GenApp-Generated Science Gateway for NMR Data Analysis and Determination of Structural Ensembles of Biological Macromolecules

Matthew R. Casertano
Montgomery Blair High School
Silver Spring, MD, USA
matthewryancasertano@gmail.com

Emre H. Brookes
Dept. Chemistry & Biochemistry
University of Montana
Missoula, MT, USA
emre.brookes@mso.umt.edu

Raquel Gama Lima Costa
Chemical Physics Program and
Dept. Chemistry & Biochemistry
University of Maryland
College Park, MD, USA
rgama@umd.edu

David Fushman
Dept. Chemistry & Biochemistry
University of Maryland
College Park, MD, USA
fushman@umd.edu

ABSTRACT

Knowledge of the structure and motions of proteins and nucleic acids is required for understanding how these molecular machines work and for development of therapeutics to combat diseases. Painting an adequate and reliable portrait of these inherently dynamic biological macromolecules presents a significant challenge due to the need to integrate experimental data from various sources and to treat these systems as structural ensembles because the experimental data are a convoluted average of contributions from multiple conformations [1]. In particular, Nuclear Magnetic Resonance (NMR) spectroscopy is a powerful versatile experimental technique that provides vital information on molecular structure and dynamics. Several approaches and software packages have been developed to tackle the complicated analysis of experimental NMR data in order to extract important structural information from underdetermined structural data. However, many of them are offline packages or command-line applications that require users to set up the run time environment and also to possess certain programming skills, which inevitably limits accessibility of this software to a broad scientific community. To address these current limitations in NMR data analysis and applications, we developed a science gateway designed for the NMR/structural biology community. The NMRSuite gateway enables comprehensive analysis of distance and orientation-dependent effects caused by paramagnetic tagging, including pseudo-contact shifts (PCS), residual dipolar couplings (RDC), and paramagnetic relaxation enhancement (PRE). The analysis is performed both for a single-structure model and for a conformational ensemble. The latter option supports two ensemble treatments: the Sparse Ensemble Selection (SES) method and the Maximum Entropy (MaxEnt) method [1]. Both methods are integrated with an additional module, Predict, which enables ab initio prediction of the relevant experimental NMR data for a given molecular structure. To make these analytical tools broadly available, we used the GenApp framework for scientific gateways [2] to transform the original software packages into a science gateway that provides advanced computational functionalities, streamlines cloud-based data input, output, and storage, and offers interactive 2D and 3D plotting and visualizations. All the aforementioned modules were originally written as standalone programs in Matlab or Java, while NMRSuite is written almost entirely in a single language, Python, to simplify future development and maintenance. All code for NMRSuite is available in a GitHub public repository, allowing anyone to read it and/or suggest improvements, as well as offering a valuable reference for others who wish to create a GenApp gateway hosted, as NMRSuite, on Jetstream or other publicly available resources. This gateway will assist researchers by providing simple, yet customizable, tools for analyzing the structure, dynamics, and function of biological macromolecules on a single easily accessible website. NMRSuite modules have been successfully used in research and in teaching Biomolecular NMR courses at the University of Maryland. As a next step, we plan to integrate NMRSuite with the SASSIE-web gateway [3] to enable joint analysis of NMR and small-angle scattering data and provide access to more analysis tools under a single website. We hope that our work will inspire other researchers to deploy their work with a scientific gateway.

Keywords—Science Gateway, GenApp, Generalized Application Framework, Biomolecular NMR, Conformational Ensemble, Structural Biology, Biophysics

REFERENCES

[1] K. Berlin, C. A. Castañeda, D. Schneidman-Duhovny, A. Sali, A. Nava-Tudela, and D. Fushman, "Recovering a Representative Conformational Ensemble from Underdetermined Macromolecular Structural Data", J. Am. Chem. Soc. (2013) 135, pp. 16595-16609

[2] A. Savelyev, E. Brookes, "GenApp: Extensible tool for rapid generation of web and native GUI applications", Future Generation Computer Systems (2017). https://doi.org/10.1016/j.future.2017.09.069. https://genapp.rocks

[3] J.E. Curtis, S. Raghunandan, H. Nanda, S. Krueger, "SASSIE: a program to study intrinsically disordered biological molecules and macromolecular ensembles using experimental scattering restraints", Comput. Phys. Commun. (2012) 183, 382–389. https://sassie-web.chem.utk.edu/sassie2/