Structural bioinformatics

PINE-SPARKY.2 for automated NMR-based protein structure research

Woonghee Lee* and John L. Markley*

National Magnetic Resonance Facility at Madison, Biochemistry Department, University of Wisconsin-Madison, Madison, WI 53706, USA

*To whom correspondence should be addressed.

Abstract

Summary: Nuclear magnetic resonance (NMR) spectroscopy, along with X-ray crystallography and cryo-electron microscopy, is one of the three major tools that enable the determination of atomic-level structural models of biological macromolecules. Of these, NMR has the unique ability to follow important processes in solution, including conformational changes, internal dynamics and protein–ligand interactions. As a means for facilitating the handling and analysis of spectra involved in these types of NMR studies, we have developed PINE-SPARKY.2, a software package that integrates and automates discrete tasks that previously required interaction with separate software packages. The graphical user interface of PINE-SPARKY.2 simplifies chemical shift assignment and verification, automated detection of secondary structural elements, predictions of flexibility and hydrophobic cores, and calculation of three-dimensional structural models.

Availability and implementation: PINE-SPARKY.2 is available in the latest version of NMRFAM-SPARKY from the National Magnetic Resonance Facility at Madison (http://pine.nmrfam.wisc.edu/download_packages.html), the NMRbox Project (https://nmrbox.org) and to subscribers to the SBGrid (https://sbgrid.org). For a detailed description of the program, see http://www.nmrfam.wisc.edu/pine-sparky2.htm.

Contact: whlee@nmrfam.wisc.edu or markley@nmrfam.wisc.edu

Supplementary information: Supplementary data are available at Bioinformatics online.

1 Introduction

Numerous different groups and institutions have developed the software packages in current use in the field of biomolecular NMR, and many utilize different nomenclatures, data input procedures, and even computer operating systems. These differences have impeded research progress, particularly by non-experts. The Integrative NMR package (Lee et al., 2016a) offered a partial solution by integrating NMRFAM-SPARKY (Lee et al., 2015) with APES for peak picking (Shin et al., 2008), PINE for automated assignment (Bahrani et al. 2009), ARECA (Dashti et al. 2016) for validation of peak assignments; TALOS-N for shift based torsion angle restraints (Shen and Bax, 2013), CS-Rosetta (Shen et al., 2008), for structure determination from chemical shifts, AUDANA (Lee et al., 2016b) and PONDEROSA-CS (Lee et al., 2014) for automated structure determination from NOE spectra, and data visualization by NDP-PLOT and an enhanced mode of the PyMOL software package (The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC.).

However, as part of this package, the original PINE-SPARKY (Lee et al., 2009) was cumbersome. Users had to pick peaks from NMR spectra, generate a set of peak list files, open a web browser, visit the PINE web page and submit generated peak list files one-by-one for each experiment. To import and verify chemical shifts, the user had to wait for an email notice, download and unpack the compressed results, use the PINE2SPARKY converter to apply PINE probabilistic assignments to SPARKY projects, and use PINE-SPARKY extensions to create the actual assignment labels. Only after following these steps could the user carry out further analysis, such as validation of chemical shift referencing by LACS.
PINE-SPARKY.2 supports 19 different NMR experiments. The user specifies the NMR experiments with spectral data to be analyzed by clicking the Add button from the spectrum list (Fig. 1D). Peaks need to be identified in the spectra to be analyzed, and this can be accomplished with the automated peak-picking program APES (two-letter-code ae). Then, an assignment job is submitted to the PINE web server (Supplementary Fig. S1) by clicking the Submit button. A unique Key identifier generated by the PINE-SPARKY importer (Fig. 1F) handles cross-talk between PINE-SPARKY2 and the PINE web server. PINE-SPARKY2 checks the status file from the URL associated with the Key, and the PINE web server updates the status of the PINE job in the status file. Predictions of secondary structures (PECAN), referencing errors (LACS), hydrophobicities (PACSY), torsion angles and flexibilities (TALOS-N), and 3D structures (CS-ROSETTA) are executed sequentially by BASH and PYTHON scripts based on the chemical shifts with the highest probabilities given by PINE (described more fully in the manual). By clicking the Check button, the PINE-SPARKY importer retrieves the results. Replacement of previously downloaded results can be accomplished by clicking the Browse button before clicking on Check. The PINE-SPARKY importer automatically sets up visualization of the PINE results and incorporates them into the current project. It asks a series of interactive yes/no questions to determine whether the user wants to (i) download the results in the PINE sub-directory under working directory; (ii) visualize secondary structures determined by the PECAN algorithm (Supplementary Fig. S2A); (iii) visualize PINE probabilities for spin system assignments (Supplementary Fig. S2B); (iv) visualize chemical shift referencing analysis by the LACS algorithm (Supplementary Fig. S2C); (v) visualize hydrophobic core residues predicted from PACSY database (Supplementary Fig. S2D; Lee et al., 2012); (vi) visualize RCI S2 (random coil index order parameter; Berjanskii and Wishart, 2005); and/or (vii) generate PINE
probabilistic labels and accept the most probable ones with $P > 0.5$ (Supplementary Fig. S3). Then the Completeness Counter (two-letter-code cm) can be used to find unassigned resonances (Fig. 1G).

As a test of PINE-SPARKY.2, we used data from three multidimensional NMR spectra (2D $^1$H, $^{15}$N-HSQC, 3D CBCA(CO)NH and HNCACB) from the small (110 amino acid residue) protein AeSCP-2 (BMRB Entry 16662) as inputs for automated assignment and fed the assignment results into CS-Rosetta for structure determination from chemical shifts alone. We compared the resulting structure (Fig. 1H) with that determined manually from NOE data (PDB ID 2KSH; Singarapu et al., 2010). Following superposition, the pairwise backbone RMSD for the two structures was 1.21 Å and the all-heavy-atom RMSD was 2.14 Å (Supplementary Fig. S4; see the Supplementary Material for details).

Acknowledgements

PINE-SPARKY2 utilizes the CS-Rosetta web server service provided by BioMagResBank (https://csrosetta.bmrb.wisc.edu/csrosetta); we are grateful to Jon Wedell for its maintenance.

Funding

Supported by the United States National Institutes of Health (P41GM103399). Conflict of Interest: none declared.

References


