Structural bioinformatics

APART: Automated Preprocessing for NMR Assignments with Reduced Tedium

Norma H. Pawley, Jason D. Gans and Ryszard Michalczyk∗
Los Alamos National Laboratory, Bioscience Division, MS G758, Los Alamos, NM 87545, USA

Received on June 11, 2004; revised on September 2, 2004; accepted on September 19, 2004
Advance Access publication September 23, 2004

ABSTRACT
Motivation: High-throughput NMR structure determination is a goal that will require progress on many fronts, one of which is rapid resonance assignment. An important rate-limiting step in the resonance assignment process is accurate identification of resonance peaks in the NMR spectra. Peak-picking schemes range from incomplete (which lose essential assignment connectivities) to noisy (which obscure true connectivities with many false ones). We introduce an automated preassignment process that removes false peaks from noisy peak lists by requiring consensus between multiple NMR experiments and exploiting a priori information about NMR spectra. This process is designed to accept multiple input formats and generate multiple output formats, in an effort to be compatible with a variety of user preferences.
Results: Automated preprocessing with APART rapidly identifies and removes false peaks from initial peak lists, reduces the burden of manual data entry, and documents and standardizes the peak filtering process. Successful preprocessing is demonstrated by the increased number of correct assignments obtained when data are submitted to an automated assignment program.
Availability: APART is available from http://sir.lanl.gov/NMR/APART.html
Contact: npawley@lanl.gov; rmichalczyk@lanl.gov
Supplementary information: Manual pages with installation instructions, procedures and screen shots can also be found at http://sir.lanl.gov/NMR/APART_Manual1.pdf

INTRODUCTION

Three-dimensional protein structures can be obtained using NMR spectroscopy; however, the process is generally time-consuming and expertise intensive. High-throughput NMR structure determination is a goal that will require progress on many fronts, one of which is rapid resonance assignment.

An important rate-limiting step in the resonance assignment process is accurate identification of resonance peaks in the NMR spectra. NMR spectra are noisy. Hence, both manual and automatic peak-picking schemes navigate between the extremes of reliable but incomplete picking, and noisy but complete picking. Each of these extremes complicates the assignment process: incomplete peak-picking results in the loss of essential connectivities, while noisy picking conceals the true connectivities under a combinatorial explosion of false positives. Consequently, additional processing is often applied to peak lists before data are submitted to an assignment program. The goal of such processing is to simplify the assignment process by preferentially removing false peaks from noisy peak lists. Many NMR practitioners currently perform such processing “by hand”, which is tedious and extremely user-dependent for success.

To increase efficiency, we have created a systematic and automated alternative to manual preprocessing, known as APART. The advantages of APART over manual preprocessing include convenience, standardization and reduced manual data entry and formatting. With a single function call, APART reads in peak lists and executes a series of common preprocessing steps, including calculation and application of interspectral referencing, grouping peaks and editing peaks. Peak editing is performed according to criteria such as presence in a reference peak list (typically HSQC), presence in multiple experiments, relationship to the distribution of chemical shifts reported in the BMRB and expectations for a given spin system.

Currently, two main alternatives to manual preprocessing are practiced in the NMR community. One option is NvAssign (Kirby et al., 2004). This preprocessing program is integrated with the spectral analysis software package NMRView (One Moon Scientific, Inc.; Johnson and Blevins, 1994), making it a powerful tool for anyone currently using NMRView to analyze their spectra. The other option is to skip pre-processing altogether and rely on the ability of the assignment program to discriminate noise peaks. This is a risky strategy, since most assignment programs clearly state that their performance degrades with the signal-to-noise ratio (number of true peaks/number of false peaks) of the peak lists (Bartels et al., 1997; Zimmerman et al., 1997; Hyberts and Wagner, 2003; Slupszyk et al., 2003).

For those preferring spectral analysis software other than NMRView (e.g., Sparky (Goddard and Kneller, 2004), XEASY (Bartels et al., 1995), Felix (MSI, Inc.), NMRDRAW (Delaglio et al., 1995), and so on) or assignment programs not integrated with NMRView (e.g., AutoAssign (Zimmerman et al., 1997), SMARTNOTEBOOK (Slupszyk et al., 2003), IBIS (Hyberts and Wagner, 2003), and so on), APART provides comprehensive preprocessing appropriate for incorporation into many alternate analysis/assignment pathways. For those currently omitting preprocessing altogether, APART provides rapid improvement in the signal-to-noise ratio of peak lists, increasing the probability of obtaining optimal results from semi-automated or automated assignment programs.

∗To whom correspondence should be addressed.
Fig. 1. Successful removal of false peaks from ubiquitin peak lists at three distinct noise levels. All true peaks present in the initial peak lists are retained throughout the filtering process. Change in signal-to-noise ratios with successive filtering steps: (A) 0.17 → 1.3 → 3.3, (B) 0.58 → 2.6 → 7.5, (C) 3.7 → 12 → 27 and (D) 2.6 → 7.5 → 15. Note that at least a factor of two in signal-to-noise is gained with each step of the filtering process. Spectra of the 8.5 kDa (76 amino acids) ubiquitin protein were collected on a Bruker 500 MHz Avance spectrometer, at 278 K. Peak lists were generated from the spectra using the automated peak-picking function of NMRDraw at three different contour levels (details of NMRDraw parameters are available in the Supplementary information). A total of 552 ‘true’ peaks were present in the spectra at the lowest contour level, 548 were present at 2× lowest contour and 517 were present at 4× lowest contour. In all the three cases, 13 peaks present in the spectra were not automatically picked, due to overlap. This is a function of the resolution of the spectra more than the quality of the peak-picking tool (e.g. the resolution of the CBCACONH spectrum is 0.62 ppm/pt in the 15 N dimension and 0.47 ppm/pt in the 13 C dimension). The success of the optional interactive filter is expected to be user-dependent. Hence, the sample result shown in the figure should be interpreted qualitatively.

RESULTS

The development of APART has focused on facilitating the resonance assignment process by simplifying, integrating and standardizing preprocessing steps, and by supporting a variety of analysis architectures, as summarized below:

- APART performs thorough, stand-alone preprocessing. Starting from automatically or manually picked peak lists, APART executes a complete series of processing steps. The processing includes formatting the output for the analysis/assignment program of the users’ choice.
- APART standardizes and documents the peak filtering process. Every processing step executed by APART, including interactions with the user, is documented. Relevant intermediate states are archived.
- APART accommodates user preferences for input/output. APART accepts several predefined input formats, including NMRDraw (Delaglio et al., 1995), NMRView (Johnson and Blevins, 1994) and Sparky (Goddard and Kneller, 2004). In addition, other input formats can be defined by the user ‘on the fly’ [e.g. XEasy] (Bartels et al., 1995). Implemented output formats include filtered versions of any input format, including those defined by user [e.g. XEasy/IBIS] (Hyberts and Wagner, 2003), and several predefined output formats, including AutoAssign (Zimmerman et al., 1997), MONTE (Hitchens et al., 2003), PACES (Coggins and Zhou, 2003) and Smartnotebook (Slupsky et al., 2003).

The success of APART is defined by its ability to substantially reduce the presence of false peaks in peak lists while retaining the true peaks. The performance of APART on ubiquitin peak lists with three distinct...
noise levels is shown in Figure 1. In each case, the signal-to-noise ratio of the peak lists improves appreciably, and all true peaks present in the initial peak lists are retained throughout the filtering process.

The improvement in the signal-to-noise ratio of the peak lists following application of APART translates into improvements in obtaining resonance assignments. For example, processing peak lists with APART before submitting them to the AutoAssign program yields a dramatic improvement in the performance for noisy peak lists: at the lowest contour level, the number of correct assignments increases from 5 (for ‘raw’ peak lists) to 36 (for APART-processed peak lists). Preprocessing yields modest improvement even for peak lists with very little noise: at the highest contour level, the number of correct assignments increases from 56 (for ‘raw’ peak lists) to 59 (for APART-processed peak lists). Hence, to obtain good results, it is useful to apply APART even to the most ideal data, and it is essential to apply APART to non-ideal data.

ACKNOWLEDGEMENTS

We thank Theresa A. Ramelot and Ryan McKay for their valuable comments and discussion. This research was supported by the Los Alamos National Laboratory LDRD program, grant X1VE.

REFERENCES


