Bioinformatics, 31(8), 2015, 1325–1327
doi: 10.1093/bioinformatics/btu830
Advance Access Publication Date: 12 December 2014
Applications Note

OXFORD

Structural bioinformatics

NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy

Woonghee Lee*, Marco Tonelli and John L. Markley*

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

*To whom correspondence should be addressed.

Associate Editor: Anna Tramontano

Received and revised on October 22, 2014; accepted on December 8, 2014

Abstract

Summary: SPARKY (Goddard and Kneller, SPARKY 3) remains the most popular software program for NMR data analysis, despite the fact that development of the package by its originators ceased in 2001. We have taken over the development of this package and describe NMRFAM-SPARKY, which implements new functions reflecting advances in the biomolecular NMR field. NMRFAM-SPARKY has been repackaged with current versions of Python and Tcl/Tk, which support new tools for NMR peak simulation and graphical assignment determination. These tools, along with chemical shift predictions from the PACSY database, greatly accelerate protein side chain assignments. NMRFAM-SPARKY supports automated data format interconversion for interfacing with a variety of web servers including, PECAN , PINE, TALOS-N, CS-Rosetta, SHIFTX2 and PONDEROSA-C/S.
Availability and implementation: The software package, along with binary and source codes, if desired, can be downloaded freely from http://pine.nmrfam.wisc.edu/download_packages.html. Instruction manuals and video tutorials can be found at http://www.nmrfam.wisc.edu/nmrfam-sparky-distribution.htm.
Contact: whlee@nmrfam.wisc.edu or markley@nmrfam.wisc.edu
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

SPARKY (https://www.cgl.ucsf.edu/home/sparky) remains the most popular computer program for NMR operations, such as peak-picking and peak assignment (Supplementary Fig. S1), despite that fact that its originators have not released a new version since 2001 (Goddard and Kneller, 2008, SPARKY 3). SPARKY supports user-defined enhancements, and we have used these to develop new tools in support of our packages for automated protein assignment and structure determination. The added features support (i) interfacing with servers offering new technologies, (ii) tools for data visualization and verification and (iii) new protocols for maximizing the efficiency of NMR data analysis. We have refined these SPARKY enhancements through their use in our annual workshops by participants with varying experience in protein NMR spectroscopy. We describe here the new package, which we have named 'NMRFAM-SPARKY'.

2 Implementation

We first replaced Python 2.4 by 2.7 and Tcl/Tk 8.4 by 8.6 to enable the implementation of features related to web services and user interfaces to platforms including Windows, Mac OS X, and Linux. We had to use Tcl/Tk 8.5 for the Mac OS X version, because the 8.6 version was incompatible with X11 libraries. Several modifications of the C++ and Python codes were needed to achieve stability. Once the new version compiled with Python and Tcl/Tk, we updated old functions and implemented new functions (Supplementary Table S1) found to be desirable from our workshops and research projects. New functions are located under the NMRFAM menu and are categorized by sub-menus identified by two-letter codes (Table 1).

Dummy Graph (two-letter-code: dg) offers a generalized version of the *Pine Graph Assigner* utility from PINE-SPARKY (Lee *et al.*, 2009) that does not require use of PINE (Bahrami *et al.*, 2009) or the

Table 1. New and updated NMRFAM-SPARKY functions

Sub-menu	Functions	Two-letter-code
Automated BB and SC assignment: PINE	>>> Automated BB and SC assignment: PINE	n1
	Export to PINE(BMRB) input file	ep
	ARECA list	ar
PINE visualization: PINE-SPARKY	>>> PINE validation: PINE-SPARKY	n2
	Run PINE2SPARKY converter	p2
	Pine assigner	ab
	Pine graph assigner	pp
	Assign the best by Pine	pr
	Select all floating labels	se
Superfast assignment with PACSY	>>> Superfast assignment on-the-fly referencing PACSY	n3
	Dummy graph	dg
	Transfer and simulate assignments	ta
	Untag _s	ut
	Center and Untag _s	cu
	>>> Structure based chemical shift predictor: SHIFTX2 from Wishart Lab.	E1
Automated 3D structure calculation with PONDEROSA	>>> Automated structure calculation: PONDEROSA	n4
	Run Ponderosa Client	cp
	Update Ponderosa	up
	Generate Distance Constraints for PONDEROSA	gd
	Cyana2Sparky format	cy
	XEASY, DYANA format	xe
	Extract phi-psi and accessible surface info from PDB with STRIDE	sr
Structural predictions	>>> Structural predictions	n5
	Export to PECAN and go PECAN webserver	n6
	3D structure prediction with CS-Rosetta (BMRB)	ce
	phi-psi prediction with TALOS-N (NIH)	tl
	Secondary structure prediction with PSIPRED (UCL)	PP
Utilities	Backbone peak picking by APES	ae
	PINE sequence formatting	fp
	Easy pipe2ucsf	Pu
	Easy bruk2ucsf	Bu

PINE2SPARKY converter. Sequence information and assignment situations are illustrated graphically in real time so as to notify users of any missing or wrong assignments that need to be fixed. Another useful function is *Transfer and Simulate Assignments* (two-letter-code: ta). We used statistics from the PACSY database Lee *et al.*, 2012), a relational database management system incorporating PDB (Berman *et al.*, 2007), BMRB (Ulrich *et al.*, 2008), SCOP (Murzin *et al.*, 1995), STRIDE (Frishman and Argos, 1995) and MolProbity (Chen *et al.*, 2010), to populate this Python module, which simulates resonance frequencies dynamically on the basis of conditions such as amino acid and atom type, preceding and following residues, secondary structure, pH and temperature. This tool enables a new *predict-and-confirm* approach to resonance assignments that is much faster than the traditional *pick-and-assign* method because the user no longer needs to refer to a table of BMRB-derived chemical shift. Furthermore, if a 3D structure is available, it is possible to determine assignments by using SHIFTX2 prediction to simulate the possible assignments to be checked. NMRFAM-SPARKY contains a structure predictions menu that includes PECAN (Eghbalnia *et al.*, 2005), TALOS-N (Shen and Bax, 2013) and CS-Rosetta (Shen *et al.*, 2009). These tools can be used to predict the 3D structure so that side chain and NOE assignments can be refined through combined use of CS-Rosetta and SHIFTX2 (Han *et al.*, 2011). In favorable cases, the user can complete the chemical shift assignments and then use the NMRFAM-SPARKY interface to PONDEROSA-C/S (Lee *et al.*, 2014) to carry out an NOE-based structure determination in seamless fashion.

NMRFAM-SPARKY contains updated chemical shift statistics which greatly improve the accuracy and reliability of predictions. We expanded the number of resonance types (amino acid type plus atom type) from 273 to 276 (Supplementary Table S2A). By basing the statistics on the current BMRB chemical shift archive, we achieved statistical significance for 27 of these resonance types, which were not considered to be statistically meaningful because they previously were based on fewer than 30 examples (Supplementary Table S2B). The average chemical shifts of three of the resonance types changed by more than one SD (Supplementary Table S2C). The SDs for chemical shifts of 28 resonance types were refined by ≥30% (Supplementary Table S2D), whereas those of an equivalent number of resonances types were broadened by ≥30% (Table Supplementary S2E).

3 Results and Conclusions

Subsequent to announcing the availability of NMRFAM-SPARKY at workshops, at scientific meetings, and on the NMRFAM website, the package was downloaded 155 times from sites that provided unique e-mail addresses over the period May 21, 2014 to September 17, 2014. We expect that use of NMRFAM-SPARKY will increase the success rate of automated assignment runs on the PINE server and automated structure determinations on the PONDEROSA-C/S server, because our analysis shows that most failures stem from incorrectly formatted input. For similar reasons, data input through NMRFAM-SPARKY should improve the reliability and successful use of other programs or web servers, such as TALOS-N and CS-Rosetta.

Acknowledgements

We are grateful to Dr Thomas Goddard for making the source code availability for SPARKY3. We also thank the participants in the NMRFAM workshops who inspired us to develop many of the new functions described here.

Funding

United States National Institutes of Health NIH [P41GM103399].

Conflict of Interest: none declared.

References

Bahrami,A. *et al.* (2009) Probabilistic interaction network of evidence algorithm and its application to complete labeling of peak lists from protein NMR spectroscopy. *PLoS Comput Biol*, **5**, e1000307.

Bartels,C. *et al.* (1995) The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J. Biomol. NMR*, **6**, 1–10.

Berman,H. *et al.* (2007) The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res.,* **35**, D301–D303.

Chen,V.B. *et al.* (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 12–21.

Eghbalnia,H.R. *et al.* (2005) Protein energetic conformational analysis from NMR chemical shifts (PECAN) and its use in determining secondary structural elements. *J. Biomol. NMR*, **32**, 71–81.

Frishman,D. and Argos,P. (1995) Knowledge-based protein secondary structure assignment. *Proteins*, **23**, 566–579.

Goddard,T.D. and Kneller,D.G. (2008) *SPARKY 3*. University of California, San Francisco, CA.

Han,B. *et al.* (2011) SHIFTX2: significantly improved protein chemical shift prediction. *J. Biomol. NMR*, **50**, 43–57.

Johnson,B.A. and Blevins,R.A. (1994) NMR View: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR*, **4**, 603–614.

Lee,W. *et al.* (2012) PACSY, a relational database management system for protein structure and chemical shift analysis. *J. Biomol. NMR*, **54**, 169–179.

Lee,W. *et al.* (2009) PINE-SPARKY: graphical interface for evaluating automated probabilistic peak assignments in protein NMR spectroscopy. *Bioinformatics*, **25**, 2085–2087.

Lee,W. *et al.* (2014) PONDEROSA-C/S: client-server based software package for automated protein 3D structure determination. *J. Biomol. NMR.*, **60**, 73–75.

McGuffin,L.J. *et al.* (2000) The PSIPRED protein structure prediction server. *Bioinformatics*, **16**, 404–405.

Murzin,A.G. *et al.* (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, **247**, 536–540.

Shen,Y. *et al.* (2009) De novo protein structure generation from incomplete chemical shift assignments. *J. Biomol. NMR*, **43**, 63–78.

Shen,Y. and Bax,A. (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR*, **56**, 227–241.

Ulrich,E.L. *et al.* (2008) BioMagResBank. *Nucleic Acids Res.*, **36**, D402–D408.