

# Center for Eukaryotic Structural Genomics

## Technology Dissemination Report

<b>CESG Tech Report No.</b>	014
<b>Title</b>	PINE (Probabilistic Inference Network of Evidence)
<b>Research Unit</b>	NMR Spectroscopy
<b>Authors</b>	Bahrami, A., Eghbalnia, H.R., Wang, L., and Markley, J.L.
<b>Primary Contact</b>	<a href="http://miranda.nmrfam.wisc.edu/PINE/">http://miranda.nmrfam.wisc.edu/PINE/</a>

**PINE Server v.1.0**

PINE accepts, as input, the sequence of the protein plus peak lists (or spin systems) from a variety of NMR experiments and offers automatic **backbone** and **sidechain assignments**, detection and automated correction of potential **referencing problems** or **inconsistent assignments** and **secondary structure determination**.

- Acceptable formats for input data are
  - Amino acid sequences and peak lists in **Sparky format**, **XEASY format**, or **PINE simple format**. Sample files and brief instructions are available [here](#).
  - If you have already analyzed the spin systems, you can upload your **spin system file in correct format** in the spin system field at the end of the webpage. In that case please do NOT upload peak lists, however you still need to upload the **sequence file in correct format**.
- PINE also offers assignments for proteins that have been **selectively labeled** or have prior **partial assignments**.
- Instruction for PINE output files can also be found [here](#).
- Assignment **results**, including secondary structure identification, are returned via **email**.
- Depending on the size of the protein, the quality of input data, and number of current jobs, **run times vary** from one hour to one day.

A service offered by the National Magnetic Resonance Facility at Madison. Updated JULY 2007.

**Upload files to PINE Server**

Please click here for **INSTRUCTIONS**. Your results will be returned to the below email address.

- This form supports upload of **up to 15 files** with corresponding experiment names. You can leave **unnneeded** file names **blank**.
- A short instruction page is available [here](#).

Upload files to PINE Server

Name\*

E-mail\*

Sequence [Sequence File in ONE or THREE letters](#)

sequence

2D-Experiments [Peak Picked Data](#)

HSQC (N15)

C13-HSQC

3D-Experiments [Peak Picked Data](#)

HNCO

CBCA(CO)/NH

HNCACB

### Summary

PINE, which is available from a webserver at <http://miranda.nmrfam.wisc.edu/PINE/>, is the implementation of a multiple step probabilistic data analysis platform for protein NMR spectroscopy. As of 12/08 PINE has been used more than 1300 times by external Users. PINE incorporates the capabilities of separate probabilistic tools: PISTACHIO [1] (automated backbone and sidechain assignment), PECAN[2] (secondary structure determination), and LACS [3] (referencing offset and outlier detection). The input to PINE is the amino acid sequence and sets of peak lists generated from one or more of the standard types of protein NMR experiments; these can be either probabilistic (e.g., peak lists generated by HIFI-NMR) or traditional peak lists generated by popular NMR data analysis tools. PINE takes into account the interconnectedness of different stages of analysis. PINE begins with a set of local statistical potentials. It then proceeds iteratively until a stationary state for a consistent global similarity measure is achieved. The resulting software enables a seamless and robust integration of multiple steps in the NMR structure determination pipeline. PINE provides as output a probabilistic assignment of backbone and sidechain signals and the secondary structure of the protein. At the same time, it identifies, verifies, and if needed rectifies, problems related to referencing, assignment, or outlying data. PINE can make use of prior information supplied from selective labeling or spin system assignments derived independently by other means. The performance of PINE is much superior to that of the individual tools used sequentially.

#### Publications:

- [1] Eghbalnia, H.R., Bahrami, A., Wang, L., Assadi, A., and Markley, J.L. (2005) *J Biomol NMR* 32(3):219-33.
- [2] Eghbalnia, H.R., Wang, L., Bahrami, A., Assadi, A., and Markley, J.L. (2005) *J Biomol NMR* 32(1):71-81.
- [3] Wang, L., Eghbalnia, H.R., Bahrami, A., and Markley, J.L. (2005) *J Biomol NMR* 32(1):13-22.

<b>Acquiring the Technology</b>	Available from: <a href="http://miranda.nmrfam.wisc.edu/PINE/">http://miranda.nmrfam.wisc.edu/PINE/</a>
<b>Other Acknowledgements</b>	Partial support from NIH Grant P41 RR02301.

Center for Eukaryotic Structural Genomics (CESG), University of Wisconsin-Madison Biochemistry Department, 433 Babcock Drive, Madison, WI 53706-1549; phone: 608.263.2183; fax: 608.890.1942; email: [cesginfo@biochem.wisc.edu](mailto:cesginfo@biochem.wisc.edu); website: <http://www.uwstructuralgenomics.org>. This research funded by NIH / NIGMS Protein Structure Initiative grant U54 GM074901