Spectroscopy 17 (2003) 559–568
IOS Press

Superposition of chemical shifts in NMR spectra can be overcome to determine automatically the structure of a protein

D. Auguin [a], V. Catherinot [a], T.E. Malliavin [b,*], J.L. Pons [a] and M.A. Delsuc [a]
[a] *Centre de Biochimie Structurale, INSERM U 414, CNRS UMR 5048, Université de Montpellier I, Faculté de Pharmacie, 15, av. Ch. Flahault, F-34060 Montpellier Cedex 2, France*
[b] *Laboratoire de Biochimie Théorique, UPR 9080, IBPC, 11, rue P. et M. Curie, F-75005 Paris, France*

Abstract. We are here addressing the problem of the automatic determination of a protein structure at atomic resolution, by using only the signal recorded on three spectra: 2D ^{15}N HSQC, 3D ^{15}N NOESY-HSQC and TOCSY-HSQC. A modified version of the neural network RESCUE (J.L. Pons and M.A. Delsuc, *J. Biomol. NMR* **15** (1999), 15–26), N15-RESCUE, is developed in order to predict the amino-acid type from only the ^{15}N, HN, Hα and Hβ chemical shifts. The spatial distances between protein residues are estimated by automatic comparison of columns extracted from a 3D ^{15}N NOESY-HSQC spectrum, using the FIRE method (T.E. Malliavin, P. Barthe and M.A. Delsuc, *Theor. Chem. Accts* **106** (2001), 91–97). The predictions provided by both FIRE and N15-RESCUE methods are then used for the determination of a preliminary NMR structure of the protein p8. A mean RMSD value of 2.31 ± 0.86 Å is observed between the coordinates of heavy atoms from helices αI and αII, and the αIII helix is taking random orientations with respect to the other helices. This random orientation is a consequence of the lack of predicted proximities between αIII and αII, and is in agreement with other independent observations made on p8 structure.

Keywords: Nuclear magnetic resonance, NMR, structure prediction, computer-aided assignment, data processing, neural network, chemical shift index

1. Introduction

Methods permitting the determination of protein structure through an automatic processing of their NMR spectra are of particular interest in the frame of structural proteomics studies as they could permit a rapid determination of the global fold, and could eventually allow high-throughput structure determination.

In the case of a double-labeled sample, methods for the automatic determination of spin system and of sequential assignment [1,2] have been proposed. But in other cases, assignment is still very dependent on a manual spectral analysis, thus presenting a major bottleneck for the use of NMR in structural proteomics. Also, the automatic assignment methods for double-labeled sample rely on the use of about 10 different NMR experiments [2,3] which makes the automatic assignment methods more extensive to use and more prone to errors in case of missing spectra.

For a ^{15}N single-labeled protein, a possible way towards gaining information on protein geometry is by direct analysis of the 3D ^{15}N NOESY-HSQC experiment which displays spatial proximity information while usually presenting little spectral superposition for the amide protons. A method for automatic processing of 3D ^{15}N NOESY-HSQC, the protocol FIRE, was recently proposed [4], and was shown

*Corresponding author. Fax: +33 1 58 41 50 26; E-mail: Therese.Malliavin@ibpc.fr.

to predict more than 70% of sequential NOE (nuclear Overhauser effect) contacts between residues of proteins with α and β secondary structures. FIRE is working through the calculation of a match matrix from the columns of a 3D ^{15}N NOESY-HSQC. The larger is the (i, j) element of the match matrix, the more closer should be the residues i and j in the protein. The match value is calculated by automatic comparison of the signal intensities between the columns of residues i and j in the 3D ^{15}N NOESY-HSQC spectrum. The match matrix is then filtered using a threshold, in order to keep, in each matrix row and column, a number of non-null match values into the range 4–4.5. Indeed, this range is close to the mean number of neighbors for a given residue in a protein (i.e., 4.5 neighbors for each residue).

In addition to the correctly predicted proximities, FIRE usually predict additional correlations between residues which are far away in the 3D space (false positive correlations). On the other hand, correlations between residues close in space are missing (false negative correlations). The false positive correlations come from the fortuitous chemical shift superpositions in the ^{1}H spectrum. These superpositions may then create positive match values for otherwise unrelated amide signals. On the other hand, the false negative correlations appear because the presence of false positive correlations induces an artificial increase of the threshold used to filter the match matrix.

A neural network, RESCUE [5], was designed to predict the type of a protein residue from its ^{1}H chemical shifts. This network achieved a mean rate of success of more than 80% on a test set of 8033 assigned amino-acid entries from the BMRB database [6]. The RESCUE network is here extended to the N15-RESCUE neural network which uses the ^{15}N, HN, Hα and Hβ chemical shifts for the prediction.

Both methods, FIRE and N15-RESCUE, are here used to compute a preliminary 3D structure of the p8$^{\mathrm{MTCP1}}$ protein, a protein of 68 amino-acids encoded by the MTCP1 oncogene [7]. The p8 main structural motif (8) consists of two anti-parallel helices, αI and αII spanning residues 8 to 20 and 29 to 40, strapped in an α-hairpin motif by the two disulfide bridges 7–38 and 17–28. The third helix αIII, spanning residues 48 to 63, is connected to the double-helix motif by a loop from residue 41 to 46, and a third disulfide bridge 39–50 links the top of helix αIII to the tip of helix αII. NMR relaxation [9], dipolar coupling measurements [10] and molecular dynamics simulations [11] showed that helix αIII displays an internal mobility in the nanosecond timescale, and that its orientation with the other helices is not well-defined.

2. Materials and methods

The 2D ^{15}N HSQC, 3D ^{15}N NOESY-HSQC and 3D ^{15}N TOCSY-HSQC spectra were recorded on p8$^{\mathrm{MTCP1}}$ with an AMX600 Bruker spectrometer, using previously described acquisition parameters [9]. Processing and handling of assignment data were realized with the Gifa assignment module [13,16].

The neural network N15-RESCUE was designed in the same way than the network RESCUE previously described to process the ^{1}H chemical shifts [5]. The network is based on a classical perception design [14] in which the input data (chemical shifts) are presented to the input layer, and results (the amino acid types) are obtained from the output layer. The optimized topology is a 3-layer network with 4 hidden neurons and 7 output neurons. The transfer and the reliability functions used here are the same than previously used [5]. The chemical shifts are entered into a fuzzy logic grid consisting of n entries, located at the positions δ_k regularly sampling the chemical shift axis with a spacing of $\Delta\delta$. For ^{1}H, the grid collapses the spectral data into 32 input neurons, spanning the -2 ppm to $+14$ ppm range in intervals of 0.5 ppm. For ^{15}N, the grid collapses the spectral data into 5 input neurons, spanning the 106.5 ppm to 131.5 ppm range in intervals of 5 ppm.

The network was tested on 18631 spin systems, extracted from the BRMB databank as described previously [5]. Each spin system was containing ^{15}N, HN, Hα and Hβ chemical shifts, according to the amino-acid type. This restriction was intended to allow the processing by N15-RESCUE of 3D ^{15}N TOCSY-HSQC spectra, in which the sensitivity does not permit the observation of chemical shifts for nuclei further than Hβ in the sidechain.

The 3D structure of p8^{MTCP1} was calculated, from the N15-RESCUE and FIRE results, with the CNS software version 1.0 [15], using a protocol of simulated annealing. The distance and dihedral angle measurements were applied through harmonic potential terms. A high-temperature (50 000 K) simulated annealing protocol (1000 steps) in torsion space, starting from an extended chain conformation, was used for the first stage of structure calculations. In this stage, the NOE and dihedral angle restraints were included with constant forces of 150 kcal Å^{-2} and 100 kcal rad^{-2}, respectively. The first stage was followed by an a annealing stage to cool down the system to 0 K by steps of 250 K. At each temperature step, 1000 steps of molecular dynamics were run with constant forces for distance and dihedral angle restraints of 150 kcal Å $^{-2}$ and 200 kcal rad^{-2}, respectively. Structures were then subjected to 10 cycles of 200-step Powell minimizations during which restraints with force constants of 75 kcal Å^{-2} and 400 kcal rad^{-2} were applied on inter-hydrogen distances and dihedral angles respectively.

3. The N15-RESCUE efficiency

The neural network N15-RESCUE determines the type of a protein residue from the chemical shift values of the following nuclei: ^{15}N, HN, Hα and Hβ. It was first tested on the set of 18631 spin systems extracted from the BRMB database [6]. The results obtained with the neural network were clustered: seven groups (IV, A, G, T, LKREQM, FYWHDNC (AMX) and S) are considered according to structural criteria and to the results of N15-RESCUE. The amino acids I and V were clustered together (group IV). Then, the amino-acids whose the sidechain contains more than two CH_2 groups, were put in the same group (LKREQM). Finally, the amino-acids for which the sidechain contains some excess of negative charges (π or n electrons) attached on the β carbone, are forming a last group (FYWHDNC or AMX).

The mean percentage of prediction success for all amino-acids is 77.2%, and the prediction rates for each amino-acid group are all larger than 70% (Table 1). The correlation matrix between target and found spin system groups (Fig. 1) shows almost no systematical error. The good prediction rates show that the ^{15}N, HN, Hα and Hβ chemical shifts (which correspond to the information available on a 3D ^{15}N TOCSY-HSQC) are sufficient to predict the amino-acid type. As the 3D ^{15}N TOCSY-HSQC and ^{15}N

Table 1
Success rates obtained with N15-RESCUE on the different groups of amino-acids

Group	Success rate (%)
All groups	77.2
IV	75.86
A	86.69
G	83.66
T	73.05
LKREQM	70.13
FYWHDNC (AMX)	79.33
S	71.10

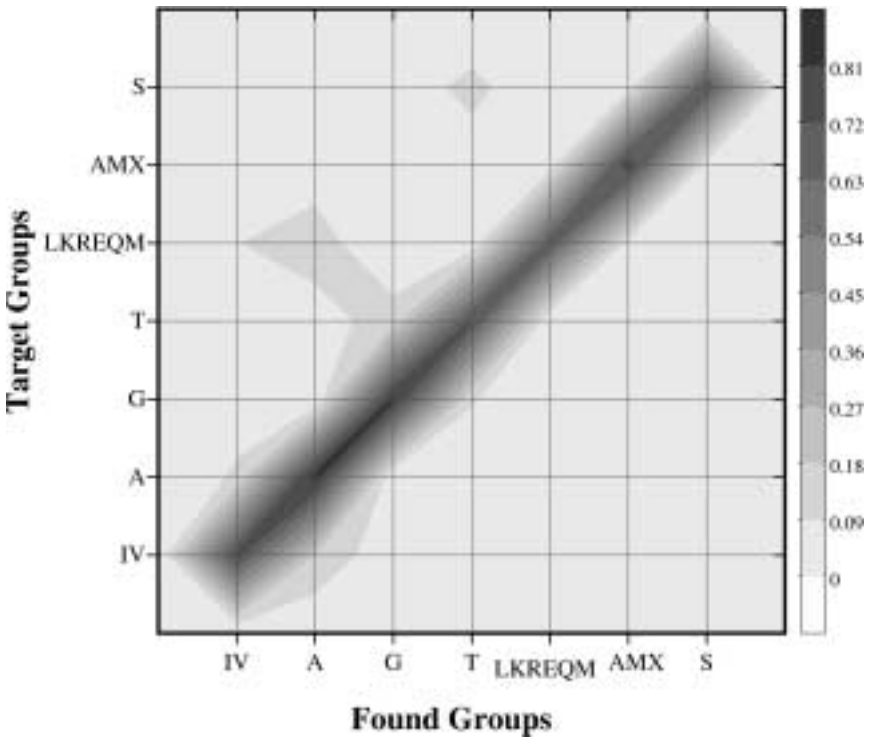

Fig. 1. Correlation matrix between the spin systems predicted with N15-RESCUE (Found groups) and true spin systems (Target groups). The set of chemical shifts used was selected from the BRMB databank as described in Materials and methods.

NOESY-HSQC spectra can be easily superposed, they are convenient to be used together in an automatic procedure.

The neural network was then used on a LTP (Lipid Transporter Protein) of 67 residues (J.L. Pons, F. de Lamotte, M.-F. Gautier and M.A. Delsuc in preparation); 44 spins systems were submitted to the neural network, from which 42 were correctly predicted. The N15-RESCUE network was also used on 65 spin systems (^{15}N, HN, Hα and Hβ chemical shifts) of p8. The prolines 2, 4 and 43 were excluded from the analysis. The neural network was able to correctly predict 38 spin systems. The false results were mainly LKREQM (12 predictions) and S (13 predictions) groups. Four IV and one FYWHDNC groups were also incorrectly predicted. The incorrectly predicted residues in the sequence are: Ala (5 cases over 5 residues), Cys (5 cases over 6), Val (3 cases over 3), Ile (2 cases over 2), Lys (2 cases over 9), Leu (3 cases over 3), Gln (2 cases over 8), Gly (2 cases over 2), Asn (2 cases over 2) and Phe (1 case over 1). The cysteine residues 39 and 50 are forming a disulfide bridge, and Cys 17 and 38 are taking part to disulfide bridges with Cys 28 and 7. The change of electronic environment due to the disulfide bridge induces certainly a variation of the chemical shift and can explain the inability of the neural network to predict the cysteine spin systems. All the residues Ala are predicted to be in the LKREQM group; this is not surprising, because of the correlation observed between A and LKREQM groups (Fig. 1) for the BMRB chemical shifts.

The residues for which 100% false prediction was made (Ala, Val, Ile, Leu, Gly, Phe, Asn) are clustered in the sequence (12, 14, 17–18, 20–21, 30–32, 35, 38–40, 45, 48–50, 52–53, 60–61, 66): the perturbations on chemical shifts due to the electronic environment and thus the difficulties for the neural network are propagating through the 3D space. Some residue groups are located close to disulfide bridges (17–18,

20–21, 30–32, 35, 38–40, 48–50, 52–53), others are at the C-terminal part (60–61, 66). These residues are belonging to the helix αIII, which exhibits conformational changes [9–11], and consequently perturbations of chemical shifts.

4. Calculation of p8 match matrix

The FIRE method [4] was used to estimate the similarity of the NOE columns extracted from the 3D ^{15}N NOESY-HSQC spectrum. A filtering window was applied to each column corresponding to a peak on the 2D ^{15}N HSQC spectra (i.e., to a residue), in order to cancel out the spectral intensities located at the water frequency. As a matter of fact, the water signal is observed on almost all the columns, and if not removed, induces a bias into the result of FIRE. Then a noise σ value was measured on each column, and all the column intensities smaller than 5σ were canceled. Finally, the columns containing less than 2 peaks, and the columns for which the mean intensity is smaller than 0.0005 times the mean intensity of all the columns, were discarded. The remaining columns were peak-picked; the central location of each peak was replaced by 1, and the other locations were set to zero. The obtained columns $C^{\text{tf}}(k)$ are thus formed from values 0 and 1. For each column pair $C^{\text{tf}}(i)$, $C^{\text{tf}}(j)$, we calculated a value M_{ij} which quantifies the spectral similarity (match) between the residues i and j:

$$M_{ij} = 2\langle C^{\text{tf}}(i) \mid C^{\text{tf}}(j)\rangle / \left(\|C^{\text{tf}}(i)\|^2 + \|C^{\text{tf}}(j)\|^2\right). \tag{1}$$

$\langle|\rangle$ is the scalar product between the two column vectors, $\|\|$ the column vector norm. If $C^{\text{tf}}(i)$ and $C^{\text{tf}}(j)$ are both null columns, M_{ij} is set equal to 0.

M_{ij} takes its value between 0 and 1. Two identical columns will produce an M_{ij} value equal to 1, whereas columns having no peak facing each other will give a null M_{ij} value. The M_{ij} value does not depend on the order of columns. A symmetric square matrix M, the match matrix, is built from the M_{ij} values. The M matrix is then filtered by applying a threshold γ such that for each residue i, the number of residues j producing a match value M_{ij} larger than γ, is in the 4–4.5 range. In all the following calculations, all the non-null match values in the filtered match matrix are replaced by 1.

The match matrix M is compared to a proximity matrix Δ, obtained from the parameters Δ_{ij}. Each Δ_{ij} value is calculated between all the non-proline residues i and j as the minimum distance between an amide hydrogen and the hydrogens of the other residue:

$$\Delta_{ij} = \text{Min}(\text{H}_i, \text{H}_j)\big(d(\text{HN}_i, \text{H}_j), d(\text{HN}_j, \text{H}_i)\big), \tag{2}$$

where H_i, H_j are the hydrogens, and HN_i, HN_j the amide hydrogens of residues i and j; d is the 3D Euclidean distance; $\text{Min}(\text{H}_i, \text{H}_j)$ is the minimum value on the overall set of hydrogens from residues i and j. The exchangeable hydrogens, aromatic hydrogens and Hε hydrogens from lysines are excluded from the calculation. Δ verifies the distance properties [16] and corresponds to the maximum distance information which can be extracted from an isolated 3D NOESY-HSQC. The proximity matrix Δ is built from the Δ_{ij} values, by replacing all the Δ_{ij} larger (resp. smaller) than 4 Å by zero (resp. one).

The matrix M of p8 (Fig. 2a) is very similar to the corresponding proximity matrix Δ (Fig. 2b); 88% of sequential, 40% of medium – range and 9% of long-range correlations are properly predicted. All the predicted long-range correlations are false except for the pair of residues (14, 32), which is marked with a star (Fig. 2b). On the other hand, long-range proximities observed in the Δ matrix between residues

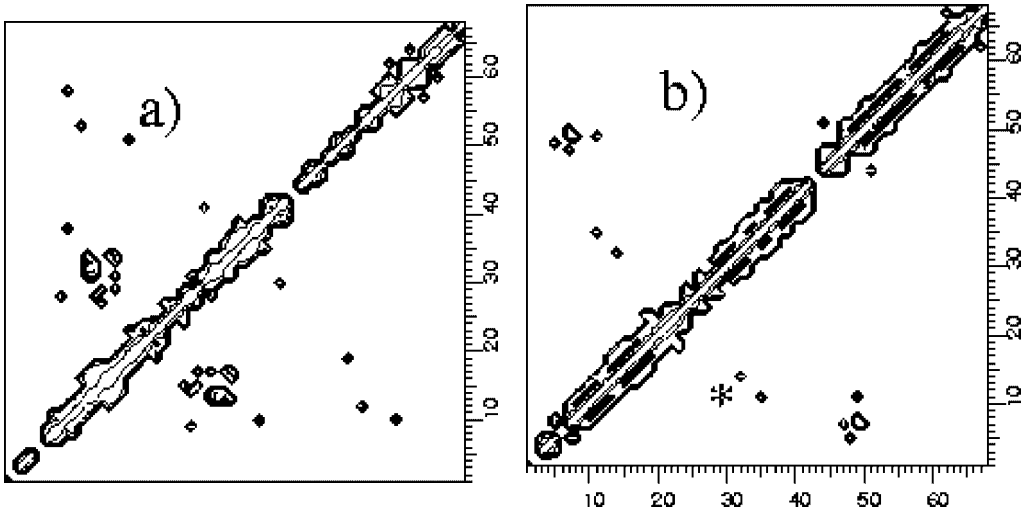

Fig. 2. Comparison of the match (a) and the Δ proximity matrices (b) obtained for p8. The proximity between residues i and j is equal to 1 if the corresponding Δ_{ij} value is smaller than 4 Å, and is equal to 0 otherwise. The proximity value between residues 14 and 32 is marked with a star. The x and y axes display the residue number.

5–11 and 47–50 are not predicted by FIRE. Two sub-matrices including residues 1–5 and residues 44–47 (Fig. 2a) show no correlations with other protein residues: the separation of these sub-matrices from the rest of the matrix is due to the break in correlation graph produced by prolines 6 and 43. It is thus not possible to determine directly from the match values, the p8 fold in 3D space, and the sequential assignment of p8 has first to be determined.

5. Determination of the p8 sequential assignment

The match values and the N15-RESCUE predictions were used together to determine the p8 sequential assignment. For a given permutation $P = (p_1, p_2, \ldots, p_N)$ of the spin systems extracted from the 3D ^{15}N NOESY-HSQC, the function F is defined as:

$$F = \sum_{i=1}^{N} P_{\text{RESCUE}}(i, p_i) + \sum_{k=1}^{N} \sum_{l=1}^{k} (M_{pk,pl} - \mu_{kl})^2, \tag{3}$$

where: N is the number of protein residues, $P_{\text{RESCUE}}(i, p_i)$ is the probability given by N15-RESCUE that the residue p_i in the permutation has the amino-acid type of the residue i in the sequence, M_{kl} is the (k, l) element of the match matrix, and μ_{kl} is the (k, l) element of a model proximity matrix.

The model proximity matrix describes the a priori hypothesis, which can be made on the protein proximity Δ matrix. Without any hint of the protein structure, we used as a minimal-knowledge hypothesis, a band-diagonal matrix, where:

$$\begin{aligned} &\mu_{kl} = 1.0 \quad \text{for } |k - l| \leqslant 3, \\ &\mu_{kl} = 0.0 \quad \text{otherwise}. \end{aligned} \tag{4}$$

The sequential assignment of p8 is determined in the following way. The permutation of the spin systems is modified in order to minimize the F function, by a Metropolis algorithm [17]. The starting

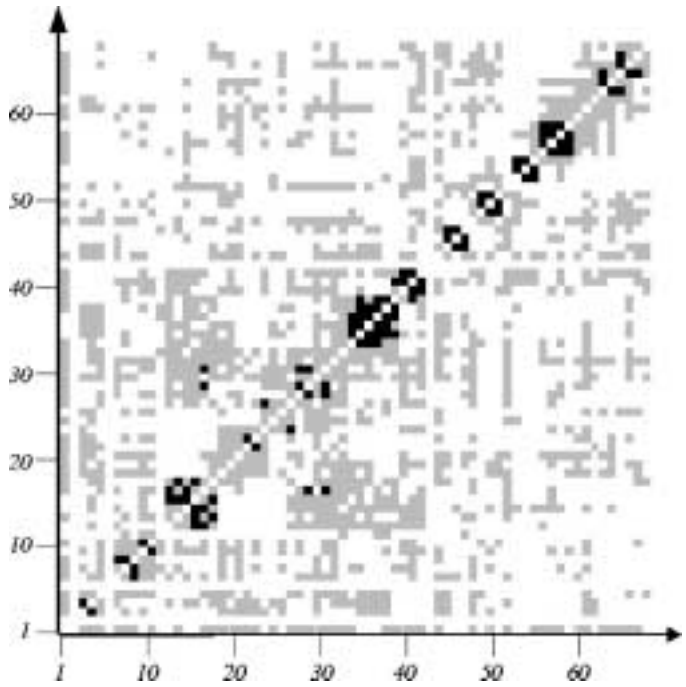

Fig. 3. Superposition of the match matrices corresponding to the best 100 values of the function F. The (i, j) element of the matrix shows the sum of the match values predicted between the residues i and j. White is 0, light gray corresponds to a sum value smaller than 70, and black to a sum value larger than 70. The x and y axes display the residue number.

permutation is chosen randomly. At each step, two residues are inverted into the permutation. The F_{new} value of the function F is calculated for the new permutation and compared to the old value F_{old} of the function. If $F_{\text{old}} \leqslant F_{\text{new}}$, the old permutation is kept, otherwise the new permutation is accepted with a probability $\exp(-(F_{\text{old}} - F_{\text{new}})/T)$, where T is the system temperature.

5000 random permutations were starting points for the minimization. For each starting permutation, 10 000 steps of Metropolis algorithm were performed, with a temperature value of 0.007. The 500 final values of the function F were into the 0.48–0.34 range, and the permutations giving rise to the first 100 best values for F (into the 0.48–0.40 range) were kept for analysis.

The sum of the match matrices obtained for the 100 best permutations is shown in Fig. 3; the match values larger than 70 are drawn in black, the other non-null match values are in gray. The threshold of 70 permits to detect the long-range proximities between helices αI and αII, for residue pairs 17–31, and 17–29. The corresponding spatial Δ_{ij} value is 3.8 Å between the residues 17 and 31, and 9.3 Å between the residues 17 and 29. Other long-range proximities (between helices αII and αIII) are unattainable anyway, because they are absent in the initial match matrix. The other false positive correlations observed in the M matrix (Fig. 2a) correspond to values smaller than 70 in the superposed match matrices (Fig. 3).

6. Calculation of a p8 preliminary 3D structure

The positive values displayed in the match matrices for more than 70% of the 100 best permutations, were used to produce 53 distance restraints between Cα carbones with lower bounds of 2.0 Å and upper bounds of 6.0 Å. The chemical shift index (CSI) values [18] were calculated for each residue from the

best 100 permutations obtained with the Metropolis algorithm. The CSI value was calculated from the difference between the residue Hα chemical shift and the corresponding random coil chemical shift of the amino-acid [19]. It was set equal to 1 if the difference is larger than 0.3 ppm, and to -1 if the difference is smaller then -0.3 ppm.

CSI values of 1 displayed in more than 70% of the 100 best permutations, were used as a prediction of α secondary structure, for sequences regions 5, 7–8, 10–19, 21–22, 25, 29–42, 45, 52 and 54. No CSI values of -1 were found. According to the CSI values found, distances restraints (66 restraints) for O–HN $(i+4, i)$ (1.8–2.6 Å), Hα–HN (i, i) (3.1–3.9 Å), Hα–HN $(i, i+2)$ (4.0–4.8 Å), Hα–HN $(i, i+3)$ (3.0–3.8 Å), HN–HN $(i, i+3)$ (2.2–3.4 Å), HN–HN $(i, i+2)$ (3.8–4.6 Å), and restraints $-150/-100°$ on the ϕ dihedral angles (94 restraints) were added to the previous restraint list. The distance restraints were applied on i only if CSI values of 1 are obtained for the residues i and $i+4$. The dihedral angle restraints were applied on the angle ϕ of each residue i with a CSI value of 1. The total number of distance restraints was 203, and the number of dihedral angle restraints was 74.

Ten conformers of p8 were calculated using the list of 330 restraints previously determined; all the conformers display no restraint violations, and were further analyzed. The 7–38 region of the sequence corresponding to the two antiparallel helices αI and αII, displays mean RMSD value of 2.31 ± 0.86 Å between the coordinates of backbone atoms, which is consistent with the determination of a preliminary structure at atomic resolution. On the other hand, the superposition of the conformer coordinates by fitting on the αI and αII regions, shows that the orientation of the helix αIII is not determined (Fig. 4). This incertitude on the position of the third helix is a consequence of the lacking match values between residues of αIII and residues of αI–αII. Nevertheless, as dipolar coupling measurements [10] and molecular dynamics simulations [11] showed an internal mobility of the helix αIII, the incertitude on its position is an intrinsic feature of the protein structure.

7. Conclusion

Two methods to perform the automatic analysis of 3D ^{15}N NOESY-HSQC and TOCSY-HSQC experiments, were here presented and applied on a 68-residue protein. First, we implemented a neural network N15-RESCUE to predict the amino-acid type from the ^{15}N and ^{1}H chemical shifts of a protein residue. This network was tested on a set of chemical shifts from the BRMB databank, and allowed the mean correct prediction of 77.2% of amino-acid groups. This percentage makes possible the use of this method in real-life cases. N15-RESCUE consists of a set of programs written in the perl language. A CGI program implementing the approach can be used from our Web site at: http://www.infobiosud.cnrs.fr/rescue.html.

Second, the results obtained with N15-RESCUE and FIRE methods on p8^{MTCP1} were used to: (i) determine the protein sequential assignment (ii) calculate a preliminary 3D structure of the protein. The sequential assignment was determined, despite several false positive correlations in the match matrix. This is due to the favorable shapes of the Δ and of the match matrices, which are almost diagonal-band. This feature made possible the convergence of the procedure, as the match matrix is close to a priori hypothesis on the μ matrix.

A preliminary 3D structure of p8 was then calculated, using: (i) the long-range proximities predicted with the FIRE method, (ii) the prediction of the type of secondary structure, based on the CSI values [18] calculated using the sequential assignment and the Hα chemical shifts. The RMSD between the calculated set of conformer coordinates, is 2.31 ± 0.86 Å in the helices αI and αII of the protein, which form the most rigid part of the structure. This value of RMSD is typical of a preliminary 3D structure, and

Fig. 4. Ten superposed p8 conformers obtained, using the simulated annealing protocol, and displayed with ribbons. The helices αI and αII were superimposed in order to minimize the RMSD between the coordinates of their backbone coordinates. The helix αIII, at the C terminal end of the protein, exhibits a wide range of orientations with respect to the hairpin formed by αI and αII.

indicates a convergence of atomic coordinates along the NMR restraints. In the calculated conformers, the helix αIII can take all the possible orientations with respect to αI and αII, which is consistent with independent experimental observations on p8 [9,10].

The automatic assignment methods are now giving rise to much interest [20–22], in particular in the frame of structural proteomics projects. A major obstacle to the development of these methods on a large scale, is the presence of chemical shifts superpositions, which produces ambiguities and false positive correlations in automatic methods. We are here showing a favorable example, where such ambiguities can be overcome during the automatic spectral processing, in order to produce a preliminary structure.

Acknowledgements

The authors gratefully acknowledge Drs. Christian Roumestand and Philippe Barthe for providing the experimental data recorded on p8. CNRS, INSERM and Université Montpellier-1 are acknowledged for funding.

References

[1] R. Bernstein, C. Cieslar, A. Ross, H. Oschkinat, J. Freund and T.A. Holak, Computer-assisted assignment of multidimensional NMR spectra of proteins – application to 3D NOESY-HMQC and TOCSY-HMQC spectra, *J. Biomol. NMR* **3** (1993), 245–251.

[2] D.E. Zimmerman and G.T. Montelione, Automated analysis of nuclear magnetic resonance assignments for proteins, *Curr. Opin. Struct. Biol.* **5** (1995), 664–673.

[3] A. Medek, E.T. Olejniczak, R.P. Meadows and S.W. Fesik, An approach for high-throughput structure determination of proteins by NMR spectroscopy, *J. Biomol. NMR* **18** (2000), 229–238.
[4] T.E. Malliavin, P. Barthe and M.A. Delsuc, FIRE: predicting the spatial proximity of protein residues from a 3D HSQC-NOESY, *Theor. Chem. Accts* **106** (2001), 91–97.
[5] J.L. Pons and M.A. Delsuc, RESCUE: an artificial neural network tool for the NMR spectral assignment of proteins, *J. Biomol. NMR* **15** (1999), 15–26.
[6] B.R. Seavey, E.A. Farr, W.M. Westler and J.L. Markley, A relational database for sequence-specific protein NMR data, *J. Biomol. NMR* **1** (1991), 217–236.
[7] M.H. Stern, J. Soulier, M. Rozenzwajg, K. Nakahara, N. Canki-Klain, A. Aurias, F. Sigaux and I.R. Kirsch, Mtcp-1: a novel gene on the human chromosome xq28 translocated to the T cell receptor α/δ locus in mature T cell proliferations, *Oncogene* **12** (1993), 379–386.
[8] P. Barthe, Y.S. Yang, L. Chiche, F. Hoh, M.P. Strub, L. Guignard, J. Soulier, M.H. Stern, H. van Tilbeurgh, J.M. Lhoste and C. Roumestand, Solution structure of human $p8^{MTCP1}$, a cysteine-rich protein encoded by the Mtcp1 oncogene, reveals a new alpha-helical assembly motif, *J. Mol. Biol.* **274** (1997), 801–815.
[9] P. Barthe, L. Chiche, N. Declerck, M.A. Delsuc, J.-F. Lefèvre, T. Malliavin, J. Mispelter, M.H. Stern, J.M. Lhoste and C. Roumestand, Refined solution structure and backbone dynamics of ^{15}N-labeled C12A- $p8^{MTCP1}$ studied by NMR, *J. Biomol. NMR* **15** (1999), 271–288.
[10] H. Déméné, T. Ducat, P. Barthe, M.A. Delsuc and C. Roumestand, Structure refinement of flexible proteins using dipolar couplings: application to the protein $p8^{MTCP1}$, *J. Biomol. NMR* **22** (2002), 47–56.
[11] P. Barthe, C. Roumestand, H. Déméné and L. Chiche, Helix motion in protein C12A- $p8^{MTCP1}$: comparison of molecular dynamics simulations and multifield NMR relaxation data, *J. Comput. Chem.* (2003), in press.
[12] J.L. Pons, T.E. Malliavin and M.A. Delsuc, Gifa v 4: A complete package for NMR data set processing, *J. Biomol. NMR* **8** (1996), 445–452.
[13] T.E. Malliavin, J.L. Pons and M.A. Delsuc, An NMR assignment module implemented in the Gifa NMR processing program, *Bioinformatics* **14** (1998), 624–631.
[14] D.E. Rumelhart, G.E. Hinton and R.J. Williams, *Parallel Distributed Processing*, MIT Press, Cambridge, MA, 1986.
[15] A.T. Brunger, P.D. Adams, G.M. Clore, W.L. Delano, P. Gros, R.W. Grosse-Kunstleve, J.-S. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson and G.L. Warren, Crystallography & NMR system, 2000.
[16] G.M. Crippen and T.F. Havel, *Distance Geometry and Molecular Conformation*, Research Studies Press Ltd., Taunton, 1988.
[17] S. Kirkpatrick, C.D. Gelatt Jr. and M.P. Vecchi, Optimization by simulated annealing, *Science* **220** (1983), 671–680.
[18] D.S. Wishart and B.D. Sykes, The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data, *J. Biomol. NMR* **4** (1994), 171–180.
[19] J.N.S. Evans, *Biomolecular NMR Spectroscopy*, Oxford University Press, 1995.
[20] R.A. Atkinson and V. Saudek, Hypothesis: The direct determination of protein structure by NMR without assignment, *FEBS Lett.* **510** (2002), 1–4.
[21] A. Grishaev and M. Llinás, CLOUDS, a protocol for deriving a molecular proton density via NMR, *Proc. Natl. Acad. Sci.* **99** (2002), 6707–6712.
[22] A. Grishaev and M. Llinás, Protein structure elucidation from NMR proton densities, *Proc. Natl. Acad. Sci.* **99** (2002), 6713–6718.