ROTDIF-web and ALTENS: GenApp-based Science Gateways for Biomolecular Nuclear Magnetic Resonance (NMR) Data Analysis and Structure Modeling

Yuexi Chen
Dept. Chemistry & Biochemistry and Computer Science
University of Maryland
College Park. MD, USA
ychen151@terpmail.umd.edu

Cheol Jeong
Dept. Chemistry & Biochemistry
University of Montana
Missoula, MT, USA
cheol.jeong@gmail.com

Alexey Savelyev
Dept Chemistry & Biochemistry
University of Montana
Missoula, MT, USA
alexsav.science@gmail.com

Susan Krueger
NIST Center for Neutron Research
National Institute of Standards and Technology
Gaithersburg, MD, USA
susan.krueger@nist.gov

Joseph E. Curtis
NIST Center for Neutron Research
National Institute of Standards and Technology
Gaithersburg, MD, USA
joseph.curtis@nist.gov

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

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

***Abstract*—Proteins and nucleic acids participate in essentially every biochemical process in living organisms, and the elucidation of their structure and motions is essential for our understanding how these molecular machines perform their function. Nuclear Magnetic Resonance (NMR) spectroscopy is a powerful versatile technique that provides critical information on the molecular structure and dynamics. Spin-relaxation data are used to determine the overall rotational diffusion and local motions of biological macromolecules, while residual dipolar couplings (RDCs) reveal local and long-range structural architecture of these molecules and their complexes. This information allows researchers to refine structures of proteins and nucleic acids and provides restraints for molecular docking. Several software packages have been developed by NMR researchers in order to tackle the complicated experimental data analysis and structure modeling. 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. Here we present new science gateways designed for the NMR/structural biology community that address these current limitations in NMR data analysis. Using the GenApp framework for scientific gateways (https://genapp.rocks), we successfully transformed ROTDIF and ALTENS, two offline packages for bio-NMR data analysis, into science gateways that provide advanced computational functionalities, cloud-based data management, and interactive 2D and 3D plotting and visualizations. Furthermore, these gateways are integrated with molecular structure visualization tools (Jmol) and with gateways/engines (SASSIE-web) capable of generating huge computer-simulated structural ensembles of proteins and nucleic acids. This enables researchers to seamlessly incorporate conformational ensembles into the analysis in order to adequately take into account structural heterogeneity and dynamic nature of biological macromolecules. ROTDIF-web offers a versatile set of integrated modules/tools for determining and predicting molecular rotational diffusion tensors and model-free characterization of bond dynamics in biomacromolecules and for docking of molecular complexes driven by the information extracted from NMR relaxation data. ALTENS allows characterization of the molecular alignment under anisotropic conditions, which enables researchers to obtain accurate local and long-range bond-vector restraints for refining 3-D structures of macromolecules and their complexes. We will describe our experience bringing our programs into GenApp and illustrate the use of these gateways for specific examples of protein systems of high biological significance. We expect these gateways to be useful to structural biologists and biophysicists as well as NMR community and to stimulate other researchers to share their scientific software in a similar way.**

Keywords—GenApp, Generalized Application Framework, Biomolecular NMR, Structural Biology, Biophysics

I. Introduction

Proteins participate in essentially every biochemical process in living organisms, and the elucidation of protein structures and motions is essential for our understanding how these molecular machines perform their function. The problem is compounded by the inherent flexibility of proteins as macromolecules. Instead of maintaining a rigid and static structure, most proteins adapt their conformations to interact with other molecules. The biomolecular interactions lay the foundation of vital biological processes like oxygen transport, muscle contraction, immune response, cell cycle, apoptosis, etc. With the timescales of motions ranging from picoseconds

to milliseconds and longer, the study of protein dynamics demands powerful experimental techniques and adequate software for data analysis and modeling. Nuclear Magnetic Resonance (NMR) combined with spin-relaxation measurements is a powerful, state-of-the-art technique to characterize local and global dynamics of proteins and nucleic acids [1].

In the past decades, we have developed several computational approaches and software that enable analysis and modeling of NMR data, including spin-relaxation rates and residual dipolar couplings (RDCs) caused by molecular alignment [1-7]. In particular, ROTDIF is a powerful method for determining and predicting molecular rotational diffusion tensors from NMR relaxation data [2]. Its latest version uses a new multi-start convex algorithm, which significantly improved the accuracy and speed [3]. Furthermore, by integrating the ab initio prediction, diffusion-driven molecular docking [4,5], and modified DYNAMICS program [6] that includes Akaike information criterion for model selection, this package provides versatile tools for analysis of the overall and local dynamics in proteins and nucleic acids. Another package, ALTENS, allows characterization of the molecular alignment under anisotropic conditions, which enables accurate local and long-range bond-vector restraints for refining 3-D structures of macromolecules and their complexes [1,7]. However, as stand-alone packages, ROTDIF (written in Java) and ALTENS (written in Matlab) offer very limited accessibility for a broader scientific community. To address this deficiency, we used GenApp-based technology [8] to transform these software packages into both NMR based science gateways containing ROTDIF-web and ALTENS, and as modules to be incorporated into the SASSIE-web gateway [9]. This work (A) made these computational tools broadly available to researchers, (B) maintained all the features of the original packages, and (C) enhanced these programs with additional functionalities and capabilities offered by the web access (and unavailable to stand-alone programs), such as cloud-based data management and interactive data visualizations. Furthermore, integration with the existing SASSIE-web science gateway engines capable of generating/modeling huge structural ensembles of biological macromolecules, allows adequate characterization of these intrinsically dynamic molecular systems and enables integration of NMR data with data from small-angle scattering (SAS) and other biophysical methods. In this paper we present these novel gateways and illustrate their features and functionalities.

II. Results

A. Utilizing the GenApp Framework

GenApp (Generalized Application Framework) is a user-friendly framework developed by Brookes et al. to build science gateways for non-computer science (CS) researchers[8]. GenApp features a separation between developers' framework and researchers' code. To add a program to GenApp, the researcher only needs to produce a straightforward GenApp definition file and wrap or modify the program to accept input and produce output as defined to create a science gateway. Meanwhile, it is the CS experts who are in charge of maintaining and updating capabilities of the GenApp framework. So far, besides supporting any programming language or wrapper capable of handling JSON, including Python, Java, Perl, and C/C++, GenApp also enhances users' experience by enabling online interactive data visualizations and analytics.

To bring the ROTDIF programs into GenApp, we wrote definition files for each program and modified the programs' Java code to accept input and produce output as we specified in the definition files. Subsequently, there were multiple rounds of adjustments to the definition file and Java code as we converged on a final design. For ALTENS, we first chose to convert the Matlab code into Python. We could have wrapped this code directly, but to ease portability of the program, we decided against dependence on a commercial package. Subsequently, we followed the same process as for ROTDIF. The ease of adding new capabilities, such as the molecular viewer, to enhance our output was appreciated and improved the programs.

B. ROTDIF-web Gateway

ROTDIF-web (http://rotdif.genapp.rocks/rotdif/) consists of three major modules: ROTDIF & Dynamics, ELM, and ELMDOCK.

The ROTDIF & Dynamics module is designed for determining both the overall rotational diffusion (tumbling) and the local bond dynamics of a given biological macromolecule. It takes as user input experimental NMR spin-relaxation data (for 13C or 15N or combination thereof) measured at a single or multiple magnetic fields, and atom coordinates (in the Protein Data Bank format). The analysis is performed using isotropic, axially symmetric and fully anisotropic models for the diffusion tensor. To speed up convergence, the results for simpler models are used as initial guess for more complicated tensor models. This module also integrates our program, DYNAMICS [6], to perform a model-free analysis of local bond dynamics. The gateway features interactive 2D and 3D output graphics (using Plotly and Bokeh visualization tools) that enable immediate interactive inspection and analysis of the results of computation (See Fig. 1).

ELM module performs ab initio prediction of the rotational diffusion tensor by generating solvent-accessible surface of a molecule (given the atom coordinates), and uses ellipsoid approximation of the resulting shape in conjunction with principal component analysis and Perrin equations [4] to compute the complete rotational diffusion tensor (See Fig.2).

ELMDOCK performs diffusion-tensor driven rigid-body docking of macromolecules by integrating ROTDIF output

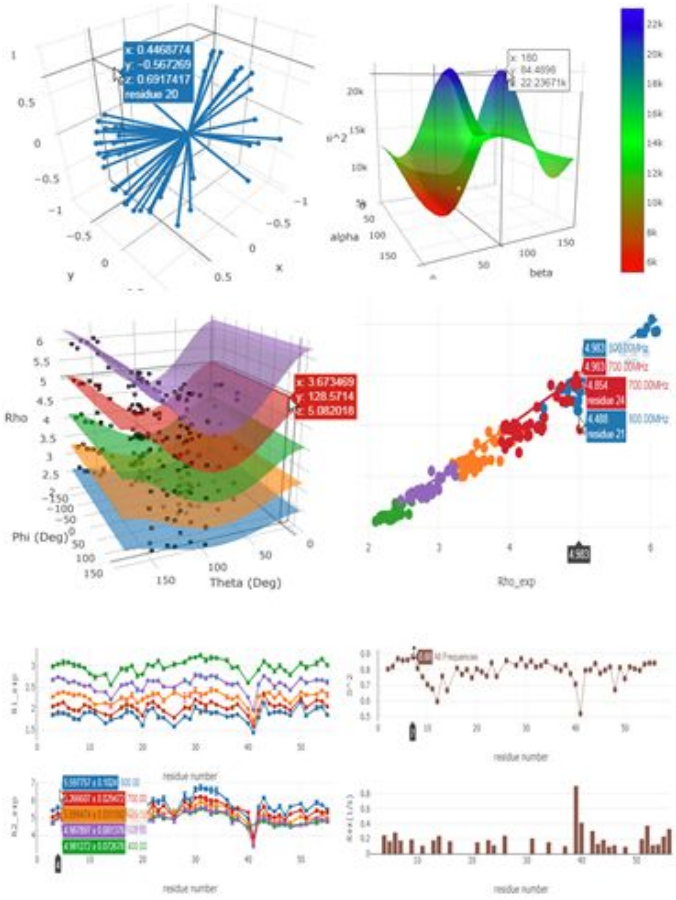

Fig. 1. Representative examples of various interactive screen outputs of the ROTDIF & Dynamics module analysis of 5-field NMR data for protein GB3.

```
Parsing pdb file....done.
Predicting tensor...done.

Dx = 1.58899 *(10^7) 1/s
Dy = 1.59822 *(10^7) 1/s
Dz = 2.26579 *(10^7) 1/s
alpha = 120.32 deg
beta = 123.19 deg
gamma = 131.25 deg
TAUc = 9.16926 ns
anisotropy = 1.42180
rhombicity = 0.02062

====Diffusion Tensor====
    [ 1.71101 -0.20427  0.15864]
D = [-0.20427  1.94654 -0.26364] *(10^7) 1/s
    [ 0.15864 -0.26364  1.79546]
```

Fig. 2. Output of the ELM module predicted rotational diffusion tensor for the docked ubiquitin dimer shown in Fig.3. The tensor prediction tool was used by ELMDOCK to guide molecular docking and validate the outcome.

and predicted rotational diffusion tensors returned by ELM with an alignment algorithm based on eigendecomposition, to create docked structures of proteins and protein complexes [4,5] (Fig. 2). In addition to output of generated atom coordinates at downloadable text files and Python macros, ROTDIF-web also allows users to immediately visualize the resulting 3D structures using biomolecular visualization tool Jmol [10] which provides multiple options for molecular image manipulations (Fig. 3).

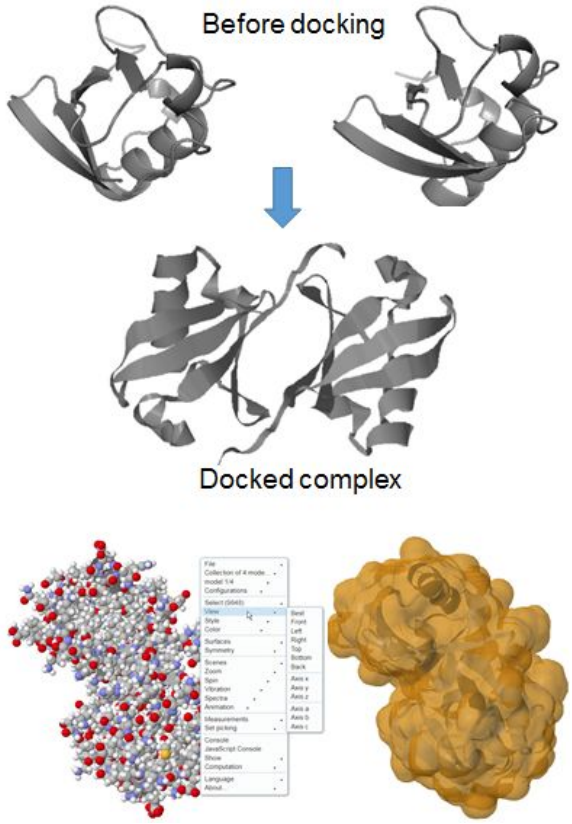

Fig. 3. Example of the ELMDOCK output. Top and middle rows: two molecules of a protein ubiquitin before and after docking. Bottom: the integration with Jmol enables immediate interactive visualization and inspection of the docked complex using various options of molecular representation and coloring.

C. ALTENS Gateway

This gateway provides tools for the analysis of experimental residual dipolar couplings (RDCs) for single or multiple sets of bond vectors (e.g. H-N, Ha-Ca, N-CO etc) separately or simultaneously, and for a single macromolecular structure or an ensemble of structures obtained experimentally or computer-generated using MD simulations. For the gateway, the ALTENS code was written *de novo* in Python, and new computational and visualization functionalities were added. The significant added functionalities include handling experimental data from multiple nuclei and the ability to perform the analysis for conformational ensembles.

Importantly, ALTENS is integrated as a module into SASSIE-web (https://sassie-web.chem.utk.edu/sassie2/), to enable seamless incorporation into the analysis of structural ensembles generated by SASSIE, in order to take into account the dynamic nature of biological macromolecules. ALTENS features interactive 2D graphics (using Bokeh interactive

visualization library) that allow users to directly inspect and analyze the program output (See Fig. 4).

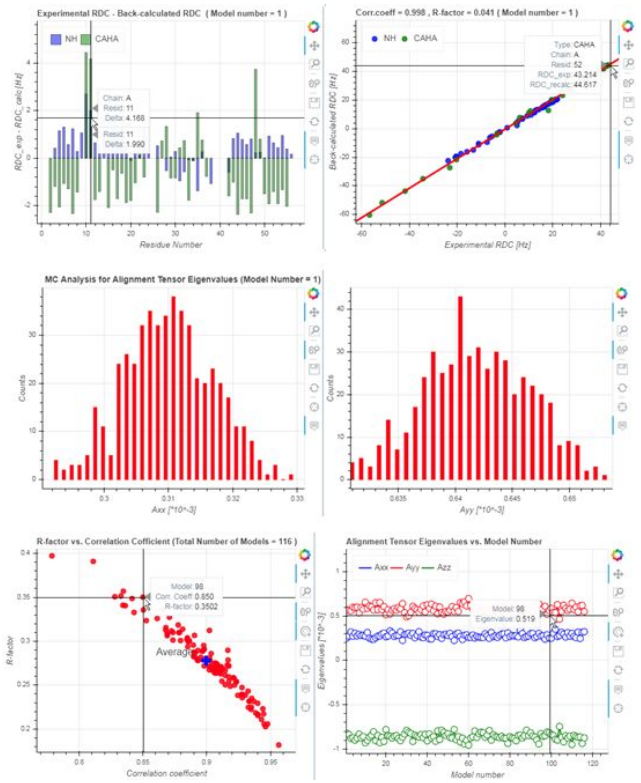

Fig. 4. Examples of interactive ALTENS output. Top row: the agreement between experimental and back-calculated RDC data for H-N and Ha-Ca bonds (analyzed together) in protein GB3. Middle row: analysis of errors in the derived alignment tensor values using Monte Carlo generated synthetic data. Bottom row: the results of RDC data analysis for ensemble of 116 structures of protein ubiquitin.

D. User Base

The science gateways described here address the needs of bio-NMR community for easily accessible web-based tools. They will also be of interest to structural biologists and biophysicists and will help "lower the barrier" for using NMR techniques in their research. Specifically, the integration into SASSIE-web connects the NMR gateways with SAS data and research community. We will expand our user base through publications, presentations and workshops at domain specific conferences. Furthermore, the described ROTDIF-web and ALTENS gateways can also be used as teaching tools: they were successfully utilized by graduate students in a NMR course given at University of Maryland.

E. Future Work

We will bring additional software packages for NMR data analysis and modeling into the GenApp framework. These include tools for analysis of paramagnetic effects (relaxation enhancement, pseudocontact shifts, etc) and for determining conformational ensembles based on experimental data from NMR, small-angle scattering, and other biophysical techniques. Finally, we plan to create a unified NMR-suite science gateway that combines the tools described in this paper with the additional software packages.

III. Conclusions

Through GenApp integration, the ROTDIF-web and ALTENS science gateways have made accessible to a broad range of researchers the software packages for analysis of NMR data that previously were available only as portable and command-line versions. Researchers will also benefit from the added online interactive data visualizations and analytics. We believe others in the structural biology and biophysics community will be able to use these new tools to advance their research capabilities and enhance research experience. We also believe that our GenApp-based ROTDIF-web and ALTENS gateways set an example that will encourage other research groups to share their scientific computer programs in a similar way.

Acknowledgment

Supported by NSF grants OAC-1739549 (to D.F.), OAC-1740087 (to E.H.B.), and by Science Gateways Community Institute summer internship to Y.C. This work benefited from CCP-SAS software developed through a joint EPSRC (EP/K039121/1) and NSF (CHE-1265821) grant. Certain commercial equipment, instruments, materials, suppliers, or software are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

References

[1] D. Fushman, R. Varadan, M. Assfalg, O.Walker, Progress in NMR Spectroscopy, (2004) 44, 189-214

[2] O. Walker, R. Varadan, D. Fushman, J. Magn. Reson. (2004) 168, 336-345

[3] K. Berlin, A. Longhini, T. K. Dayie, D. Fushman, J Biomol NMR (2013) 57, 333-352

[4] Y. Ryabov, C. Geraghty, A.Varshney, D. Fushman, J. Am. Chem. Soc. (2006) 128, 15432-15444; K. Berlin, D. P. O'Leary, D. Fushman, Proteins (2011) 79, 2268-2281.

[5] Y. Ryabov, D. Fushman, J. Am. Chem. Soc. (2007) 129, 7894-7902.

[6] D. Fushman, in Protein NMR Techniques, Third Edition, Eds. A. Shekhtman, D. S. Burz; Methods in Molecular Biology, 2012, Volume 831, 485-511, Springer Science

[7] R. Varadan, O. Walker, C. Pickart, D. Fushman, J. Mol. Biol. (2002) 324, 637-647; K. Berlin, D. P. O'Leary, D. Fushman, J. Am. Chem. Soc. (2010) 132, 8961-8972

[8] A. Savelyev, E. Brookes. 2017. 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

[9] 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.

[10] R. M. Hanson, J. Prilusky, Z. Renjian, T. Nakane, J. L. Sussman. Israel Journal of Chemistry (2013) 53, 207-216 . doi:10.1002/ijch.201300024