NvMap: automated analysis of NMR chemical shift perturbation data

Lisa Fukui and **Yuan Chen**[*]
Division of Immunology, Beckman Research Institute of the City of Hope, 1450 East Duarte Road, Duarte, CA 91010, USA

Abstract

Summary—NMR chemical shift perturbation experiments are widely used to define binding sites in biomolecular complexes. Especially in the case of high throughput screening of ligands, rapid analysis of NMR spectra is essential. NvMap extends NMRViewJ and provides a means for rapid assignments and book-keeping of NMR titration spectra. Our module offers options to analyze multiple titration spectra both separately and sequentially, where the sequential spectra are analyzed either two at a time or all simultaneously. The first option is suitable for slow or intermediate exchange rates between free and bound proteins. The latter option is particularly useful for fast exchange situations and can compensate for the lack of indicators for overlapped peaks. Our module also provides a simple user interface to automate the analysis process from dataset to peak list. We demonstrate the effectiveness of our program using NMR spectra of SUMO in complexes with three different peptides.

1 INTRODUCTION

Nuclear magnetic resonance (NMR) chemical shift perturbation experiments are widely used to characterize molecular complex formation and map binding interfaces. One use of such an approach is drug screening, where many ligands are tested for interaction with the desired protein (Lepre *et al*., 2004). With NMR, proteins can be studied in solution and individual residues correlate to specific peaks in NMR spectra. When a protein or nucleic acid forms a complex, the nuclei at the binding interface will experience changes in their environments, which lead to chemical shift perturbation. Thus, chemical shift perturbation can be used to efficiently map binding sites. In this type of study, it is important to assign the cross peaks that show chemical shift changes in the spectra of the complex, and estimate the differences in chemical shift changes between the free and bound proteins or nucleic acids (Cavanagh *et al*., 1995).

Depending on the exchange rate between the free and bound species, the appearance of peak movement may vary during the titration of one molecule with another molecule. In the case of fast exchange, cross peaks change according to the population weighted average of the free and bound chemical shifts, and thus these peaks move linearly from the peak position of the free to that of the bound. In this case, assignments of the bound spectra can be obtained reliably by following the movements of the peaks from the free to the bound positions. On the other

hand, in the case of slow exchange, separate peaks correspond to the free and bound states. In this case, chemical shift assignments cannot be completely obtained without information from 3D triple resonance experiments. In the case of intermediate exchange, severe line-broadening effects occur during titration (Cavanagh *et al.*, 1995). In all cases, tracking these chemical shift perturbations manually is tedious, especially when many peaks exhibit chemical shift changes.

A couple of automated methods have been proposed for the automation of NMR data analysis in these chemical shift perturbation studies. MUNIN (Damberg *et al.*, 2002) identifies spectra that have different peak positions without peak-picking, but it does not provide assignments. FELIX-Autoscreen (Peng *et al.*, 2004) is available as a FELIX software module to automate the analysis of titration spectra using a simulated annealing approach to identify the assignments. This is based on the assumption that the correct assignments of the complex spectrum should produce the minimum overall chemical shift difference between the free and bound spectra. This approach cannot handle the cases where cross peaks are missing, such as in intermediate exchange situations. Additionally, the assignments do not take advantage of the basic NMR principle in the fast exchange situation.

We have developed a computational approach that is fast, user-friendly and more versatile for handling the different spectral properties caused by the different exchange rates. In fast exchange cases, more than two spectra can be considered simultaneously at a time. For intermediate and slow exchange situations, the two-spectra comparison option can be chosen. Our program is incorporated into the widely used software package NMRViewJ (Johnson, 2004). In our extension, we have constructed a straightforward interactive window that can be easily accessed through the main menu of NMRViewJ.

2 METHODS

The input to NvMap consists of a reference peak list, usually the assignment of a free molecule, and one or more NMR spectra. Given the input, our module produces a new peak list for each spectrum where each picked peak corresponds to a peak in the reference spectrum of the free molecule. To assign the spectrum of a bound molecule, NvMap first utilizes the automatic peak-picking ability of NMRViewJ to conduct peak-picking. The contour level can be adjusted by users to include more peaks picked than in the original spectrum.

With the newly picked peaks, our program finds the distances between nearby (reference peak, new peak) pairs of peaks, connecting the provided peaklist with the generated peaklist. We use a greedy algorithm, a method in which we look for the best local solution at each step (Cormen *et al.*, 2001), to sequentially match these pairs of peaks. Initially, all new peaks and all reference peaks are eligible to be matched to each other. For each step, we first find the closest pair of reference and new peaks. Then the new peak is given the identification (ID) of the corresponding reference peak. After the new peak has been assigned a designated ID, both the reference peak and the new peak are removed from a list of eligible peaks. In the next step, we find the next closest pair of reference and new peaks, where both peaks are from the list of eligible peaks. NvMap continues to find pairs of peaks until the distance between these pairs reaches a threshold distance, which can be specified by the user. The new peaks with established IDs (i.e. the new peaks paired with nearby reference peaks) are members of a new peak list, ready for use in NMRViewJ.

When provided with two or more spectra, the user can choose to incorporate additional routines to recover and match unpaired peaks. These routines take advantage of the fact that peaks tend to move in a straight line with increasing concentrations of ligand.

3 RESULTS

Our NMRViewJ module is capable of taking a reference peak list and then copying the peak numbers and assignments of this peak list to any number of given spectra. NvMap includes menus for picking among the imported peak lists and datasets, and allows the user to name new peaklists. In addition, the user can specify distance and peak picking thresholds as desired.

The user has the choice of either using the same reference peak list for every dataset or to sequentially process the data, using each previously generated peak list as a reference for the next. The first choice offers convenience for those who prefer to analyze several different ligand-binding spectra at once, such as in drug screening. We expect that the sequential mode of our software will be particularly useful for analyzing fast-exchange protein titrations.

We tested this program on several NMR titration data involving SUMO and three different peptides known as PIASx-P, PML, and SAE2 (Song *et al.*, 2004). In the cases of PML (Fig. 1) and SAE2 peptides, we used a series of five titration spectra with increasing concentrations of peptides. NvMap was able to assign ~95% of the peaks correctly, including 90% of the peaks that differ in the free and bound state. By simultaneous analysis of five spectra, the program was able to resolve overlapping peaks in the middle titration points (Fig. 1). The main cause of errors was the peak picking ability of NMRViewJ; the peak picking subroutine cannot distinguish between a single peak and multiple overlapping peaks. A NMRViewJ user can overcome this difficulty by manually identifying overlapped peaks and adding them to their peak list. This user may also lower the peak picking threshold level to pick more peaks. NvMap considers multiple spectra in a titration to compensate for the lack of indicators for overlapped intermediate peaks. For PIASx-P, a ligand that participates in a slow exchange interaction, we gave NvMap only the reference peak list and complex spectrum, and NvMap was able to assign 75% of the cross peaks correctly. The incorrectly assigned peaks in PIASx-P were largely due to the lack of sufficient information for assignments from only two HSQC spectra in the slow exchange situation. Nonetheless, NvMap greatly speeds up the data analysis process for any chemical shift perturbation analysis.

Acknowledgments

REFERENCES

Cavanagh, J.; Fairbrother, WJ.; Palmer, AGI.; Skelton, NJ. Protein NMR Spectroscopy: Principles and Practice. Academic Press; 1995.

Cormen, TH.; Leiserson, CE.; Rivest, RL.; Stein, C. Introduction to Algorithms. MIT Press; Cambridge, MA: 2001. Chapter 16; p. 370-404.

Damberg CS, et al. Automated analysis of large sets of heteronuclear correlation spectra in NMR-based drug discovery. J. Med. Chem 2002;45:5649–5654. [PubMed: 12477348]

Johnson BA. Using NMRView to visualize and analyze the NMR spectra of macromolecules. Methods Mol. Biol 2004;278:313–352. [PubMed: 15318002]

Lepre CA, et al. Theory and applications of NMR-based screening in pharmaceutical research. Chem. Rev 2004;104:3641–3676. [PubMed: 15303832]

Peng C, et al. Automated evaluation of chemical shift perturbation spectra: New approaches to quantitative analysis of receptor-ligand interaction NMR spectra. J. Biomol. NMR 2004;29:491–504. [PubMed: 15243180]

Song J, et al. Identification of a SUMO-binding motif that recognizes SUMO--modified proteins. Proc. Natl Acad. Sci. USA 2004;101:14373–14378. [PubMed: 15388847]
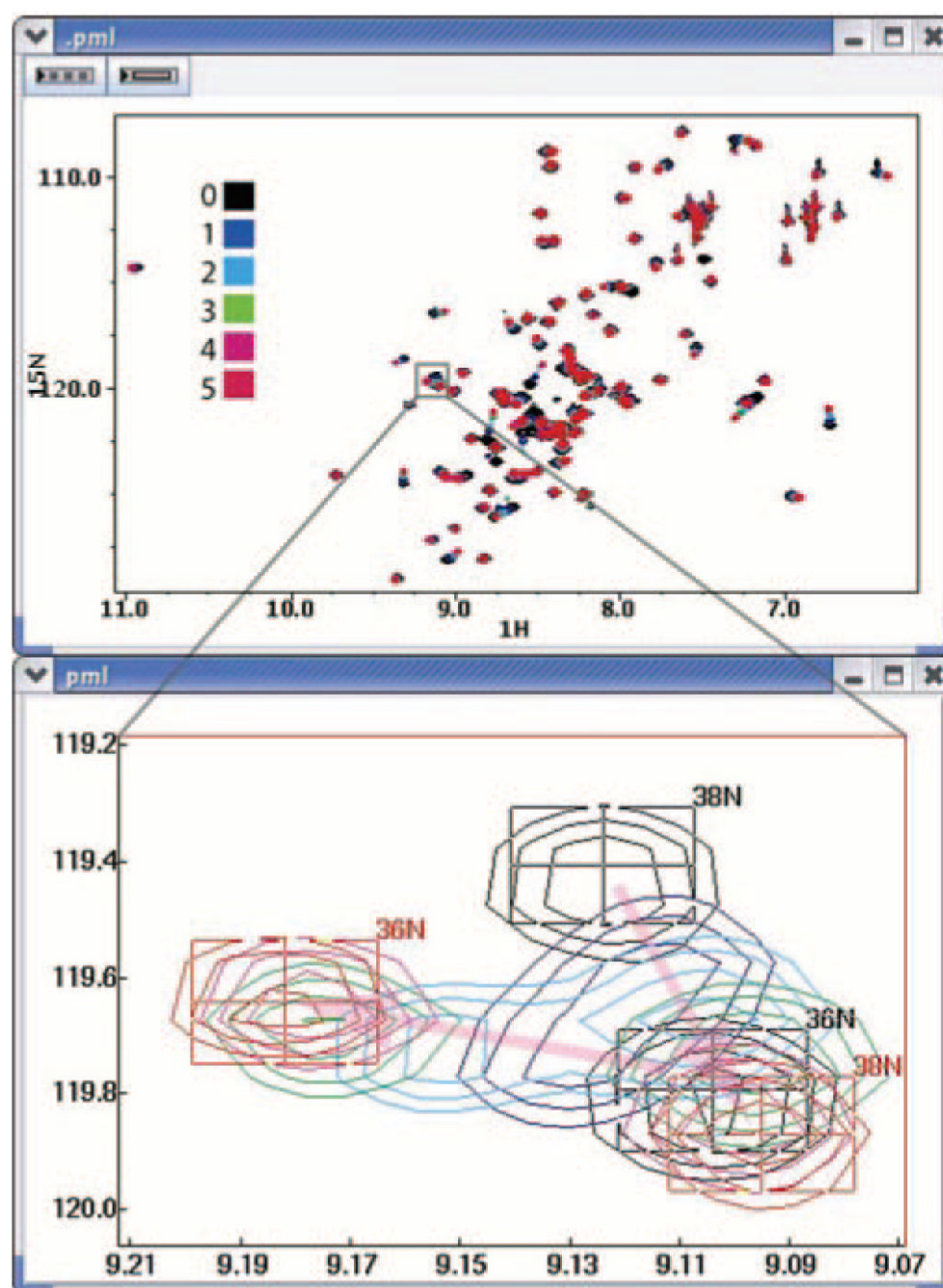
Fig. 1.
Peak movement in a NMR titration experiment. We show the overall spectrum (top) and a close-up view of two peaks' movement (bottom) for a titration experiment with SUMO and a PML peptide. Given the initial locations of peaks (black) and titration spectra (1–5), NvMap can trace peak movement through the spectra to produce related peak lists. By considering >2 spectra at a time, our program can find the correct assignment even when intermediate peaks overlap, such as in spectrum 1 in the close-up view. Note that finding the closest final peak does not necessarily generate the correct assignment in this case.