}, and so on (Fig. 2a). Methyl groups and aromatic rings appear to be an exception, since they have been equally completely assigned as the H^{β}-atoms (Fig. 2a).

A projection onto the 3D structure of FimD(25–125) visualizes the distribution of the variable extent of side-chain assignments in different regions of the protein structure (Fig. 3). Overall, ASCAN performed better on interior, buried residues than on extensively solvent-exposed residues (Table 2). If one considers only the residues exhibiting an average solvent-accessibility below 20%, then the resonance assignment completeness is close to 90% (Table 2). Furthermore, comparison with the interactive side-chain resonance assignments (Nishiyama et al. 2005) indicates that ASCAN leads to similar limitations as the interactive work of an experienced spectroscopist. With both approaches, the completeness of the side chain assignments correlates with the degree of solvent-accessibility. This is readily rationalized if one considers that a much larger number of NOEs is generally observed for interior atoms than for atoms at or near the protein surface. ASCAN thus primarily assigns side-chain atoms that are involved in numerous inter-residue NOEs.

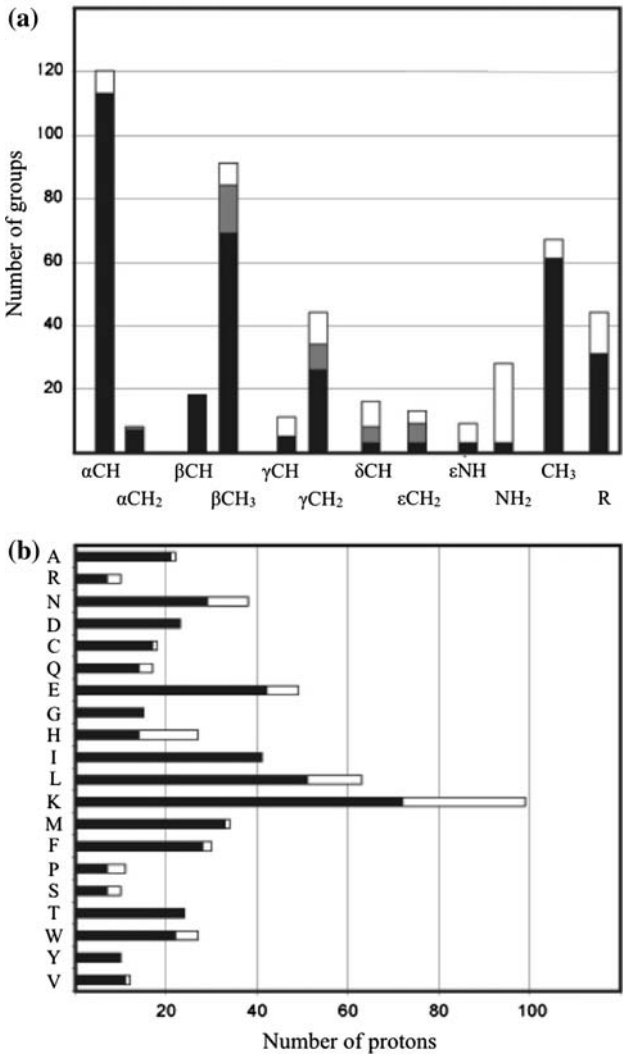

Fig. 2 Extent of the automated ASCAN side-chain resonance assignments for FimD(25–125). (**a**) Assignment extent for the different types of ^{1}H atoms, where R stands for aromatic ring protons. (**b**) Assignment extent for the different amino acid residue types. The bars represent the total number of hydrogen atoms/residues of the type indicated. Black shading represents the extent of the assignments of individual hydrogen atoms. Grey shading indicates the number of CH_2 groups for which only one ^{1}H chemical shift was assigned

The performance of ASCAN was further evaluated using the two proteins TM1442 (110 residues; Etezady-Esfarjabi et al. 2006) and BmPBP(1–128) (128 residues; Michel et al. 2005), for which nearly complete sequence-specific resonance assignments are available from conventional, interactive methods (BioMagResBank accession codes: TM1442, 5921; BmPBP(1–128), 6313), and high-quality NMR structures are also available (Protein Data Bank accession codes: TM1442, 1SBO; BmPBP(1–128), 1XFR). For these two proteins, closely similar data were obtained as those described above in detail for FimD(25–125). For BmPBP(1–128), ASCAN resulted in 85% of the

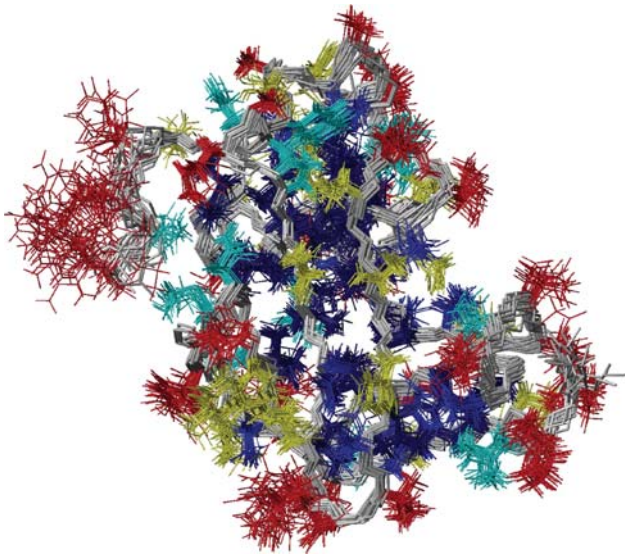

Fig. 3 Extent of automated ASCAN side-chain assignments in FimD(25–125) for residues with different average solvent accessibilities. The data of Table 2 are visualized by color-coding of the bundle of 20 conformers. The backbone is colored in gray, and the amino acid side chains are color-coded according to the solvent accessibility of the residue calculated as the average of the solvent accessibility of all hydrogen atoms of each residue, using the software MOLMOL (Koradi et al. 1996) with a solvent radius of 2.0 Å. Residues in dark blue, light blue, cyan, yellow and red have solvent accessibilities of <5%, 5–10%, 10–20%, 20–30%, and >30%, respectively

Table 2 Extent of the automated ASCAN side-chain assignments in the protein FimD(25–125) for residues with different average solvent accessibilities

Solvent accessibility [%][a]	FimD(25–125) [%][b]
0–5	86 (23)
5–10	93 (12)
10–20	83 (15)
20–30	74 (21)
30–100	69 (30)

[a] The average solvent accessibility for a given residue was calculated as the average of the solvent accessibilities of its hydrogen atoms, using the software MOLMOL (Koradi et al. 1996) with a solvent radius of 2.0 Å

[b] The extent of the assignment is given in % of all the atoms in the residues with the indicated solvent accessibilities. In parentheses, the numbers of residues are given that exhibit the indicated solvent accessibilities

expected assignments for non-labile hydrogen atoms and backbone amide protons, and the corresponding extent of the assignments for TM1442 was 81%. The extent of the interactive resonance assignments for BmPBP(1–128) and TM1442 is 95 and 89%, respectively, and the agreement between the ASCAN result and the interactive assignment is 94 and 93%, respectively.

Similar results to those for FimD(25–125) and the two additional test proteins have been obtained with high-quality NOESY data sets of several other proteins. These proteins were selected as targets in an on-going structural genomics project. They represent different molecular sizes in the range from 89 to 129 amino acid residues, as well as different secondary structure types. These results will be described in detail in a separate publication (B. Pedrini, pers. comm.).

Validation of the ASCAN performance by complete structure determination

Here, the automated ASCAN resonance assignments were used as input for an automated 3D structure determination with the software ATNOS/CANDID (Herrmann et al. 2002a, b) applied in combination with the torsion angle dynamics algorithm DYANA (Güntert et al. 1997). The NOE cross peaks that were unambiguously assigned in the seventh cycle of the ATNOS/CANDID/DYANA calculation (see Herrmann et al. 2002a, b) yielded 2,589 meaningful NOE upper distance limits as input for the final structure calculation with the program DYANA. To qualify the course of the structure calculation, values for the mean global root-mean-square deviation (RMSD) relative to the mean coordinates were calculated for the residues 27-123 in the usual bundles of 20 conformers. The bundles of conformers obtained after cycle 1 and after cycle 7, as well as the reference ATNOS/CANDID/DYANA structure bundle determined using interactively obtained side-chain resonance assignments (Nishiyama et al. 2005), are shown in Fig. 4. When using the ASCAN-based input, the correct fold was obtained already in the ATNOS/CANDID cycle 1, as documented by the RMSD and $RMSD^{drift}$ values (Herrmann et al. 2002a) of 1.12 Å and 2.12 Å calculated for the backbone atoms. The RMSD values calculated for the backbone atoms of residues 27-123 between the mean coordinates of the final bundles of 20 conformers from the ASCAN/ATNOS/CANDID calculations and the reference structure is 1.03 Å, showing that the correct structure was obtained with the ASCAN-generated assignments.

Comparable results were obtained for the two aforementioned proteins TM1442 and BmPBP(1–128). For these two proteins, the correct fold was obtained after ATNOS/CANDID cycle 1. The backbone RMSD values between the mean coordinates of the final structure bundles from the ASCAN-based structure calculations and the reference structures were 1.32 Å and 1.41 Å for TM1442 and BmPBP(1–128), respectively. Similar results were obtained when the ASCAN-based NMR structures of several other proteins selected as targets in ongoing structural genomics projects were compared with the corresponding crystal structures (manuscript in preparation; B. Pedrini, pers. comm.). Overall, the experience gained so far indicates that the

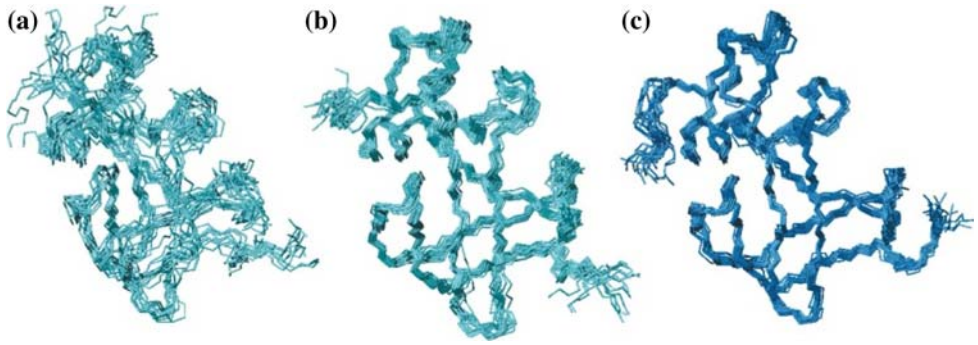

Fig. 4 Comparison of ASCAN-based structures of FimD(25–125) obtained in the first and the seventh ATNOS/CANDID cycle with a structure derived from interactive assignments. (**a**) Result of ATNOS/CANDID cycle 1 based on automated ASCAN side-chain assignments (10 best conformers). (**b**) Result of ATNOS/CANDID cycle 7 based on automated ASCAN side-chain resonance assignments (20 best conformers after energy-refinement). (**c**) Reference ATNOS/CANDID structure determination based on interactive side-chain assignments (cycle 7, 20 best conformers after energy-refinement). This structure was re-calculated using the input from Nishiyama et al. (2005)

aforementioned correlation between assignment completeness and solvent exposure is key to rationalizing the high quality of the structure determinations with ASCAN/ATNOS/CANDID. Although the completeness of the ASCAN assignments so far is lower than the extent of the assignments typically achieved with exhaustive interactive approaches, the resulting NMR structures are only marginally affected in terms of precision and accuracy of the atomic coordinates of the final structures, since ASCAN yields nearly complete side-chain assignments for the atoms involved in meaningful NOE upper distance constraints (Fig. 3).

If the ATNOS/CANDID/DYANA structure calculation should fail to satisfy the acceptance criteria for the input, which is an independent test for the result obtained with ASCAN, then manual inspection of the ASCAN result would be required. The lists of expected peak positions derived by ASCAN from the knowledge of the covalent polypeptide structure and the magnetization pathways used, could then be used as a guide also for such interactive controls.

Conclusions

The robustness of NMR structure determinations of proteins with regard to up to about 20% of the side chain resonances being left unassigned by ASCAN, is highly encouraging. This result with the protein FimD(25–125) (Table 2, Figs. 2 and 4) has been confirmed by applications with several other proteins (B. Pedrini, unpublished data). It can be rationalized by the fact that most of the ASCAN-unassigned resonances are either located near the protein surface (Fig. 3) or otherwise involved in only a small number of ^{1}H–^{1}H NOEs. Furthermore, starting from ASCAN assignments at a similar level to that obtained for FimD (25–125) (Table 2), quite extensive additional resonance assignment has in practice been achieved with only a few hours of interactive work by an experienced NMR spectroscopist (B. Pedrini, unpublished data).

It is a special advantage of the [^{1}H,^{1}H]-NOESY-based ASCAN approach that the same data sets are used for the amino acid side chain assignments and for the collection of NOE upper distance constraints. Since adjustments of polypeptide backbone chemical shifts have already been made when preparing the input for ASCAN, this eliminates the need for further chemical shift adjustments between data sets recorded with different experimental conditions, which is an intrinsically laborious procedure that may introduce unnecessary ambiguity into the NOE assignments.

ASCAN is also laid out to operate on 3D heteronuclear-resolved [^{1}H, ^{1}H]-TOCSY data sets. TOCSY data sets can be a useful supplement to NOESY. However, they should only be used in conjunction with NOESY data, since NOESY spectra have a higher data redundancy due to the presence of sequential NOE connectivities. In our experience, the ASCAN approach based on NOESY data alone is fairly robust for small and medium-sized proteins. Further studies are envisaged to evaluate the additional benefits from incorporating different types of TOCSY and/or COSY experiments into the ASCAN assignment process. This may include attempts to provide ASCAN with interactively determined incomplete sets of side-chain chemical shifts, and then to operate ASCAN for completing the partial interactive side-chain assignments.

For academic users the program ASCAN is available free of charge as part of the stand-alone ATNOS/CANDID software package (Herrmann et al. 2002a, b). Download information is available under http://www.mol.biol.ethz.ch/group/wuthrich_group/software.

Acknowledgments We thank Dr. B. Pedrini for sharing his experience with applications of ASCAN for side-chain resonance assignments in a variety of proteins selected as targets in a structural genomics project.

Financial support by the Schweizerischer Nationalfonds (project 3100-AO-113838) is gratefully acknowledged. Kurt Wüthrich is the Cecil H. and Ida M. Green Professor of Structural Biology at the Scripps Research Institute, and a member of the Skaggs Institute of Chemical Biology.

Appendix 1: Generation of expected peak pattern

The evaluation of sets of expected peak patterns (Eq. 7) and of the values of the scoring function (Eq. 12) are the core elements of the ASCAN approach. In each ASCAN iteration, these two quantities are derived from the set of already known resonance frequencies (Eq. 1). In this treatment so far, the three frequency dimensions in the Eqs. 7 and 8 were selected with the sole criteria that two out of the three frequencies of an expected peak must be restrained by previous resonance assignments, and they did not necessarily have to correspond to the frequency dimensions of the NMR experiment used. In the following, however, the notation is more specific, with the three frequency coordinates $\Delta\omega_1$, $\Delta\omega_2$ and $\Delta\omega_3$ of a cross peak in a 3D spectrum corresponding, in this order, to the indirect proton frequency, the heavy atom frequency, and the direct proton frequency. For the actual computation of $E(u)$ and $F(\Omega_{u_k})$with Eqs. 7 and 12, respectively, we distinguish three situations, depending on the extent of the previous assignments and the chemical structure of the amino acid side chain to be assigned.

Firstly, if the resonance frequency, $\Omega_{h(u)}$, of the ^{13}C or ^{15}N atom, $h(u) \in A$, that is covalently bound to an unassigned ^{1}H atom, u, with $u \in P\backslash A$, is known, then the set of assigned atom pairs, $D(u)$, used to determine the expected peak pattern for an unassigned ^{1}H atom, u, is derived from all so far assigned ^{1}H atoms, $p_i \in A$, that satisfy Eq. 6. Since $h(u) \in A$, the set of atom pairs $D(u)$ can then be generated either with Eqs. 4 or 5. Consequently, the set of expected peaks for u is composed of two subsets,

$$E(u) = E^1(u) \cup E^2(u), \tag{15}$$

which represent, respectively, the situations where either the frequency of the heavy atom bound to the previously assigned hydrogen atom, h (p_i), or of the heavy atom bound to the unassigned hydrogen atom, h (u), is known:

$$E^1(u) \equiv \left\{ \overrightarrow{e^1(u)}_i \equiv \left(\Omega_u, \Omega_{h(p_i)}, \Omega_{p_i}\right) | \; (h(p_i), p_i) \in D(u); i = 1 \ldots M^1 \right\} \tag{16}$$

$$E^2(u) \equiv \left\{ \overrightarrow{e^2(u)}_i \equiv \left(\Omega_{p_i}, \Omega_{h(u)}, \Omega_u\right) | \; (h(u), p_i) \in D(u); i = 1 \ldots M^2 \right\} \tag{17}$$

In Eqs. 16 and 17, M^1 and M^2 denote the number of peaks in the two subsets (see also Eq. 7). The scoring function, F (Eq. 12), yields a scoring value for each potential resonance frequency, $\Omega_{u_k} \in R(u)$, which is calculated as follows:

$$F(\Omega_{u_k}) = \frac{1}{|M^1|} \sum_{i=1}^{M^1} C(\overrightarrow{e^1(u)}_i, \Delta\vec{\omega}) + \frac{1}{|M^2|} \sum_{i=1}^{M^2} C(\overrightarrow{e^2(u)}_i, \Delta\vec{\omega}) \tag{18}$$

In Eq. 18 the sums run over all the elements of the two subsets of expected peaks, $\overrightarrow{e^1(u)}_i \in E^1(u)$ and $\overrightarrow{e^2(u)}_i \in E^2(u)$.In addition to the three acceptance criteria formulated with Eqs. 13 and 14, each potential resonance frequency, $\Omega_{u_k} \in R(u)$, must result from at least one confirmed match between observed peaks and expected peaks from the subset $E^2(u)$, otherwise the value of the scoring function $F(\Omega_{u_k})$ is reset to zero. This additional requirement ensures that at least one observed peak used to identify the new resonance frequency of a previously unassigned hydrogen atom, u, is compatible with the known resonance frequency of its covalently bound heavy atom, $\Omega_{h(u)}$.

Secondly if the resonance assignments for a ^{1}H atom, u, and its covalently bound heavy atom, $h(u)$, are both unknown, with $u, h(u) \in P\backslash A$, then a two-step procedure is employed, whereby initially a set of potential resonance frequencies, $R(u)$, for the unassigned proton, u, is obtained, and subsequently a set of potential resonance assignments for $h(u)$, $R(\Omega_{u_k}; h(u))$ is identified, with $\Omega_{u_k} \in R(u)$.In the initial step, an analogous procedure to the one described above is applied, where the set of expected peaks involving the resonance frequency of the so far unassigned ^{1}H atom, u, is defined by

$$E(u) \equiv \left\{ \overrightarrow{e(u)}_i \equiv \left(\Omega_u, \Omega_{h(p_i)}, \Omega_{p_i}\right) | \; (h(p_i), p_i) \in D(u); i = 1 \ldots M \right\}. \tag{19}$$

The mapping function, Q, between the sets of expected peaks, $E(u)$, and the sets of observed peaks, $O(u) \subset S$, for a ^{1}H atom, u, yields a set of potential resonance frequencies, $R(u)$ (Eqs. 9 and 10), and for each potential resonance frequency, $\Omega_{u_k} \in R(u)$, a value of the scoring function, $F(\Omega_{u_k})$ is calculated with Eq. 12. Subsequently, a set of potential resonance assignments for the covalently bound heavy atom of the proton u, $h(u)$, is determined, with $\Omega_{u_k} \in R(u)$.The set of expected peaks correlating with $h(u)$ is then defined by

$$E(\Omega_{u_k}; h(u)) \equiv \left\{ \overrightarrow{e(\Omega_{u_k}; h(u))}_j \equiv \left(\Omega_{p_j}, \Omega_{h(u)}, \Omega_{u_k}\right) | p_j \in A; j = 1 \ldots M^{\Omega_{u_k}} \right\}. \tag{20}$$

A set of observed peaks, $O(\Omega_{u_k}; h(u)) \subset S$, is extracted from the updated list of the local extrema identified by a two-

dimensional grid spanned by the potential resonance frequencies of u, $\Omega_{u_k} \in R(u)$, and all the chemical shifts of the previously assigned hydrogen atoms, $p_j \in A$.The mapping between the set of expected peaks, $E(\Omega_{u_k};h(u))$, and the set of observed peaks, $O(\Omega_{u_k};h(u))$ with Eq. 9 then yields a set of potential resonance frequencies, $R(\Omega_{u_k};h(u))$ (Eq. 10). For each of these potential resonance frequencies, $\Omega_{h(u)_j} \in R(\Omega_{u_k};h(u))$, a value for the scoring function $F(\Omega_{h(u)_j})$is then calculated with (Eq. 12). At the end of the two-stage procedure for this combined hydrogen and heavy atom assignment, a set of potential resonance frequency pairs,

$$\Omega(u,h(u)) \equiv \left\{ (\Omega_{u_k},\Omega_{h(u)_j}) | \Omega_{u_k} \in R(u), \Omega_{h(u)_j} \in R(\Omega_{u_k};h(u)) \right\}, \tag{21}$$

is obtained for the unassigned ^{1}H atom, u, and its covalently bound heavy atom, $h(u)$. For each potential pair of resonance frequencies, $\left(\Omega_{u_k},\Omega_{h(u)_k}\right) \in R(u,h(u))$, two scoring values, $F(\Omega_{u_k})$ and $F(\Omega_{h(u)_j})$, are calculated with Eq. 12, and the atom pair u, $h(u)$, is added to the list of assigned atoms if the acceptance criteria of Eqs. 13 and 14 are met simultaneously for both potential resonance frequencies of the atom pair.

Thirdly, for ^{13}C–^{1}H fragments in aromatic rings, with, $u, h(u) \in P\backslash A$, the resonance assignments for the two atoms in the ^{13}C–^{1}H moiety can be obtained in a single step if the nearest-neighbor ^{13}C–^{1}H group in the aromatic ring has previously been assigned.

The set of assigned atom pairs used to derive the expected peak pattern for the so far unassigned aromatic ^{13}C–^{1}H moiety then consists of a single element, which is composed of the two assigned atoms H^δ and C^δ of the same aromatic ring:

$$D(u,h(u)) \equiv \left\{ (C^\delta, H^\delta) \right\} \tag{22}$$

The set of expected peaks for the unassigned aromatic ^{1}H atom, $u \in P\backslash A$, and its covalently bound heavy atom, $h(u) \in P\backslash A$, is then defined by

$$E(u,h(u)) \equiv \left\{ \left(\Omega_{(C^\delta,H^\delta)_i}, \Omega_{h(u)}, \Omega_u \right) | i = 1,\ldots,M \right\}, \tag{23}$$

where the frequencies $\Omega_{(C^\delta,H^\delta)_i}$ are determined using the local extrema that match the frequency coordinates given by Eq. 24 within a tolerance window $\Delta\vec{\omega} = (\Delta\omega^p, \Delta\omega^h, \Delta\omega^p)$ (see Table 1):

$$\vec{\omega}(C^\delta,H^\delta) = \Big((\omega_{(C^\delta,H^\delta)_i}, \omega_{2,i}, \omega_{3,i}) \big| |\omega_{2,i} - \Omega_{C^\delta}| < \Delta\omega^h \wedge |\omega_{3,i} - \Omega_{H^\delta}| < \Delta\omega^p; i = 1,\ldots,M \Big) \subset S \tag{24}$$

The mapping function, Q, between the set of expected peaks in Eq. 23, $E(u, h(u))$, and the set of observed peaks, $O(u, h(u))$ (Eq. 8), then yields a set of potential resonance frequencies, $R(u, h(u))$ (Eq. 10), and for each pair of potential resonance frequencies, $\left(\Omega_{u_k},\Omega_{h(u)_k}\right) \in R(u,h(u))$, a single value for the scoring function $F(\Omega_{u_k},\Omega_{h(u)_k})$is evaluated (Eq. 12). Finally, the aromatic ring ^{13}C–^{1}H moiety with $\left(\Omega_{u_k},\Omega_{h(u)_k}\right) \in R(u,h(u))$ is added to the list of assigned atom if the acceptance criteria of Eqs. 13 and 14 are met.

Appendix 2: Use of ASCAN with supplementary input of TOCSY data

In addition to accepting an input of 3D ^{13}C- or ^{15}N-resolved [^{1}H,^{1}H]-NOESY data, ASCAN is also laid out to operate on 3D heteronuclear-resolved [^{1}H,^{1}H]-TOCSY data sets. The correlation of an unassigned atom, $u \in P\backslash A$, to the set of assigned atom pairs, $D(u)$, is then based on the TOCSY magnetization transfer pathways. Sets of assigned atom pairs, $D(u)$, arising from scalar ("through-bond") coupling of an unassigned hydrogen atom, $u \in P\backslash A$, with assigned hydrogen atoms, $p_i \in A$, are derived from the covalent structure based on the fact that only pairs of hydrogen atoms, u and p_i, which are separated by a given number, n, of covalent bonds, $n_{up_i}^{cov}$, give rise to a scalar coupling in the same spin system. All correlated pairs of hydrogen atoms determined with Eq. 25,

$$2 \le n_{up_i}^{cov} \le 3, \tag{25}$$

are used to generate the elements of $D(u)$ (Eq. 3) with the Eqs. 4 and/or 5. Overall, the treatment of [^{1}H,^{1}H]-NOESY and [^{1}H,^{1}H]-TOCSY data by ASCAN differs only by the considerations given to the different coherence transfer pathways in the two experiments, as reflected in the Eqs. 6 and 25. In practice, TOCSY data sets can be a useful supplement to the NOESY input data, and they should be used exclusively in conjunction with NOESY data.

References

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

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

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

Bartels C, Güntert P, Billeter M, Wüthrich K (1997) Automated sequence-specific NMR assignment of homologous proteins using the program GARANT. J Comp Chem 18:139–149

Buchler NE, Zuiderweg ER, Wang H, Goldstein RA (1997) Protein heteronuclear NMR assignments using mean-field simulated annealing. J Magn Reson 125:34–42

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

Eghbalnia HR, Bahrami A, Tonelli M, Hallenga K, Markley JL (2005) Probabilistic identification of spin systems and their assignments including coil-helix inference output (PISTACHIO). J Am Chem Soc 127:12528–12536

Etezady-Esfarjabi T, Placzek WJ, Herrmann T, Wüthrich K (2006) Solution structures of the putative anti-sigma-factor antagonist TM1442 from Thermotoga maritima in the free and phosphorylated states. Magn Reson Chem 44:61–70

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

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

Gronwald W, Kirchfofer R, Gorler A, Kremer W, Gansmeier B, Neidig KP, Kalbitzer HR (1998) CAMRA: chemical shift based computer aided protein NMR assignments. J Biomol NMR 12:395–405

Güntert P (2003) Automated NMR protein structure calculation. Prog Nucl Magn Reson Spectrosc 43:105–125

Güntert P, Braun W, Wüthrich K (1991) Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. J Mol Biol 217:517–530

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

Hyberts SG, Wagner G (2003) IBIS—a tool for automated sequential assignment of protein spectra from triple resonance experiments. J Biomol NMR 26:335–344

Koradi R, Billeter M, Wüthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 14:51–55

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

Kraulis PJ (1989) ANSIG—a program for the assignment of protein H-1 2D NMR spectra by interactive computer graphics. J Magn Reson 24:627–633

Malmodin D, Papavoine CHM, Billeter M (2003) Fully automated sequence-specific resonance assignments using multi-way decomposition. J Biomol NMR 27:69–79

Michel E, Damberger FF, Chen AM, Ishida Y, Leal WS, Wüthrich K (2005) Assignments for the Bombyx mori pheromone-binding protein fragment BmPBP (1–128) at pH 6.5. J Biomol NMR 31:65

Moseley HNB, Montelione GT (1999) Automated analysis of NMR assignments and structures of proteins. Curr Opin Struct Biol 9:635–642

Moseley HNB, Monleon D, Montelione GT (2001) Automatic determination of protein backbone resonance assignments from triple resonance nuclear magnetic resonance data. Methods Enzymol 339:91–108

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

Nishiyama M, Horst R, Eidam O, Herrmann T, Ignatov O, Vetsch M, Bettendorff P, Jelesarov I, Grütter MG, Wüthrich K, Glockshuber R, Capitani G (2005) Structural basis of chaperone-subunit complex recognition by the type 1 pilus assembly platform FimD. EMBO J 24:2075–2086

Orekhov VY, Ibraghimov VI, Billeter M (2001) MUNIN: a new approach to multi-dimensional NMR spectra interpretation. J Biomol NMR 20:49–60

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

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

Wüthrich K (1986) NMR of proteins and nucleic acids. Wiley, New York

Wüthrich K, Billeter M, Braun W (1983) Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton–proton distance constraints with nuclear magnetic resonance. J Mol Biol 169:949–961

Zimmerman DE, Kulikowski CA, Huang Y, Feng W, Tashiro M, Shimotakahara S, Chien C, Powers R, Montelione GT (1997) Automated analysis of protein NMR assignments using methods from artificial intelligence. J Mol Biol 269:592–610