Journal of Biomolecular NMR, 5 (1995) 154–160
ESCOM

J-Bio NMR 228

# Computer assignment of the backbone resonances of labelled proteins using two-dimensional correlation experiments

Nathalie Morelle, Bernhard Brutscher, Jean-Pierre Simorre and Dominique Marion*

*Institut de Biologie Structurale-Jean Pierre Ebel, C.E.A.-C.N.R.S., 41 Avenue des Martyrs, F-38027 Grenoble Cedex, France*

Received 7 July 1994
Accepted 22 August 1994

*Keywords:* Protein; Resonance assignment; Automated assignment; Heteronuclear assignment

**Summary**

We present ALPS (Assignment for Labelled Protein Spectra), a flexible computer program for the automatic assignment of backbone NMR resonances of $^{15}N/^{13}C$-labelled proteins. The program constructs pseudoresidues from peak-picking lists of a set of two-dimensional triple resonance experiments and uses either a systematic search or a simulated annealing-based optimization to perform the assignment. This method has been successfully tested on two-dimensional triple resonance spectra of *Rhodobacter capsulatus* ferrocytochrome $c_2$ (116 amino acids).

## Introduction

Assignment of backbone and side-chain resonances is essential for conformational and dynamic studies of proteins. This can be achieved for small proteins by homonuclear NMR (Wüthrich, 1986), or by heteronuclear NMR when $^{13}C$ and $^{15}N$ labelling is available (Kay et al., 1990; Clore and Gronenborn, 1991; Clubb et al., 1992; Grzesiek and Bax, 1992). The assignment of a large protein (≥ 10 kDa) usually takes several months, including several weeks of spectrometer time and manual spectral analysis. This process can be accelerated by cutting down the length of the data acquisition and/or automating the data interpretation. By reducing the dimensionality of the heteronuclear correlation spectra, the acquisition time can be substantially shortened (Szypersky et al., 1993a,b; Simorre et al., 1994). Recently, we have presented a complete set of 2D correlation experiments and shown its performance on medium-sized proteins (Brutscher et al., 1994); each experiment correlates three nuclei of the backbone without sensitivity loss with respect to the corresponding 3D experiment recorded in the same experimental time, and in a much more convenient form.

Many attempts to automate the spectral analysis and the sequential assignment process have been made in the past, most of these using the classical $^1H$ NOE-based approach (Cieslar et al., 1988,1990; Catasti et al., 1989; Eads and Kuntz, 1989; Van der Ven, 1990; Eccles et al., 1991; Kleywegt et al., 1991,1993; Bernstein et al., 1993; Xu et al., 1994). However, the identification of sequential connectivities from a NOESY map is very difficult due to the presence of long-range interactions and peak overlap, and this strategy has consequently met with limited success so far. On the other hand, heteronuclear NMR data from triple resonance experiments are much better suited to automation (Ikura et al., 1990). Recently, new semi-automatic assignment algorithms have been published (Meadows et al., 1994; Olson and Markley, 1994; Zimmerman et al., 1994).

For large proteins, it is necessary to collect a wide set of experimental NMR information; otherwise, the number of configurations that should be tested by the computer during the sequential assignment procedure rules out any systematic or heuristic approach. On the other hand, for smaller proteins (typically around 100 residues) with good signal dispersion in the NMR spectrum and slow relaxation, the resolution of ambiguities can be performed by the computer search instead of by numerous experiments. As the digital resolution of 2D experiments can be better than that of 3D experiments, the uncertainty in peak positions is smaller, and thus the number of ambiguities for the assignment is substantially reduced. Here we dem-

*To whom correspondence should be addressed.

0925-2738/$ 6.00 + 1.00 © 1995 ESCOM Science Publishers B.V.

onstrate that a small number of 2D heteronuclear experiments is sufficient for an automatic assignment of the backbone resonances.

The computer program ALPS (Assignment of Labelled Protein Spectra) compensates for incomplete information by searching for the correct alignment among all possible combinations in agreement with the NMR data, a prohibitive task if performed manually. As the chemical shift of carbon nuclei depends on the nature of the residue (Richarz and Wüthrich, 1978; Wishart and Sykes, 1994), its use further reduces the number of potential combinations. The standard database for our program comprises the five most sensitive experiments, i.e.: ($^1$H,$^{15}$N)-HSQC, 2D-HNCO, 2D-HN(CO)CA, 2D-H(N)COCA and 2D-HNCA. Other correlation experiments (involving correlations to $H^{\alpha}$ or $C^{\beta}$) and supplementary spin system information (obtained by $^1$H-$^1$H or $^{13}$C-$^{13}$C TOCSY experiments) can be added, without modification of the program, whenever the constraints are insufficient to yield the correct assignment.

With this approach, we were able to (re)-assign the backbone resonances of *Rhodobacter capsulatus* ferrocytochrome $c_2$, a diamagnetic protein of 116 residues (Caffrey et al., 1994) in an overall (experimental and computational) time of about one week.

## Materials and Methods

The method was tested on doubly $^{15}$N- and $^{13}$C-labelled (labelling > 95%) *Rhodobacter capsulatus* ferrocytochrome $c_2$ at a concentration of 1 mM in 45 mM $K_2HPO_4$/$KH_2PO_4$ buffer, pH = 6.0. Assignments of $^1$H and $^{15}$N (Gooley et al., 1990) as well as $^{13}$C and $^{13}$CO resonances (Caffrey et al., 1994) have been published previously. All NMR experiments were carried out on a Bruker AMX-600 spectrometer at 303 K. The computer program was written in ANSI C and the calculation times referred to were obtained on a Silicon Graphics 4D-35 workstation.

### *Experimental data set*

The following experiments were recorded, as described elsewhere (Brutscher et al., 1994): (1) ($^1$H,$^{15}$N)-HSQC with the refocusing delay of the $^1$H-$^{15}$N INEPT transfer tuned to filter out the $NH_2$ side-chain resonances (Van Doren et al., 1993); a set of (N → CO) experiments: (2) 2D-HNCO, (3) 2D-HN(CO)CA, (4) 2D-H(N)COCA, (5) 2D-HN(COCA)H; and a set of (N → $C^{\alpha}$) experiments: (6) 2D-HNCA, (7) 2D-HN(CA)CO and (8) 2D-HN(CA)H. The overall experimental times were 5 h for the 2D-HNCO, 8 h for 2D-HNCA, 2D-HN(CO)CA and 2D-H(N)COCA,

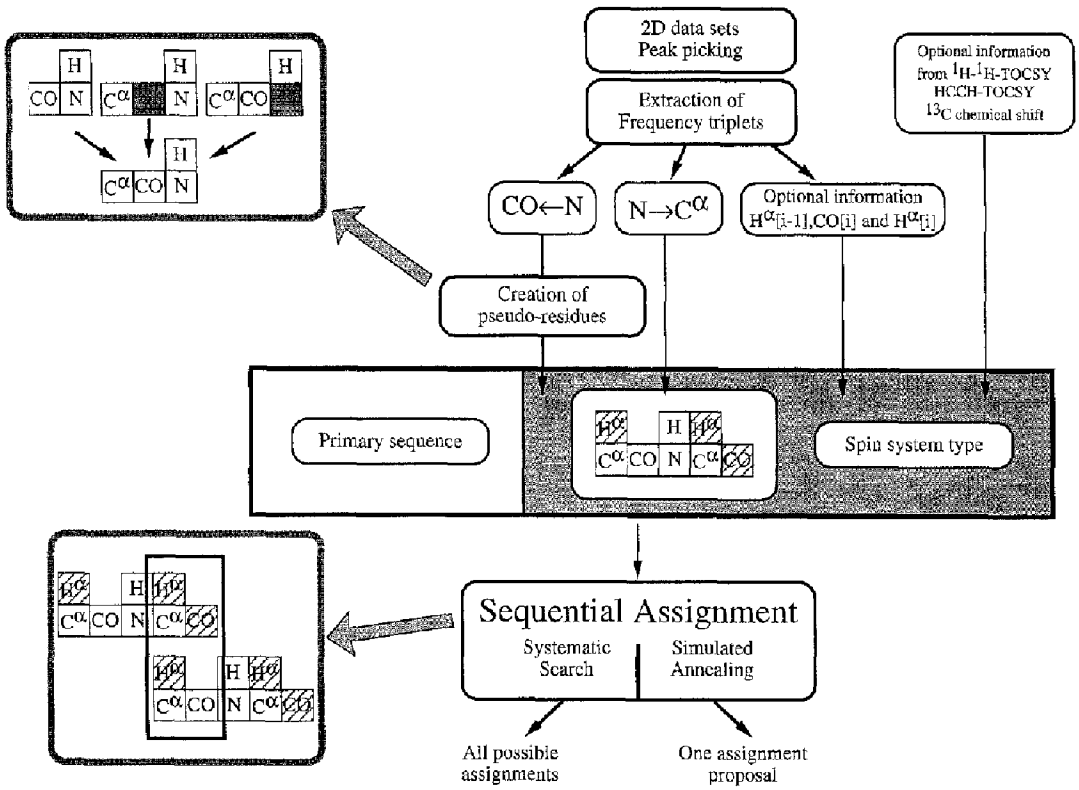


Fig. 1. Flow chart of the ALPS computer program. Following the peak picking of the 2D data sets, frequency triplets are extracted from the (N → CO) and (N → $C^{\alpha}$) transfer experiments. Pseudoresidues (illustrated in the upper left inset) are constructed using only unambiguous (N → CO) experiments and are then further extended, when possible, by additional information. The sequential assignment procedure employs two types of information: the primary sequence and the pseudoresidues (extended towards the C-terminus and complemented by spin system information). It aims at matching common parts of pseudoresidues (as shown in the lower left inset) by using either a systematic search or simulated annealing (see text).

10 h for the 2D-HN(CA)H and 18 h for the two less sensitive experiments. As discussed below, 2D-HN(CA)-CO, 2D-HN(COCA)H and 2D-HN(CA)H are optional for the automated assignment protocol described in this paper.

*Automated analysis*

The data were processed using the FELIX program, version 2.3 (Hare Research Inc., Biosym Technologies Inc.). Peak lists (frequencies and intensities) were used as input for the assignment program. The standard peak-picking routine of FELIX was followed by a quick visual check. On the basis of peak shapes only, rapid identification of overlap can eliminate ambiguities. This saves time in later processing stages.

The protocol consists of three independent steps: (i) identification of frequency triplets; (ii) construction of 'pseudoresidues'; and (iii) sequential assignment by either a systematic search algorithm or an optimization procedure. A schematic diagram of the program is presented in Fig. 1.

*(i) Identification of frequency triplets*

The third frequency in 2D triple resonance correlation experiments has to be extracted from the distance of two symmetrical peaks with respect to a given center peak (Simorre et al., 1994). Thus, two frequency lists are required, one containing the peak picking of the 2D triple resonance spectrum (MQ-peak file) to be analyzed and one with all possible center peaks (SQ-peak file).

First, a routine searches for all pairs of symmetrical (MQ) peaks of similar intensities with respect to a given SQ peak. Then these triplets (one SQ and two MQ peaks) are filtered to remove redundancies (due to peak-picking artifacts and/or accidental resonance overlap) using best-alignment criteria. Although at this stage the program might still have several triplets for a single SQ peak, it aims at retaining a unique triplet. When triplets overlap, all possible combinations are analyzed, as shown in Fig. 2: the combination using the maximal number of MQ peaks is selected. If two combinations have similar scores, their two sets of triplets are retained.

The program output is an ASCII file, listing all the correct triplets as well as the corresponding frequencies of the three nuclei. A warning is issued for each ambiguity related to the duplicate use of any peak. These cases have to be checked, either by inspection of the contour plots or by a second run of the program with slightly different parameters (the tolerance for misalignments or intensity deviations may be revised, as may the peak-picking thresholds).

For the experiments involving an $N \rightarrow C^{\alpha}$ magnetization transfer, correlations can be observed with the previous residue as well, via the $^2J_{NC\alpha}$ coupling. As a result, peaks detected in ($N \rightarrow CO$) experiments (such as HN-(CO)CA, HNCO and HN(COCA)H) may also be present in the corresponding ($N \rightarrow C^{\alpha}$) experiments. Thus, if among the triplets found in the latter experiments some are identical to those found via the $N \rightarrow CO$ transfer, a routine removes them before any further filtering occurs.

*(ii) Construction of pseudoresidues*

This step and the following could as well be run on 3D data sets, using peak-picking files with consistent $H^N$ and N chemical shift values. It comprises the construction of ($H_i^N$, $N_i$, $CO_{i-1}$, $C_{i-1}^{\alpha}$) frequency quartets, which we call *pseudoresidues*. To achieve this, the information of HNCO, HN(CO)CA and H(N)COCA experiments has to be combined. We chose the most sensitive spectrum, i.e. the HNCO, to be processed first, as it should contain the largest number of peaks with the least overlap. The frequency triplets can be extracted in the manner described above with an HSQC peak picking as SQ-peak file. As overlap may occur in the HSQC spectrum, several sets of triplets might be retained per SQ-peak. The degeneracy of the HSQC peaks is removed by the obtained HNCO frequency triplets; therefore this list is retained for further analysis of all remaining spectra by using either the $H^N$,N or the $H^N$,CO frequency pairs as SQ peaks.

The only difference between the HN(CO)CA and the H(N)COCA experiment lies in a different labeling. The corresponding triplets are associated with HNCO triplets sharing respectively exactly the same $H^N$,N and $H^N$,CO frequencies; then the $C^{\alpha}$ frequencies from both experiments are compared. The redundancy of these data sets

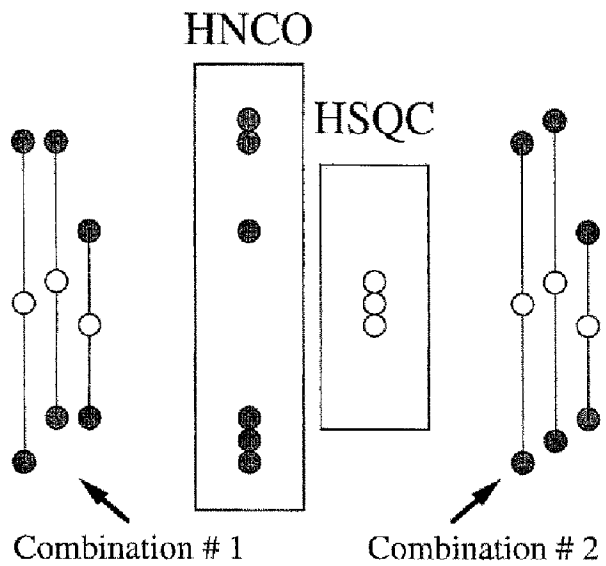


Fig. 2. Resolution of overlaps during the construction of triplets. Six peaks in the 2D-HNCO spectrum have the same $H^N$ chemical shift, whereas three peaks are present in the HSQC spectrum. If combination #1 is chosen, some peaks of the 2D-HNCO remain unused, while all peaks are assigned in combination #2. The latter is thus selected. The best alignment criterion can also be used for resolving near-overlap situations.

resolves the ambiguities for overlapping HSQC peaks, and in the other cases confirms the assignment. The probability of automatically generating a false pseudoresidue is small, as this implies correlated mistakes in all three triplet lists. This is very important, because these pseudoresidues are the elements which later have to be placed in the sequence correctly. If some HN(CO)CA or H(N)-COCA triplets have not been used, the user might reinspect the triplet lists and the plots, and accept or reject the pseudoresidues without redundancy.

Optionally, the HN(COCA)H triplets can be used here and the $H^{\alpha}_{i-1}$ frequency stored. It should be noted that the use of this data set, and of all the following, relies on the lack of ambiguities for the ($H^N$,N) pairs; as a result, when two pseudoresidues overlap on $H^N$ *and* N, two triplets are expected and both possible values have to be assigned to each of the $H^{\alpha}_{i-1}$. If the HN(COCA)H spectrum contains only one candidate (due to a lack of sensitivity), they should be left unassigned.

So far, these pseudoresidues, involving four frequencies, do not overlap and thus cannot be sequentially correlated. For this purpose, they are extended towards residue (i) using the HNCA experiment, which supplies the $C^{\alpha}_i$ frequency, and optionally the HN(CA)H and HN-(CA)CO experiments ($CO_i$ and $H^{\alpha}_i$). The $C^{\beta}_i$ and $C^{\beta}_{i-1}$ frequencies could possibly be connected in the same way. In the experiments involving $H^{\alpha}$, the transfer delay has been tuned for CH groups, so that frequently the information is not available for the glycines; it could be obtained from an extra experiment with a shorter transfer interval, but usually the $C^{\alpha}$ resonances of glycines are resolved and such an experiment is not necessary.

*(iii) Sequential assignment*

Now that the extended pseudoresidues contain all the (i) and (i–1) frequencies available, for each pseudoresidue (i), a list of all possible neighbours (j) is constructed on the basis that the $C^{\alpha}_i$ chemical shift should be close enough to the $C^{\alpha}_{j-1}$ shift (typically ± 0.3 ppm). Note that all pseudoresidues are retained as possible choices when the $C^{\alpha}_i$ information is lacking. When the chemical shifts of the $H^{\alpha}$ and CO are available for (i) and (j–1), agreement is required.

In the next step, the protein sequence is entered and the NH-terminal and proline residues are discarded. If there are less pseudoresidues than expected in the sequence, a corresponding pool of 'dummy' residues is created. As for any sequential assignment method, information on the spin system is valuable as it can significantly reduce the number of combinations to be tested. This information can be provided either by supplementary experiments ($^1$H-TOCSY, HCCH-TOCSY (Bax et al., 1990) or HCCONH-TOCSY (Grzesiek et al., 1993; Lyons and Montelione, 1993)), by chemical shift ranges or by a priori knowledge (highly conserved residues in mutant studies). In the present case, only very loose chemical shift ranges for the $C^{\alpha}_i$ and $C^{\alpha}_{i-1}$ were used to rule out some possibilities for the (i) and (i–1) amino acid types; for each residue the list of possible locations in the sequence is created on this basis. The few spin systems with characteristic shifts, i.e. glycines, alanines, isoleucines, valines and threonines, provide enough information to anchor the stretches of residues into the sequence.

The prolines correspond to gaps in the sequence around which no sequential connectivity is expected. Possibly, a constraint on the $H^{\alpha}$ and $C^{\alpha}$ values of the residues preceding prolines could be added (Olejniczak and Fesik, 1994). By now we are just limited by the chemical shift ranges, nevertheless these are quite interesting for the following residues, as their $C^{\alpha}_{i-1}$ value must fit the well-resolved downfield ranges of the proline.

The program then looks for sequences that are in agreement with all the information. Two alternative approaches were implemented, one based on a systematic search and the other on an optimization procedure.

*The systematic search* This search proceeds sequentially towards the C-terminus. At each step, the program looks for a pseudoresidue which fulfills all the following conditions: (i) it has not been placed in the sequence yet; (ii) it matches the spin system required; and (iii) it belongs to the list of possible following pseudoresidues for the preceding one, unless the residue corresponds to the N-terminus or follows a proline. (Dummy residues are forward- and backward-connected to any residue.)

If no sequence is found, this may be an indication that some data (most likely HNCA triplets) are erroneous. In such a case, the program adds a new dummy residue to the pool which will be used instead of the incorrect one, and restarts from the N-terminus. This procedure is repeated as long as no sequence has been found. The output of the program consists of all assignment proposals consistent with the data and with a minimal number of dummy residues. If there are more than one, the common parts are displayed and the discrepancies listed, as well as the discarded residues if any exist.

*The simulated annealing approach* Starting from the amino acid sequence, which is filled with the pseudoresidues either randomly (the agreement for the $C^{\alpha}$ chemical shift ranges is checked) or interactively, ALPS performs a permutation between segments of residues. The beginning of the segment and its possible new location (among those in agreement with the spin system) are randomly chosen. From its starting point, the segment is stepwise extended if sequential connectivities have already been established, but this extension is performed only towards the C-terminus and not systematically, or breaking those connectivities would never be feasible. This is a significant difference with respect to many other approaches, and is essential for global convergence. The program computes an energy corresponding to the viol-

ations of the spin system constraints and the data from the HNCA and optionally the HN(CA)CO and HN(CA)-H experiments, then accepts or rejects the permutation following a simulated annealing minimization scheme (Metropolis et al., 1953). Here the program provides one output sequence, and several runs of the optimization can be compared to evaluate the precision of the result.

We have implemented an analogous but more general simulated annealing-based sequential assignment routine in the FELIX software (available in the July 1994 FELIX-Assign release). The complete ALPS program is also available upon request from the authors.

## Results and Discussion

*Extraction of triplets*

Extracting triplets of related frequencies from the 2D triple resonance experiments for cytochrome $c_2$ was achieved nearly automatically; only a few cases had to be checked manually (about four in each experiment): either close overlap of two peaks yielded several basically equivalent triplets, or additional triplets were built from spurious peaks. Selection of multiple triplets corresponding to the same HSQC peaks was based on best alignment and intensity criteria. After this procedure, the following numbers of triplets were found in the various experiments: 111 in the HNCO and HN(CO)CA, 110 in the HNCA and H(N)COCA, 69 in the HN(CA)CO, 88 in the HN(CA)H and 75 in the HN(COCA)H. In this 116-residue cytochrome, containing four proline and an N-terminal residue, 111 triplets were expected in each experiment.

*Pseudoresidues*

The combination of the HNCO, HN(CO)CA and H(N)COCA experiments yielded 110 pseudoresidues. No incorrect pseudoresidue was generated. Spurious peaks in the HNCO and HN(CO)CA spectra were associated with an NH side-chain resonance present in the HSQC spectrum. In this case, the resulting triplet could be eliminated automatically, because it was not confirmed in the third experiment. One pseudoresidue is thus missing, which was later identified as Lys[32]. In fact, Lys[32] amide $^1H$ and $^{15}N$ resonances are missing in all experiments carried out on *Rb. capsulatus* cytochrome $c_2$. One $C_i^\alpha$ shift from the HNCA experiment was associated with each pseudoresidue; $H_{i-1}^\alpha$ chemical shifts are available for 75 of them, $CO_i$ shifts for 69 of them and $H_i^\alpha$ shifts for 88 of them.

*Spin system information*

The $C^\alpha$ chemical shift ranges (used for $C_i^\alpha$ and $C_{i-1}^\alpha$) limited the number of possible locations in the sequence for all residues to approximately one third. In particular, only 10 residues out of the whole pool were left as possible candidates for the four residues following a proline. It should be noted that for cytochromes the carbon shifts are little influenced by the heme group ring currents; therefore, they fit into the usual ranges (secondary shifts have even been used to detect secondary structure elements (Caffrey et al., 1994)), and the dispersion is not better than in other proteins.

*Sequential assignment*

Using the whole set of recorded experiments and either a systematic search or the simulated annealing algorithm, ALPS yielded one unique and correct solution for the sequential assignment. As no satisfactory candidate for residue 32 could be found, the program used a dummy residue to replace it.

For the protein under investigation, the entire process required about three nights and days of spectrometer time, one day of processing and peak picking, and one or two days to obtain and verify all triplet lists. From then on the result was obtained with no more human intervention in a few minutes of calculation.

Once the program was proved to run properly, we investigated its robustness by reducing or altering the information provided. The following features were tested: (i) reduction of additional information of the spin systems; (ii) incorrect connectivity with the $C_i^\alpha$; and (iii) missing pseudoresidues.

*(i) Reduction of additional information* Table I shows the evolution of the number of potential neighbours per residue when the set of available experiments is reduced. If either the $H_i^\alpha$ or $CO_i$ chemical shifts (from respectively the HN(CA)H and HN(CA)CO experiments) are missing, the program still unambiguously finds the correct solution. It is worth noting that the $CO_i$ information is slightly more efficient than the $H_i^\alpha$ data in reducing the number of connectivities. This phenomenon arises presumably from two factors: first, a larger amount of information available on the CO resonances (all $CO_{i-1}$ shifts are obtained compared to only 75 $H_{i-1}^\alpha$ shifts) and secondly, the strong correlation of the $C^\alpha$ and $H^\alpha$ shifts ($C^\alpha$ intricacy is more likely to be settled by CO shifts than by $H^\alpha$ shifts).

TABLE I
EVOLUTION OF THE SYSTEMATIC SEARCH PROGRAM AND AMOUNT OF INFORMATION USED

| Data used[a] | $H_i^\alpha + CO_i$ | $CO_i$ | $H_i^\alpha$ | None |
|---|---|---|---|---|
| Average number of neighbours per residue | 1.37 | 1.74 | 1.94 | 3.90 |
| Number of tests in the arborescence | 23 933 | 128 064 | 783 765 | [b] |

[a] ($N \rightarrow C^\alpha$) experiments used in addition to HNCA: HN(CA)H for $H_i^\alpha$, HN(CA)CO for $CO_i$.
[b] The test lasted several days and was not completed.

TABLE 2
EVOLUTION OF SIMULATED ANNEALING WHEN MISSING RESIDUES ARE REPLACED BY DUMMY RESIDUES[a]

| No. of missing residues[b] | Unambiguously assigned residues[c] | No. of distinct generated sequences | CPU time (min:s)[d] |
|---|---|---|---|
| 0 | 111 (100%) | 1 | 1:25 |
| 1 | 111 (100%) | 1 | 1:47 |
| 2 | 111 (100%) | 1 | 2:07 |
| 5 | 108 (97%) | 3 | 5:30 |
| 10 | 105 (95%) | 8 | 15:15 |
| 15 | 79 (71%) | 10 | 19:00 |

[a] Average of 10 program runs.
[b] Number of residues that were randomly removed in addition to $Lys^{32}$.
[c] Including dummy residues.
[d] CPU time for one run on an SGI 4D-35 computer. The number of iterations and steps per iteration were adjusted in order to reach the global minimum, when a larger configurational space had to be explored.

If neither $H_i^\alpha$ nor $CO_i$ shifts are available, the program fails to determine the correct assignment. With an average of almost four possible neighbors per residue, the calculation time becomes inhibitory for the systematic search; for the simulated annealing, the energetic space is poorly defined, leading to a large number of significantly different solutions with no constraint violation. However, the loss of connectivity restraints could be compensated by additional spin system information. This information was gained from other data (3D (H)CCH-TOCSY experiments, homonuclear and $^{15}N$-edited TOCSY) already available in our laboratory. Tests were made by providing spin system assignment for a randomly chosen set of pseudoresidues: when about 30% of the spin system assignments were provided, the correct sequence was definitely found. In fact, a smaller number of crucial residues may be sufficient.

*(ii) Incorrect connectivity with $C_i^\alpha$* If the spectra are crowded and the sensitivity poor, a few triplets might be incorrect, especially in the $N \rightarrow C^\alpha$ experiments (a doubled number of peaks and lower S/N ratio). To evaluate the sensitivity of the program to such problems, we artificially modified the data and introduced one or two errors in the HNCA triplets; upon inspection of the whole arborescence the systematic program did not find any solution and it had to add one dummy residue per incorrect connectivity. In the case of accidental overlaps, some permutations of residues yielded additional, slightly different sequences. Inspection of the shifts of the remaining residues allowed identification of the error and replacement in a correct sequence. Similarly, the simulated annealing approach yielded the 'correct' sequence, but this involved a violation of the erroneous connectivity constraint, which was detected in the final energy. However, when the number of errors becomes too large, many alternative sequences (with very similar energies) might be found, and the user is advised to fix the mistakes in the data set.

*(iii) Missing pseudoresidues* Another conceivable stumbling block is the absence of correlations for some residues, due for instance to shorter $T_2$ relaxation of some spins. Then an identical number of dummy pseudoresidues has to be added to the pool and each one may fit anywhere in the sequence. Moreover, such a dummy residue does not carry any information on the nature of its neighbours and can thus be anchored to any pseudoresidue. Consequently, the duration of the systematic search exponentially increases (from less than one second to days), as it is not possible to stop the exploration of erroneous paths before their ultimate end. Our practical tests on cytochrome $c_2$ showed that, in the absence of additional information, the systematic search fails when more than two residues are missing.

The simulated annealing method was tested as an alternative which is more tolerant with a larger number of missing residues. Simulations were carried out by progressively spoiling the data set. As shown in Table 2, the program yielded a high rate of correct assignment when up to 10 or 15% of the residues were missing. Then several runs of simulated annealing yielded distinct sequences, but these had significant common parts whose assignment was thus reliable. The program was also able to locate where in the sequence dummy pseudoresidues were needed, employing the consistency of the overall information. Generally, this algorithm seems to be less sensitive to the absence of a few localized signals than to the suppression of a 2D experiment.

## Conclusions

The method proved efficient in assigning the backbone resonances of cytochrome $c_2$ using very short experimental and computational times. The little human intervention that is required is mostly devoted to checking the triplets extracted from 2D experiments of reduced dimensionality. The first, systematic method yields directly all possible solutions, but is limited by computation time. The simulated annealing approach provides a more powerful alternative when the data set is incomplete. We believe that this method will show its optimal efficiency on labelled proteins of intermediate size ($\approx$ 100–120 residues), as

both the acquisition and interpretation time will be minimized. Furthermore, because the set of experiments that the program can use is very variable, we hope that ALPS will solve cases of larger proteins as well.

## Acknowledgements

We wish to thank Dr. M. Caffrey for helpful discussions on cytochrome $c_2$, and Prof. M. Cusanovich (University of Arizona) for his contribution in the preparation of the labelled sample. This work was supported by the Centre National de la Recherche Scientifique, the Commissariat à l'Énergie Atomique, Biosym Technologies Inc., San Diego and the Bruker company (Wissembourg). N.M. is the recipient of a CEA-Biosym fellowship and B.B. of a CEA-Bruker fellowship.

## References

Bax, A., Clore, G.M. and Gronenborn, A.M. (1990) *J. Magn. Reson.*, **88**, 425–431.

Bernstein, R., Cieslar, C., Ross, A., Oschkinat, H., Freund, J. and Holak, T.A. (1993) *J. Biomol. NMR*, **3**, 245–251.

Brutscher, B., Simorre, J.-P., Caffrey, M.S. and Marion, D. (1994) *J. Magn. Reson. Ser. B*, **105**, 77–82.

Caffrey, M.S., Brutscher, B., Simorre, J.-P., Fitch, J., Cusanovich, M.A. and Marion, D. (1994) *Eur. J. Biochem.*, **221**, 63–75.

Catasti, P., Carrara, E. and Nicolini, C. (1989) *J. Comput. Chem.*, **11**, 805 818.

Cieslar, C., Clore, G.M. and Gronenborn, A.M. (1988) *J. Magn. Reson.*, **80**, 119–127.

Cieslar, C., Holak, T.A. and Oschkinat, H. (1990) *J. Magn. Reson.*, **87**, 400–407.

Clore, G.M. and Gronenborn, A.M. (1991) *Prog. NMR Spectrosc.*, **32**, 43–92.

Clubb, R.T., Thanabal, V. and Wagner, G. (1992) *J. Magn. Reson.*, **97**, 213–217.

Eads, C.D. and Kuntz, I.D. (1989) *J. Magn. Reson.*, **82**, 467–482.

Eccles, C., Güntert, P., Billeter, M. and Wüthrich, K. (1991) *J. Biomol. NMR*, **1**, 111 130.

Gooley, P.R., Caffrey, M.S., Cusanovich, M.A. and MacKenzie, N.E. (1990) *Biochemistry*, **29**, 2278–2290.

Grzesiek, S. and Bax, A. (1992) *J. Magn. Reson.*, **96**, 432–440.

Grzesiek, S., Anglister, J. and Bax, A. (1993) *J. Magn. Reson. Ser. B*, **101**, 114–119.

Ikura, M., Kay, L.E. and Bax, A. (1990) *Biochemistry*, **29**, 4659–4667.

Kay, L.E., Ikura, M., Tschudin, R. and Bax, A. (1990) *J. Magn. Reson.*, **89**, 496–514.

Kleywegt, G.J., Boelens, R., Cox, M., Llinas, M. and Kaptein, R. (1991) *J. Biomol. NMR*, **1**, 23–47.

Kleywegt, G.J., Vuister, G.W., Padilla, A., Knegtel, R.M.A., Boelens, R. and Kaptein, R. (1993) *J. Magn. Reson. Ser. B*, **102**, 166–176.

Lyons, B. and Montelione, G.T. (1993) *J. Magn. Reson. Ser. B*, **101**, 206–209.

Meadows, R.P., Olejniczak, E.T. and Fesik, S.W. (1994) *J. Biomol. NMR*, **4**, 79–96.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E. (1953) *J. Chem. Phys.*, **21**, 1087–1092.

Olejniczak, E.T. and Fesik, S.W. (1994) *J. Am. Chem. Soc.*, **116**, 2215–2216.

Olson Jr., J.B. and Markley, J.L. (1994) *J. Biomol. NMR*, **4**, 385–410.

Richarz, R. and Wüthrich, K. (1978) *Biopolymers*, **17**, 2133–2141.

Simorre, J.-P., Brutscher, B., Caffrey, M.S. and Marion, D. (1994) *J. Biomol. NMR*, **4**, 325–333.

Szypersky, T., Wider, G., Bushweller, J.H. and Wüthrich, K. (1993a) *J. Am. Chem. Soc.*, **115**, 9307–9308.

Szypersky, T., Wider, G., Bushweller, J.H. and Wüthrich, K. (1993b) *J. Biomol. NMR*, **3**, 127–132.

Van der Ven, F.J.M. (1990) *J. Magn. Reson.*, **86**, 633 644.

Van Doren, S.R., Kurochkin, A.V., Ye, Q.-Z., Johnson, L.L., Hupe, D.J. and Zuiderweg, E.R.P. (1993) *Biochemistry*, **32**, 13109–13122.

Wishart, D.S. and Sykes, B.D. (1994) *J. Biomol. NMR*, **4**, 171–180.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York, NY.

Xu, J., Strauss, S.K., Sanctuary, B.C. and Trimble, L. (1994) *J. Magn. Reson. Ser. B*, **103**, 53–58.

Zimmerman, D., Kulikowski, C., Wang, L., Lyons, B. and Montelione, G.T. (1994) *J. Biomol. NMR*, **4**, 241–256.