Journal of Structural and Functional Genomics **2:** 93–101, 2002.

CABM Symposium

Rapid analysis of protein backbone resonance assignments using cryogenic probes, a distributed Linux-based computing architecture, and an integrated set of spectral analysis tools

Daniel Monleón[1], Kimberly Colson[2], Hunter N. B. Moseley[1], Clemens Anklin[2], Robert Oswald[3], Thomas Szyperski[4] & Gaetano T. Montelione[1,*]

[1]*Center for Advanced Biotechnology, and Medicine and Department of Molecular Biology and Biochemistry, Rutgers University, 679 Hoes Lane, Piscataway, NJ 08854, USA;* [2]*Bruker Instruments Inc., 44 Manning Road, Billerica, MA 01821, USA;* [3]*Department of Molecular Medicine, Cornell University, C3 167 Veterinary Medical Center, Ithaca, NY 14853, USA;* [4]*Department of Chemistry, University of Buffalo, The State University of New York, 816 Natural Sciences Complex, Buffalo, NY 14260, USA; *Author for correspondence (fax: 732-235-5633; tel.: 732-235-5321; e-mail: guy@cabm.rutgers.edu)*

Received 31 January 2002; Accepted 2 May 2002

Key words: AutoAssign software, automatic spectral analysis, high-throughput NMR, parallel processing, peak picking and filtering, structural proteomics

Abstract

Rapid data collection, spectral referencing, processing by time domain deconvolution, peak picking and editing, and assignment of NMR spectra are necessary components of any efficient integrated system for protein NMR structure analysis. We have developed a set of software tools designated AutoProc, AutoPeak, and AutoAssign, which function together with the data processing and peak-picking programs NMRPipe and Sparky, to provide an integrated software system for rapid analysis of protein backbone resonance assignments. In this paper we demonstrate that these tools, together with high-sensitivity triple resonance NMR cryoprobes for data collection and a Linux-based computer cluster architecture, can be combined to provide nearly complete backbone resonance assignments and secondary structures (based on chemical shift data) for a 59-residue protein in less than 30 hours of data collection and processing time. In this optimum case of a small protein providing excellent spectra, extensive backbone resonance assignments could also be obtained using less than 6 hours of data collection and processing time. These results demonstrate the feasibility of high throughput triple resonance NMR for determining resonance assignments and secondary structures of small proteins, and the potential for applying NMR in large scale structural proteomics projects.

Abbreviations: BPTI – bovine pancreatic trypsin inhibitor; LP – linear prediction; FT – Fourier transform; S/N – signal-to-noise ratio; FID – free induction decay

Introduction

Resonance assignments provide the basis for analysis of protein structure and dynamics by NMR spectroscopy [1–3]. The use of multidimensional triple resonance NMR for determination of protein resonance assignments [4–7] has become standard in many laboratories. Indeed, for small proteins ($\leq$ 25 kDa) triple resonance NMR data analysis is, in many cases, a routine task. However, the processes of data collection, referencing, time domain deconvolution, organizing, and manual analysis of the spectral data still

takes a significant amount of time and effort. Over the last few years a number of laboratories have developed automatic or semiautomatic methods for NMR data analysis [7, 8–13]. Although such software greatly accelerates the process of going from high-quality peak lists to resonance assignments, the data collection, referencing, Fourier transformation, peak picking, and peak list editing processes still constitute the major portion of the time required for protein structure determination.

Recent developments in cryogenic probe technology for NMR spectroscopy [14, 15] provide significant improvements in signal-to-noise ratios for protein samples in aqueous buffered solutions, and allow much shorter data collection times for most of the triple resonance experiments designed for protein resonance assignments. For example, Cowburn and co-workers have demonstrated that all the data required for analysis of backbone resonance assignments of small proteins can be collected in just a few days using triple resonance NMR cryoprobes [16]. However, high-throughput structure analysis requires both rapid data collection and rapid and automated NMR 'data processing'. We include in the term 'data processing' the several data reduction and organizational processes involved in converting raw time domain multidimensional NMR data into properly referenced peak lists suitable for analysis with automated resonance assignment and structure generation programs.

The use of parallel and/or distributed computer systems provides an important approach to reducing the time required for 'data processing'. Anticipating these developments, some software packages for multidimensional Fourier transformation of NMR data, like NMRPipe [17], have been designed for use with distributed computing systems of many different processor architectures. While most UNIX-based parallel systems run on very expensive hardware configurations, the Linux operating system provides a very efficient and cheap platform for many scientific software packages. The low cost of Linux workstations and associated network and storage accessories required for implementing an efficient distributed system allows even modest academic laboratories to implement such configurations for high-throughput NMR data analysis.

In this study, we demonstrate the combined use of triple resonance cryoprobes, efficient data collection strategies, 'data processing' using a cluster of Linux-based processors, and our newly developed AutoProc and AutoPeak software in combination with the programs NMRPipe [17], Sparky [18], and AutoAssign [9, 13], to provide rapid determination of protein resonance assignments and secondary structure. These results provide a basis for considering high throughput resonance assignments and secondary structure analysis as a component of an integrated program in structural proteomics [19].

Methods and materials

Uniformly $^{13}C,^{15}N$-enriched BPTI was prepared as described elsewhere [20]. This construct includes an N-terminal methionine residue, resulting in a total of 59 amino acids. Triple resonance spectra for backbone resonance assignments were collected using a Bruker AVANCE NMR spectrometer system equipped with a triple resonance cryoprobe. The NMR sample was maintained at 30 ± 0.1 °C, while coil temperature was maintained at 30 K. Samples were prepared for NMR measurements in 5-mm diameter Shigemi NMR tubes using a sample volume of 190 μl. The protein concentration was 0.9 mmol/L at pH 5.8 in aqueous solvent containing 5% D_2O, 20 mM sodium acetate, 25 mmol/L $CaCl_2$, and 0.02% NaN_3.

Results and discussion

Data collection

The combination of cryoprobe and automated analysis technologies can provide a robust approach for rapid protein structure analysis by NMR. We demonstrate this here using the canonical protein NMR test sample, bovine pancreatic trypsin inhibitor (BPTI). Although BPTI provides particularly nice NMR spectra, similar quality data have been obtained using the cryoprobe on other protein samples. Data collection was carried out at 500 MHz in two different runs. The key parameters for these data collection runs are summarized in Table 1. For data collection I, six 3D triple-resonance spectra were collected over 36 h. For data collection II, the same six 3D spectra were collected over 5 h. The primary aim of data collection II was to explore the feasibility of recording the required NMR spectra in an unprecedented short period of time, i.e. in less than six hours. To accomplish this goal, we (i) recorded four of the six 3D data sets with

Table 1. Triple resonance data sets collected to explore various high-throughput resonance assignment strategies.

	^{15}N-HSQC	HNcoCA	HNCO	HNCA	HNCACB	HNcoCACB	HNcaCO
Data Collection I							
No. of points							
Collected	1024, 128	1024, 20, 40	1024, 20, 40	1024, 20, 40	1024, 20, 40	1024, 20, 40	1024, 20, 40
After LP	1024, 256	1024, 40, 80	1024, 40, 80	1024, 40, 80	1024, 40, 80	1024, 40, 80	1024, 40, 80
After zero filling	1024, 256	1024, 64, 128	1024, 64, 128	1024, 64, 128	1024, 64, 128	1024, 64, 128	1024, 64, 128
No. of scans	1	2	2	2	8	8	8
Spectral width							
(ω_1, ω_2,	6720, 1613	6720, 1613,	6720, 1613,	6720, 1613,	6720, 1613,	6720, 1613,	6720, 1613,
ω_3; Hz)		4032	2777	4032	9433	9433	2777
Recycle delay (s)[a]	1.15	1.05	1.05	1.05	1.15	1.15	1.05
Collection time (h)	0.1	2.4	2.4	2.4	10	10	9.3
Data Collection II							
No. of points							
Collected	1024, 128	1024, 20, 26	1024, 20, 20	1024, 20, 26	1024, 20, 60	1024, 20, 60	1024, 20, 20
After LP	1024, 256	1024, 40, 52	1024, 40, 40	1024, 40, 52	1024, 40, 120	1024, 40, 120	1024, 40, 40
After zero filling	1024, 256	1024, 64, 64	1024, 64, 64	1024, 64, 64	1024, 64, 128	1024, 64, 128	1024, 64, 64
No. of scans	1	1	1	1	1	1	4
Spectral width							
(ω_1, ω_2,	6720, 1613	6720, 1613,	6720, 1613,	6720, 1613,	6720, 1613,	6720, 1613,	6720, 1613,
ω_3; Hz)		4032	2777	4032	9433	9433	2777
Recycle delay (s)[a]	1.15	0.75	0.75	0.75	0.75	0.75	0.75
Collectiont time (h)	0.1	0.5	0.4	0.5	1.0	1.3	1.7

[a] Recycle day = acquisition time plus interincrement delay.

only a single scan per acquired free induction decay (FID), (ii) carefully optimized the sweep widths and maximal evolution times in the indirect dimensions, and (iii) employed unusually short relaxation delays between scans, trading optimum signal-to-noise ratios (S/N) for shorter data collection time requirements.

Cryoprobe performance

In order to document the performance of the Bruker triple-resonance cryoprobe used in this work on an aqueous biological sample, S/N ratios were measured and compared with those obtained on the same 500-MHz spectrometer system using a conventional triple-resonance probe. Using a standard 0.1% ethylbenzene sample, the cryoprobe provided a S/N (200-Hz noise region) of 3415 : 1, while the conventional probe provided a S/N of 970 : 1. Using the ^{15}N,^{13}C -enriched BPTI sample described in the Methods and materials section, 3D HNcoCACB spectra were run using identical collection parameters on the cryoprobe and conventional probes. A comparison of the HN-C projections of these two HNcoCACB spectra is shown in Figure 1, together with a representative trace through the spectrum. Analysis of cross peaks throughout these spectra demonstrate a range of S/N enhancements of 2.5 to 3.5, with an average value of 3.3. Accordingly, the triple-resonance cryoprobe measurements can be carried out on aqueous protein solutions at reasonable ionic strength using approximately 10-fold shorter data collection periods than the corresponding experiments collected using a conventional triple-resonance probe.

Automated referencing and spectral processing

Prior to Fourier transformation of the time domain data, the NMR spectra were automatically referenced using AutoProc. Details of the algorthithms of AutoProc and their implementations will be presented elsewhere. The AutoProc software also generates NMRPipe [17] processing scripts for multidimensional Fourier transformation based on user-specified phasing, window functions and baseline correction.

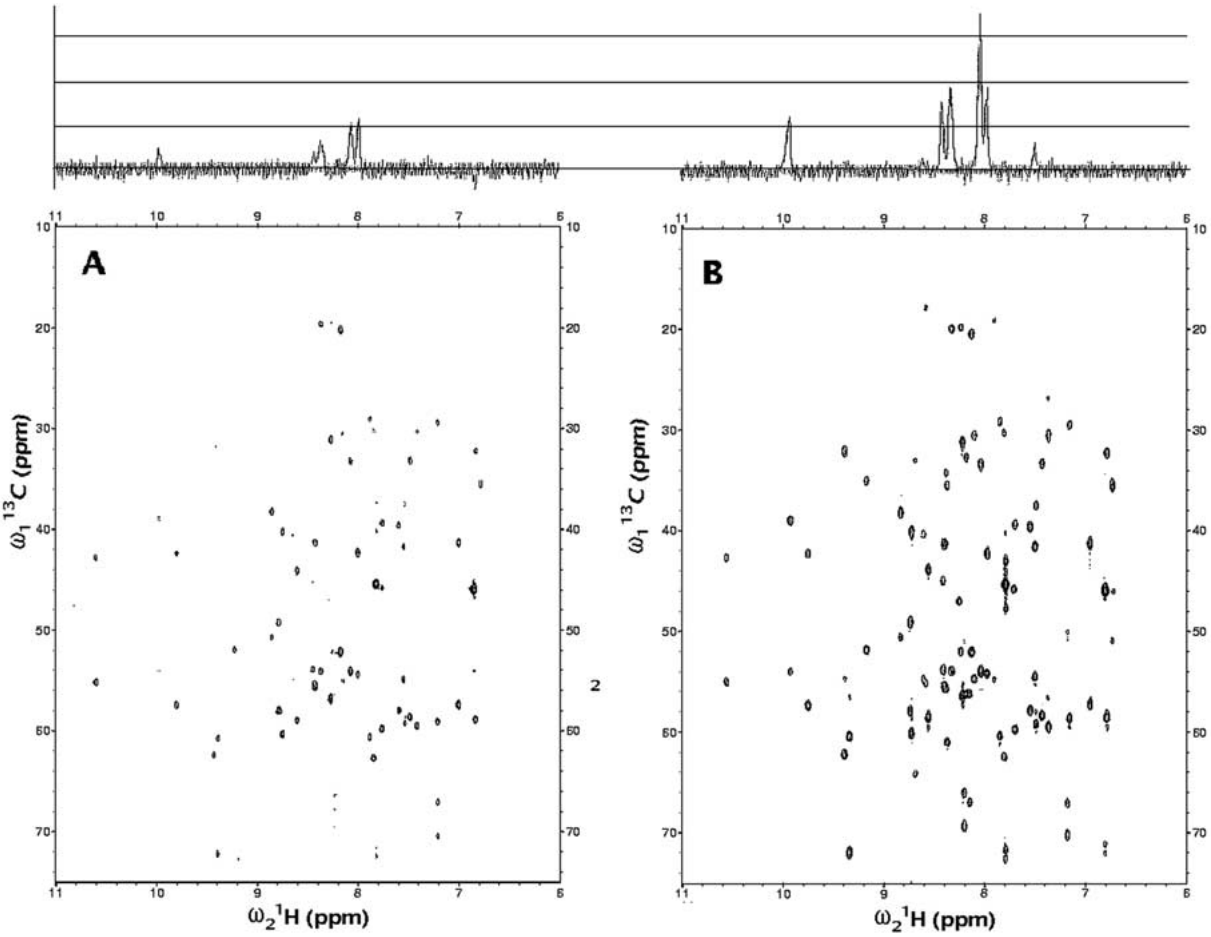

Figure 1. Comparison of skyline projections of HNcoCACB spectra recorded for 0.9 mmol/L ^{15}N,^{13}C-enriched BPTI in solutions containing 20 mmol/L NaAc, 25 mmol/L $CaCl_2$, and .02% NaN_3 at pH 5.8 and 30 °C. These 3D spectra were each collected at 500 MHz over a 1.3-h period (data collection parameters listed in Table 1, data collection II) using either (**A**) a conventional triple-resonance probe or (**B**) a triple-resonance cryoprobe. Cross-sections through these projections demonstrate a S/N enhancement of ~ 3 for the cryoprobe compared with the conventional probe.

These NMRPipe scripts are optimized for parallel processing on a Linux architecture. In this particular study, all data were linear predicted, zero filled, and Fourier transformed using parallel processing methods on our Linux-based cluster of 44 Pentium III CPUs. The total processing time for all six triple-resonance spectra described in Table 1 was about 1 h. This corresponds to an average of about 10 min for each 3D spectrum (including linear prediction). Details of hardware and software configurations of our Pentium III PC cluster for parallel NMR processing are available upon request (guy@cabm.rutgers.edu).

Digital solvent signal suppression [21], linear prediction [22, 23] to double the number of points in all the indirect dimensions, shifted (0.4π) sine-bell window functions, and zero filling were applied to these data. The use of linear prediction greatly improves the resolution in overlapped peaks, allowing the peak picking and filtering methods described below to better identify peak frequencies. Polynomial baseline corrections of 4th order were also applied in all the indirect dimensions after Fourier transformation. The general use of baseline correction improves the quality of the 2D projections required for the peak filtering approach described below. Once all of the data were Fourier transformed, several 2D projections of these 3D spectra were generated and used to evaluate their overall quality. These 2D projections served as filters for the peak editing software, AutoPeak.

Automated peak list editing

Peak-picking and editing were carried out in an automated manner using Sparky software [18] for peak-picking and AutoPeak software for peak list editing. In this study, we first performed an initial peak picking at low intensity level (0.9 times the heuristic white noise level) over each of the 3D spectra using the $^{15}N–H^N$ *restricted peak picking* feature of the program Sparky [18]. This process provides a peak list containing only peaks of the multidimensional spectra that match (e.g. within tolerances of $\delta_N = \pm 0.2$ ppm, $\delta_{HN} = \pm 0.02$ ppm) in their ^{15}N and H^N resonance frequencies to some peak in a high-resolution ^{15}N–HSQC spectrum that has been carefully manually picked. We refer to these $^{15}N–H^N$ peak frequencies as the 'N–H root peak list' [9].

In the second step, we further filter this 'N–H filtered' 3D peak list using a 'C–H root peak list', generated by a similar manual interactive analysis of appropriate 2D $^{13}C–H^N$ projections of 3D HNCACB and/or 3D HNcoCACB spectra. In principle, this 'C–H root peak list' contains the resonances of every peak in the projections of these 3D HNCACB and 3D HNcoCACB spectra. This approach was sufficiently robust to handle most cases of peak overlap in these projections. After this filtering process, each peak in the final peak list matches ^{15}N and H^N peak frequencies derived from the $^{15}N–H^N$ projection and ^{13}C peak frequencies derived from the $^{13}C^{\alpha/\beta}$-H^N projections. The probability of an artifact peak in the 3D spectra passing this filtering process is very low. On the other hand, the otherwise significant time required for editing the 3D peak lists is greatly reduced because only two or three 2D spectra (or 2D projections of 3D spectra) require manual editing of automatically generated peak lists. The manual editing of these 2D spectra can be done in tens of minutes. Of course, the efficiency and reliability of this peak filtering approach depends strongly on the quality of the spectra and on the quality of the resulting 'root peak lists' used as filters.

In order to generate better root peak lists, we explored different approaches for generating the required 2D projections of 3D HNCACB and 3D HNcoCACB spectra. In particular, we evaluated both *skyline* and *sum* 2D projections of these 3D spectra. The *skyline projection* is generated by computing the profile of 3D peak intensities onto a selected 2D plane; the *sum projection* is generated by computing a projection of the sum of peak intensities onto a 2D plane. The differences between these methods was limited to only a few peaks in some spectra. However, AutoAssign generally provided better results for the peak lists filtered using the *sum projection* as the basis for the 'root peak list'.

As one might expect, accurate and consistent referencing of spectra is critical for the performance of these peak editing methods. A relative shift between spectra provides almost empty filtered peak lists for normal filtering tolerances, while inaccuracy in the absolute referencing of the whole set of spectra can result in amino acid typing errors in AutoAssign. For this reason, accurate and consistent spectral referencing provided by the AutoProc process is critical for proper performance of the AutoPeak editing process. In addition, AutoPeak utilizes algorithms for inter-spectral registration of peak lists, applying small global adjustments to the shift lists to optimally align them with one another, prior to peak picking and peak list editing.

Strategies of triple resonance data collection

We also explored several different strategies of data collection and processing in order to evaluate approaches for reducing data collection time requirements. AutoPeak provides automatic generation of peak lists for specific triple-resonance experiments required for backbone assignments from the encoded information of other experiments. This allows for data collection strategies using fewer than the seven data sets summarized in Table 1. For example, it is relatively easy to extract the peaks corresponding to a 3D HNCA spectrum from a 3D HNCACB spectrum collected with delays tuned to provide different signs (i.e., positive or negative intensities) for C^α and C^β resonances (i.e. C^α/C^β phase-discriminated experiments; [24, 25]). The 3D HNCACB and 3D HNcoCACB spectra used in this study were both collected with such phase discrimination of C^α and C^β cross peaks. The combinations of strategies tested using various subsets of the six triple-resonance experiments of Table 1 are summarized in Table 2. Representative sequential connectivity maps generated from these automated analyses are shown in Figure 2. A complete summary of results for various analysis strategies used with both data collection strategies is presented in Table 3.

In all the data analysis strategies that we explored, assignments of atoms in residues around Cys-14 and Cys-38 were incomplete. This is attributable to broad-

Table 2. Different strategies used for generating the peak lists required for AutoAssign analysis.

Data analysis strategy	A	B	C	D
HSQC	HSQC	HSQC	HSQC	HSQC
HNCO	HNCO	HNCO	HNCO	HNCO
HNCA	a	a	HNCA	HNCA
HNcoCA	b	c	HNcoCA	HNcoCA
HNCACB	HNCACB	HNCACB	HNCACB	HNCACB
HNcoCACB	HNcoCACB	HNcoCACB	HNcoCACB	HNcoCACB
HNcaCO			HNcaCO	
Collection I (h)	22.5	22.5	36.6	27.3
Collection II (h)	2.8	2.8	5.5	3.8
LP/FT (h)	1	1	1	1
Peak picking and editing (h)	1	1	1	1
Total Time I (h)	24.5	24.5	38.6	29.3
Total Time II (h)	4.8	4.8	7.5	5.8

[a] HNCA generated from a C^{α}/C^{β} phase-discriminated HNCACB spectrum.
[b] HNcoCA data is generated from a C^{α}/C^{β} phase-discriminated HNCACB spectrum as described above in footnote a together with HNcoCACB data. This strategy does not require recording a HNcoCACB spectrum with C^{α}/C^{β} phase-discrimination.
[c] HNcoCA generated from a C^{α}/C^{β} phase-discriminated HNcoCACB spectrum.

Table 3. Numbers of backbone spin systems assigned[a] and numbers of spin system assignment errors[b] using the AutoProc/NMRPipe/Sparky/AutoPeak/AutoAssign process with various strategies and data sets

Strategy	A	B	C	C + G37	D	D + G37
Data collection I						
Total time (h)	24.5	24.5	38.6	38.6	29.5	29.5
Sum projection	43(4)	39(1)	50(0)	53(0)	50(0)	53(0)
Skyline projection	38(3)	36(0)	48(0)	50(0)	48(0)	50(0)
Data collection II						
Total time (h)	4.8	4.8	7.5	7.5	5.8	5.8
Sum projection	36(4)	38(5)	45(1)	46(0)	45(1)	46(0)
Skyline projection	40(3)	36(1)	46(3)	50(0)	46(3)	50(0)

[a] A backbone spin system is counted as 'assigned' if an assignment is reported for the $^{13}C^{\alpha}$ resonance.
[b] Number of spin system assignment errors reported in parentheses. Deviations of C^{α} or ^{15}N chemical shift larger than 0.2 ppm and deviations of H^{N} chemical shift larger than 0.02 ppm with respect the assignments reported in BMRB (entry BMRB-4968) are considered errors in the overall spin system assignment.

ening of many resonances belonging to amino acids in these polypeptide segments resulting from dynamic conformational exchange between low-energy conformers of the Cys-14–Cys-38 disulfide bond [26, 27], and the high density of proline residues in polypeptide segment Pro-8 to Pro-13. Assignments in this region of the protein are difficult to make even when using fully manual analysis methods.

Water suppression

Although the quality of these spectra collected using the Bruker triple resonance cryoprobe was quite good, we did not obtain excellent water suppression in most of the data sets from Data Collection II involving single scan experiments. Because of this, we were unable to observe cross peaks associated with residue glycine-37 [$\delta(H^{N}) = 4.36$ ppm]. In order to evaluate the performance of our methods on spectra collected with better water suppression, we compared the performance of automated analysis with and without manual insertion of the peaks corresponding to Gly-37 triple resonance correlations that were missing in the original data. The insertion of these additional peaks in the input peak lists provided one to six more backbone spin system assignments than the

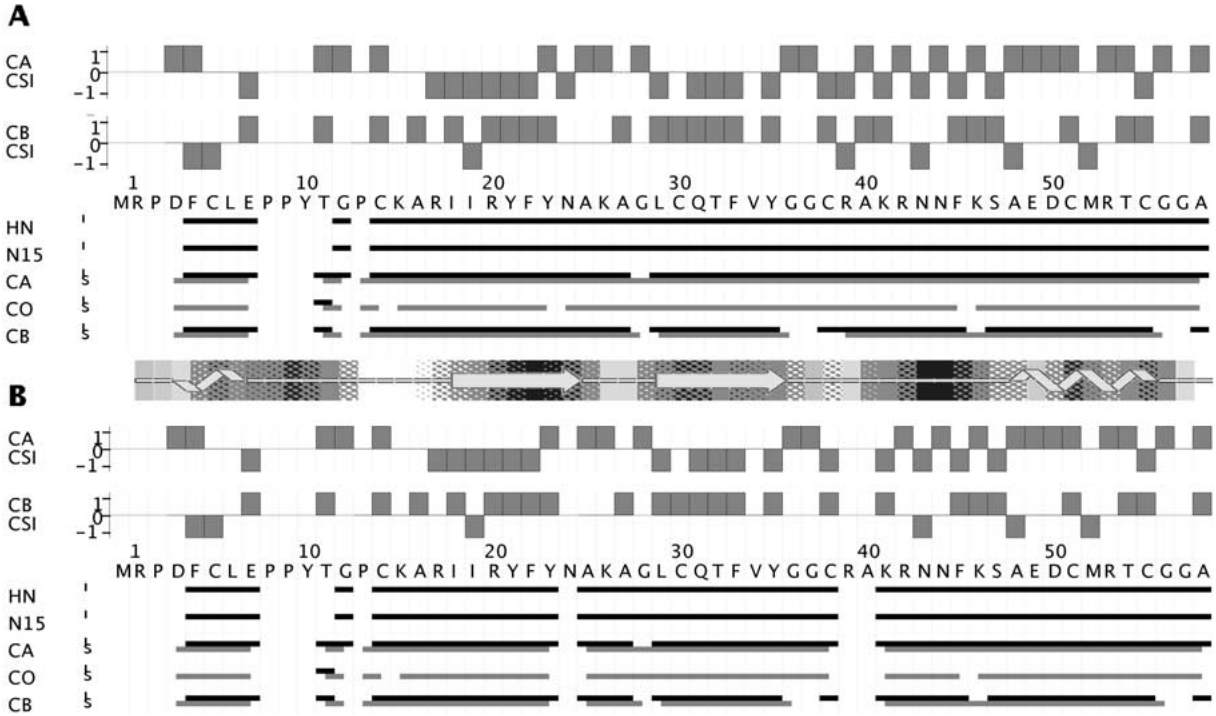

Figure 2. (**A**) Triple-resonance connectivity map and Wishart plot for the best results obtained with the AutoProc/NMRPipe/Sparky/AutoPeak/AutoAssign process using data collection I parameters and Strategy D+G37 with sum projection to generate 'C–H root peak list'. The total time required for data collection and processing was ~30 h. Also shown is a comparison with the secondary structure calculated from the 3D structure of BPTI deposited in the Protein Data Bank (PDB code 5pti). (**B**) AutoAssign triple-resonance connectivity map and Wishart plot for the best results using data collection II parameters and Strategy D+G37 with *skyline* projection to generate 'C–H root peak list'. The total time required for data collection and processing was ~6 h.

original data (compare results of strategies C and C+G37, and D and D+G37 in Table 3).

Comparison of rapid data collection and processing strategies for automated analysis of backbone resonance assignments with AutoAssign

As shown in Table 3, some of the simpler data collection strategies that we evaluated provided reasonably extensive backbone assignments with relatively short (3–23 h) data collection periods. Successful AutoAssign analysis of the most minimal of these subsets of experiments (i.e. 2D HSQC, 3D HNCO, 3D HNCACB and 3D HNcoCACB) requires exceptionally complete and high-quality triple-resonance data. However, for this BPTI sample, even this very compact data collection strategy was reasonably successful. For example, using data collection strategies A and B, extensive resonance assignments with relatively few errors were obtained automatically in about 5 h (Table 3). Previous efforts in automated analysis of resonance assignments using similar data for BPTI required several days for Fourier transformation, peak picking and peak editing in order to provide the input files required for AutoAssign analysis, and resulted in backbone spin system assignments for only 46 amino acids [13].

Strategies A and B of Tables 2 and 3 are approaches in which the C^{α} and C^{β} resonances are distinguished on the basis of C^{α}/C^{β} sign information (i.e. positive or negative peak intensities) in the HNCACB and HNcoCACB spectra, providing for shorter overall data collection times. In both Strategies A and B, HNCA-type data, required by AutoAssign in order to identify C^{α} resonances, are generated from a C^{α}/C^{β} phase-discriminated HNCACB spectrum. In strategy A, the HNcoCA input file, used by AutoAssign to identify sequential $C_i^{\alpha} \rightarrow H_{i+1}^{N}$ connections, is generated with AutoPeak by comparing the HNCA data with HNcoCACB data recorded without C^{α}/C^{β} phase discrimination, thus identifying the sequential (HNcoCA-type) peaks in the HNCA data set. In strategy B, the required HNcoCA data are generated with AutoPeak by simply removing C^{β} cor-

relations (i.e. negative peaks) from a C^{α}/C^{β} phase-discriminated HNcoCACB spectrum.

In comparing these data collection/analysis strategies (Table 3), Strategy A provides more backbone spin system assignments but also more spin system assignment errors than strategy B. This behaviour can be explained by the fact that, in strategy A, both HNCA and HNcoCA peak lists are generated from a single common spectrum (i.e. by appropiate editing of the 3D HNCACB data), providing higher consistency for interspectrum matching and therefore, in this case, more resonance assignments. Strategy B, on the other hand, uses both the C^{α}/C^{β} phase-discriminated 3D HNCACB and 3D HNcoCACB spectra to generate the required 3D HNCA and 3D HNcoCA peak lists, making use of more actual experimental information and providing more accurate results. The most reliable strategies are the ones that use actual data from HNCA and HNcoCA spectra (i.e. strategies C and D). In our view, strategies C and D are preferable, as collecting the actual 3D HNCA and 3D HNcoCA spectra, which are relatively high-sensitivity experiments, does not require a large increase in the total data collection time. In addition, using strategies C and D ensures optimum sensitivity for detecting C^{α} resonances (using HNCA and HNcoCA spectra) and allows the HNCACB and HNcoCACB data collection parameters to be tuned to optimize C^{β} resonance intensities.

For this particular sample, the establishment of intraresidue C^{α}–C' connectivities using 3D HNcaCO data did not provide more accurate or complete resonance assignments. This is a consequence of the relative completeness of the other spectra. As the HNcaCO is the least sensitive of the spectra listed in Table 1, it is noteworthy that at least in some cases the most time consuming of the basic spectra can be omitted without severely compromising the accuracy or completeness of the results. In fact, the only improvement observed for addition of HNcaCO peak lists was for strategy C+G37 in Data Collection II runs (Table 3), providing one additional spin system assignment for a 25% (ca. 1.5 h.) increase in data collection time.

Automated analysis of protein secondary structure

Figure 2A shows the Wishart plot [28] characterizing backbone secondary structure from resonance assignments obtained with AutoAssign using data collection I spectra and strategy D + G37. This secondary structure analysis based on C^{α} and C^{β} chemical shift data are compared with the actual secondary structure reported for a crystal structure of BPTI (Protein Data Bank, entry 5pti). These results demonstrate automated processing and analysis of NMR data collected in about 28 h and processed in about 2 h to provide extensive backbone resonance assignments and secondary structure for this small protein. Similar resonance assignment and secondary structure accuracy (shown in Figure 2B) was obtained using data collection II parameters and strategy D + G37, requiring about 4 h of data collection and about 2 h of data processing and analysis. Successful and rapid analysis was also achieved without artificially supplementing the data sets with the obscured resonances of residue Gly-37.

Conclusions and future prospects

We demonstrate in this study the automatic determination of backbone resonance assignments and secondary structure for a 59 amino acid protein in less than 2 days of data collection, processing, peak picking, peak list editing, and analysis time. This acceleration relies on the high sensitivity afforded by a triple-resonance cryogenic probe, and is enabled by the integration of software packages AutoProc, AutoPeak, NMRPipe [17], Sparky [18], and AutoAssign [9, 13]. In particular, the AutoPeak software provides flexibility for using various combinations of triple-resonance spectra in AutoAssign analysis. In this favorable case, extensive backbone resonance assignments and reliable identification of secondary structure was completed using less than 6 h of data collection, processing, peak picking, and analysis time. The use of pulse sequences designed for the short measurement times and small number of scans per FID made feasible by cryoprobes, such as reduced-dimensionality NMR experiments [e.g. see References 29 and 30], as well as the use of shorter recycling delays and numbers of points in indirect dimensions, can provide further reductions in data collection times. Hardware and pulse sequence design providing improvements in water suppression when using cryogenic probes, especially for data sets collected with only one scan per increment, will be critical for these applications. Finally, this study demonstrates the feasibility of high-throughput analysis of small protein structures using NMR, and the potential for applying NMR in large-scale structural proteomics projects.

Acknowledgements

The AutoProc, AutoPeak, and AutoAssign software packages described in this paper are available to non-commercial users at (www-nmr.cabm.rutgers.edu). This work was supported by grants from the National Institute of Health (P50 GM62413 and R01 GM56233), The New Jersey Commission on Science and Technology (99-2042-007-13), the Spanish Science and Education Ministry (postdoctoral Fellowship EX-29179910 to DM) and the National Science Foundation (postdoctoral Fellowship DBI-9974200 to HNBM).

References

1. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
2. Wüthrich, K. (1995) *Acta. Cryst.* **D51**, 249–270.
3. Clore, G. M., and Gronenborn, A. M. (1991) *Science* **252**, 1390–1399.
4. Montelione, G. T., and Wagner, G. (1990) *J. Magn. Reson.* **87**, 183–188.
5. Ikura, M., Kay, L. E., and Bax, A. (1990) *Biochemistry* **29**, 4659–4667.
6. Kay, L. E. (1995) *Prog. Biophys. Mol. Biol.* **63**, 277–299.
7. Montelione, G. T., Rios, C. B., Swapna, G. V. T., and Zimmerman, D. E. (1999) *Biol. Magn. Reson.* **17**, 81–130.
8. Zimmerman, D. E., Kulikowski, C. A., Wang, L. L., Lyons, B. A., and Montelione, G. T. (1994) *J. Biomol. NMR* **3**, 241–256.
9. Zimmerman, D. E., Kulikowski, C. A., Feng, W. *et al.* (1997) *J. Mol. Biol.* **269**, 592–610.
10. Leutner, M., Gschwind, R. M., Liermann, J., Schwarz, C., Gemmecker, G., and Kessler, H. (1998) *J. Biomol. NMR* **11**, 31–43.
11. Güntert, P., Salzmann, M., Braun, D., and Wüthrich, K. (2000) *J. Biomol. NMR* **18**, 129–137.
12. Moseley, H. N. B., and Montelione, G. T. (1999) *Curr. Opin. Struct. Biol.* **9**, 635–642.
13. Moseley, H. N. B., Monleón, D., and Montelione, G. T. (2001) *Meth. Enzymol.* **339**, 91–108.
14. Styles, P., and Soffe, N. F. (1984) *J. Magn. Reson*. **60**, 397–402.
15. Marek, D. (1993) US Patent 5,247,256. *RF Receiver Coil Arrangement for NMR Spectrometers*.
16. Goger, M., McDonnell, J., and Cowburn, D. (2000) In: *Abstract Book, 41st Experimental NMR Conference*, Asilomar, CA, pp. 267.
17. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR* **6**, 277–293.
18. Goddard, T. D., and Kneller, D. G. (1999) *SPARKY 3*, University of California, San Francisco.
19. Montelione, G. T., Zheng, D., Huang, Y. J., Gunsalus, K. C., and Szyperski, T. A. (2000) *Nat. Struct. Biol.* **7**, 982–985.
20. Jansson, M., Li, Y. C., Jendeberg, L., Anderson, S., Montelione, G. T., and Nilsson, B. (1996) *J. Biomol. NMR* **7**, 131–141.
21. Marion, D., Ikura, M., and Bax, A. (1989) *J. Magn. Reson.* **84**, 425–430.
22. Barkhuijsen, H., De Beer, R., Bovée, W. M. M. J., and Van Ormondt, D. (1985) *J. Magn. Reson.* **61**, 465–481.
23. Zhu, G., and Bax, A. (1992) *J. Magn. Reson*. **98**, 192–199.
24. Grzesiek, S., and Bax, A. (1993) *J. Biomol. NMR* **3**, 185–204.
25. Rios, C. B., Feng, W., Tashiro, M., Shang, Z., and Montelione, G. T. (1996) *J. Biomol. NMR* **8**, 345–350.
26. Otting, G., Liepinsh, E., and Wüthrich, K. (1993) *Biochemistry* **32**, 3571–3582.
27. Szyperski, T. A., Luginbühl, P., Otting, G., Güntert, P., and Wüthrich, K. (1993) *J. Biomol. NMR* **3**, 151–164.
28. Wishart, D. S., and Sykes, B. D. (1994) *J. Biomol. NMR* **4**, 171–180.
29. Szyperski, T., Wider, G., Bushweller, J. H., and Wüthrich, K. (1993) *J. Am. Chem. Soc.* **118**, 9307–9308.
30. Szyperski, T., Banecki, B., Braun, D., and Glaser, R. W. (1998) *J. Biomol. NMR* **12**, 25–37.