Center for Eukaryotic Structural Genomics

Protein Structure Initiative

PINE-SPARKY: VISUALIZATION AND REFINEMENT OF NMR ASSIGNMENTS

Woonghee Lee,[2] William M. Westler,[2] Arash Bahrami,[1,2] Hamid R. Eghbalnia,[1,‡] John L. Markley[1,2]

[1]Center for Eukaryotic Structural Genomics (CESG) and [2]NMRFAM (National Magnetic Resonance Facility at Madison),
University of Wisconsin-Madison, Department of Biochemistry, 433 Babcock Drive, Madison, WI USA 53706-1549 http://www.uwstructuralgenomics.org

Abstract

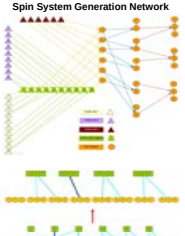
PINE-NMR [1] is a protein NMR assignment package that accepts, as input, the amino acid sequence of a protein and peak lists associated with defined NMR experiments and provides, as output, probabilistic backbone and side chain assignments and an analysis of the secondary structure. PINE-NMR can accommodate prior information about assignments or stable isotope labeling schemes. As part of the analysis, PINE-NMR identifies, verifies, and rectifies problems related to chemical shift referencing or erroneous input data. PINE-NMR achieves robust and consistent results that have been shown to be effective in subsequent steps of NMR structure determination. As an adjunct to PINE-NMR, we have developed PINE-SPARKY, a software package that complements the automated approach in cases where the input peak list data may need to be reconsidered. PINE-SPARKY provides for the rapid and efficient visualization and refinement of PINE-NMR results in the context of the experimental data that underlay the peak lists. PINE-SPARKY is based on the very popular SPARKY package [2]. The current version of PINE-SPARKY consists of a converter (PINE2SPARKY), which takes the output from a PINE-NMR analysis and transforms it into SPARKY input, and a number of SPARKY extensions. Possible alternative assignments given by PINE are incorporated into SPARKY save files by the PINE2SPARKY converter and visualized as labels in SPARKY's spectrum view. Three SPARKY extensions (PINE Assigner, PINE Graph Assigner, and Assign the Best by PINE) serve to manipulate the labels. The functions provided by these extensions make the assignment process easy and fast. The software supports rapid peak-by-peak or atom-by-atom assignment confirmation. Alternatively, automatic selection by the highest probability is enabled. PINE Assigner lists all possible assignments for a peak selected in the dialog box and enables the user to choose among these. A window in PINE Graph Assigner provides a graphical display of all atoms in a selected residue along with all the atoms in the residues adjacent to it. It displays a ranked list of PINE-derived connectivity assignments to any selected atom. Assign the Best-by-PINE allows the user to choose a probability threshold and to automatically fix all assignments above that threshold; following this operation, only the less certain assignments need to be examined visually. Once assignments are fixed at a certain level, further refinement can be obtained by rerunning PINE-NMR. We will demonstrate how the PINE-NMR and PINE-SPARKY duo speed up the task of protein structure determination. The PINE-NMR server and PINE-SPARKY (which operates on all platforms that support SPARKY) are freely available at http://pine.nmrfam.wisc.edu/PINE-SPARKY We welcome comments from users.

What is PINE?

PINE is a program that uses the sequence of a protein in conjunction with peak lists (or spin systems) from a variety of NMR experiments to generate the following: (1) automatic backbone and sidechain assignments, (2) detection and automated correction of potential referencing problems or inconsistent assignments, and (3) secondary structure determination.

General Overview of Probabilistic Network Defined by PINE

Spin System Generation Network

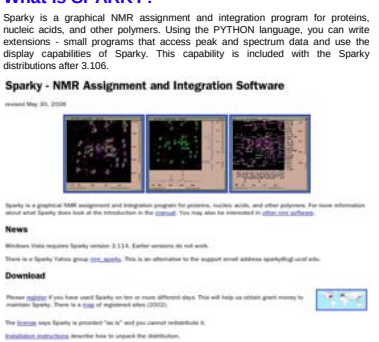


PINE Homepage: http://pine.nmrfam.wisc.edu

What is SPARKY?

Sparky is a graphical NMR assignment and integration program for proteins, nucleic acids, and other polymers. Using the PYTHON language, you can write extensions - small programs that access peak and spectrum data and use the display capabilities of Sparky. This capability is included with the Sparky distributions after 3.106.

Sparky - NMR Assignment and Integration Software



SPARKY Homepage: http://www.cgl.ucsf.edu/home/sparky/

What is the Goal of PINE-SPARKY?

Using unassigned peaks (which may be generated using SPARKY), our PINE server provides assignments for each peak along with the probability that the assignment is correct. In cases of ambiguity, the PINE server provides multiple possible assignments ranked by probability. PINE-SPARKY enables users to visualize the PINE assignments with reference to the underlying spectral data, so that correct choices can be made. We chose SPARKY, because it currently is the most widely used visualization tool for protein NMR.

To make the PINE output work in SPARKY, we did following:

(1) Developed a converter for PINE output to work as input to SPARKY. The converter allows multiple PINE assignments for each peak to be input into SPARKY.
(2) Developed SPARKY PYTHON extensions for manipulating and visualizing the information from PINE.

Components of PINE-SPARKY

(1) PINE2SPARKY converter:
Converts multiple assignments from PINE to SPARKY labels.

(2) SPARKY extensions for visualization and manipulation:
PINE Assigner
PINE Graph Assigner
Assign the Best by PINE

New extensions developed

PINE2SPARKY Converter

1. The "PINE2SPARKY" converter handles multiple probabilistic PINE assignments.
PINE does not provide absolute assignments for each peak. It provides possible assignments including the probability that the assignment is correct. Multiple possible assignments are converted into labeled objects, and the labeled objects are incorporated into SPARKY save files. Because a peak cannot be attached to several labels at once in SPARKY, the labels must include position information. The information is saved not only in SPARKY save files but also in newly created files with names that have "_pl" at the tail of the save files, because the labels are temporary and can be changed later.

The format of the labels is similar to the following:

possible assignments[probability]:(peak positions)
ex) A14CA-A14N-A14H[0.988]:(60.342,120.231,7.764)

A single peak can have many labels, and once the user determines which assignment is correct, the incorrect labels can be removed. The probable ranks are color coded: in the default settings the coolest color (blue) indicates the highest probability. This color coding can be redefined by the user.

2. User Interface
The user interface of PINE2SPARKY consists of roughly four parts: (1) SPARKY input, (2) PINE output, (3) Sparky output, and (4) controls.

(1) Sparky input: If you have a peak picked SPARKY project and save files, they can be set here. The PINE2SPARKY converter will use them as templates.

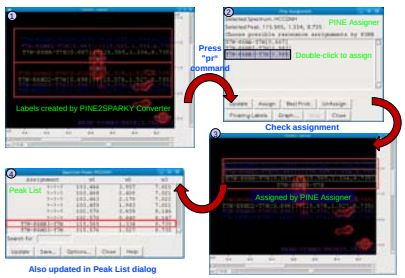
(2) PINE output: You will need a protein sequence file and at least one of the native PINE probabilistic output files. The sequence file can use either the three- or one-letter code.

(3) Sparky output: SPARKY output files are automatically set when you set SPARKY input files. Or you can set them manually if ne SPARKY files are generated without templates.

(4) Controls:
(a) Preferences: Tolerances and first index settings.
(b) Spectrum identifier: Set spectrum types to generate proper artificial spin systems.
(c) Options: The color of the labels for one peak is determined by the ranks of the probabilities. In the default settings, blue is set to represent the highest probability and purple is the next highest. If you want to change colors for ranks, press the "Option..." button and change them.
(d) PINE2SPARKY: After the initial setting up, probable assignments from PINE are converted into PINE labels in the Sparky save files by pressing the "PINE2SPARKY" button. PINE labels in Sparky save files are not eternal: one can change the position of the labels or remove them entirely. The time it takes to convert a file is dependent on how many peaks are present in the project.

SPARKY Extensions: PINE Assigner

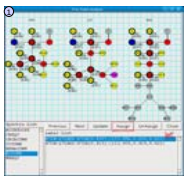
The SPARKY extension, "PINE Assigner", is a dialog box that lists all possible PINE assignments the selected peak. To select the most likely assignment, the user clicks on the desired label in the list. *Command "pr"*



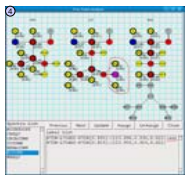
We recommend you use "PINE Assigner" and "Peak List" dialog together. Typing the "lt" command will bring up the "Peak List" dialog box, and you can select a peak in the "Peak List" by double-clicking on the desired peak. Then, type "pr" or select "PINE Assigner" from the extension tab to open the "PINE Assigner" window.

SPARKY Extensions: PINE Graph Assigner

"PINE Graph Assigner" is a graphical diagram window that displays all the atoms from the selected residue as well as the atoms in the residues adjacent to it. The window also contains a list of the NMR experiments PINE used for the assignment. When one selects an individual atom from the structure and a spectrum, "PINE Graph Assigner" displays a ranked list of PINE-derived connectivity assignments to that atom from that spectrum.



(1) First, select a peak from the spectrum you want to assign. Because "PINE Graph Assigner" works atom by atom, rather than peak by peak, PINE Graph will recognize the corresponding atom in a particular residue. Open up "PINE Graph Assigner" by typing the "pp" command.
(2) Check the peak position in the spectrum by double-clicking in the Label List. If you decide it is correct, press the "Assign" button.
(3) Assignment is established and it can be seen in the spectrum and also in the PINE Graph Assigner.
(4) Move to another atom to be assigned.

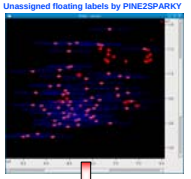


SPARKY Extensions: Assign the Best by PINE

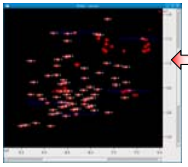
When you use "PINE Assigner" to assign a peak, you can select the assignment with the highest probability. Certainly, this is faster than assigning every peak by selecting and assigning peaks one-by-one the spectrum view. However, if you want to automatically assign the labels with the highest probability, you can use the "Assign the best by PINE" extension.

"Assign the Best by PINE" performs assignment with the highest probability for all of the residues represented in a peak. You can set a minimal threshold for probability, and the function will only assign peaks that have higher probabilities than the minimal threshold. This function will help you to save time by reducing repetitive operation. "Assign the best by PINE" is executed by typing "ab" or by selecting it in the menu from "Extensions" tab.

Unassigned floating labels by PINE2SPARKY



Most probable assignments



"ab" command

Type minimal threshold and press "Assign"

Enter the minimal threshold from 0.0 to 1.0. The default is 0.9. Then click the "Assign" button and wait until the assignment process is finished. You will see assigned peaks and selected floating labels selected. The floating labels can easily be removed by pressing the "Delete" key; the result is shown on the left.

PINE-SPARKY on Ubiquitin Protein

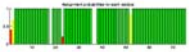
Collected Experiment Sets

2D Experiments	N-HSQC, C-HSQC
3D Experiments	CBCA(CO)NH, CC(CO)NH, HBHA(CO)NH, HC(CO)NH, HCCH-TOCSY, ^{13}C-NOESY, ^{15}N-NOESY

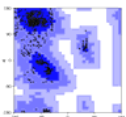
Protocols

(1) PROCESSING: NMRpipe, NMRdraw[3]
(2) PEAK PICKING: NMRdraw except NOESYs
ATNOS for ^{13}C-NOESY, ^{15}N-NOESY
(3) PINE-NMR: guessing probable assignments
(4) PINE2SPARKY converter
(5) SPARKY extensions:
(a) "Assign the best by PINE" with 0.9 threshold.
(b) Assign atom-by-atom with "PINE Graph Assigner."
(c) Assign atoms showing large deviations peak-by-peak with "PINE Assigner."
(6) Running FAPSY[4](Automated Protein Structure Suite; http://waltz.yonsei.ac.kr/asps)
(a) TALOS[5] for calculating angle constrains.
(b) CYANA[6] for calculating protein structures.
(7) MOLMOL[7] to visualize and calculate RMSDs

PINE result



TALOS result

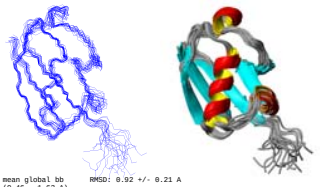


Results

How "Assign the Best by PINE" with 0.9 Threshold Worked?

	Ideal assignments(%)	Automatically Assigned(%)	Manually Assigned(%)	Auto/Manual(%)
All assignments	[illegible]	[illegible]	[illegible]	[illegible]
Backbone assignments	[illegible]	[illegible]	[illegible]	[illegible]
Sidechain assignments	[illegible]	[illegible]	[illegible]	[illegible]

Ensemble of 20 Structures Generated in *One Hour* by the Above Protocol



mean global bb (0.46.. 1.63 A) RMSD: 0.92 +/- 0.21 A
mean global heavy (1.22.. 2.33 A) RMSD: 1.74 +/- 0.19 A

Conclusions

(1) PINE-SPARKY is a software package for the rapid and efficient visualization and refinement of PINE-NMR results. PINE-SPARKY is based on the very popular SPARKY package.

(2) PINE-SPARKY consists of a PINE2SPARKY converter and three SPARKY extensions.

(3) The PINE2SPARKY converter takes the output from a PINE-NMR analysis and transforms it into SPARKY input. Possible alternative assignments given by PINE are incorporated into SPARKY save files by the PINE2SPARKY converter.

(4) The visualized labels in SPARKY's spectrum view are handled by three SPARKY extensions- PINE Assign (peak-by-peak assignment), PINE Graph Assigner(atom-by-atom assignment), and Assign the Best by PINE (choose highest probability).

(5) We tested the PINE-SPARKY package with PINE-NMR results from the protein ubiquitin; and it took only one hour from raw data to a structure.

References
(1) Bahrami A, Assadi A, Markley JL, and Eghbalnia H. (2009) *PLoS Comp Biol*, in press
(2) TD Goddard and DG Kneller, SPARKY 3, University of California, San Francisco
(3) Delaglio F, Grzesiek SW, Vuister G, Zhu G, Pfeifer J, and Bax A. (1995) *J Biomol NMR*
(4) Lee W, Jung J, and Lee W. (2007) Yonsei University, Seoul
(5) Cornilescu G, Delaglio F, and Bax A. (1999) *J Biomol NMR*
(6) Guntert P. (2004) *Methods Mol Biol*
(7) Koradi R. Billeter M, Wüthrich K. (1996) *J Mol Graph*

Acknowledgement
This work was funded by Protein Structure Initiative Grant U54 GM074901 (J.L. Markley, PI, G.N. Phillips, and B.G. Fox, Co-Investigators) and P41 RR02301 (NMRFAM).

Contact Information
Woonghee Lee
whlee@nmrfam.wisc.edu
NMRFAM (National Magnetic Resonance Facility at Madison)
http://pine.nmrfam.wisc.edu/PINE-SPARKY

CESG is supported by the National Institute of General Medical Sciences through the Protein Structure Initiative NIH/NIGMS grant numbers 1 U54 GM074901 (JLM; 07/01/05-06/30/10) and P50 GM064598 (JLM; 01/01/02-08/31/05).

NMRFAM is supported by the National Center for Research Resources P41 RR02301 (JLM).