nature portfolio

Peer Review File

NMR Data Processing, Visualization, Analysis and Structure Calculation with NMRFx

Corresponding Author: Dr Bruce Johnson

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I have to say that bio nmr is not my field, I am from a small molecules background. In that sense, my comments are not the definitive opinion, more an outsider's view. I think the manuscript is well written and the software seems useful. What I would ask for is a clearer distinction if things apply for small molecules or macro molecules. I understand that there are some small molecule aspects as well (e. g. the introduction gives "Assignment of the natural product taccalonolide E" as a use case), but later on it is not always clear what things refer to. For example, 2.10 Prediciton - is that for macro molecules only, or for small molecules, or both? So overall, a clear statement of the application domain would be good. It's probably obvious for the authors, but not necessarily for the reader.

Reviewer #2

(Remarks to the Author)
In this paper, the authors present a software package equipped with a wide variety of functions for NMR measurements, reaching a high level of completeness.

Modern NMR spectroscopy is already a highly established, state-of-the-art methodology for signal analysis and structural modeling of biomolecules such as proteins and nucleic acids (DNA, RNA). By leveraging isotope labeling, diverse pulse sequence developments, and magnetic alignment techniques such as RDC and PCS, as well as long-range restraints from PCS (pseudo-contact shifts) and PRE (paramagnetic relaxation effects), NMR provides an unusually rich variety of structural information compared to other spectroscopic methods. At the same time, objective measurement, analysis, and interpretation require deep theoretical background knowledge—including organic chemistry, elementary quantum mechanics, and thermodynamics—as well as computational sciences such as mathematical optimization, molecular dynamics, statistics, and information theory.

Unlike X-ray crystallography or cryo-EM, NMR structural modeling proceeds through intermediate steps of signal assignment and structural information extraction, followed by computational model construction. Whereas X-ray or cryo-EM directly provide structural maps, NMR inherently leaves more room for human interpretation. Depending on the case, analyses may stop at signal assignment for ligand interaction or dynamics studies, or proceed directly to full 3D structural modeling.

For these reasons, full automation of NMR analysis systems is realistically very difficult, despite many attempts. In practice, users require not only semi-automated analysis but also flexible functions that allow visual verification of data.

The software includes data processing functions (1D, 2D, 3D, and 4D FFT), even for the highly labor-intensive NUS (Non-Uniform Sampling) method, implemented within the same package. This allows users to begin with standard NMR data sets (FIDs and measurement parameters), then seamlessly analyze spectral data sets for all NMR analyses. Importantly, it also allows users to easily return to the processing stage if adjustments are required—for example, to correct poor phasing, slight offset mismatches in individual dimensions, or inverted data increments in 3D spectra acquired using States-TPPI or Echo/Anti-Echo methods. These issues frequently arise even in routine processing.

Moreover, the system enables automatic NMR signal detection (or import of peak tables generated by external software such

as DEEP-Picker), automatic signal assignment, and consistency analysis between experimentally observed data and computationally modeled NMR structural coordinates using RDCs.

A particularly noteworthy feature is that the software also implements effective functions for RNA NMR analysis, which has long been a challenge in structural biology.

Although recent advances in deep learning have improved methods for predicting chemical shifts from biomolecular structural coordinates, this remains an extremely difficult task. Nonetheless, the approach reported in this paper for predicting RNA NMR chemical shifts—through sequence dependency and statistical refinements—achieves good predictive performance using strategies distinct from those applied to proteins.

The tool developed by this group is especially valuable in that it can visually represent sequence-specific assignment information, making it easier to identify problems, while cleverly avoiding a fully “black-box” design. This enables safer and more reliable RNA NMR data interpretation and structural modeling, making the tool highly practical for signal assignment and model building.

Another strength of this software is its ease of installation and wide compatibility across operating systems (Windows, macOS, Linux).

The use of Git for recording analysis histories is also highly effective, enabling review of workflows and improving efficiency.

Although CCP-NMR is a well-known comparable software, the present tool is more refined, particularly in its powerful support for RNA NMR signal assignment and structural analysis in solution—a feature of great current interest.

While artificial intelligence systems such as AlphaFold2 have succeeded in accurately predicting protein structures from sequence data, their training on crystal structures limits them to predicting only the most stable conformations. They provide no information on dynamic states or folding stability, and structural biologists should pay greater attention to this limitation. NMR remains the only technique capable of directly providing atomic-level information on motion and stability. This software, through its diverse functions—signal linearity, chemical shift prediction, and RDC consistency—can contribute to identifying data suggestive of dynamics.

In the future, many AI-assisted NMR data analysis tools will surely appear. However, the fact that this software achieves robust automation without turning the crucial task of signal assignment into a complete black box, while still allowing easy visual confirmation, is a highly commendable strength.

The software pursues modern usability and stability, supports a wide range of formats, and is filled with functions designed to assist many users. It can also be expected to flexibly incorporate future technical advances.

Making such a large-scale software, which must have required enormous effort and time, freely available to the academic community represents a highly significant contribution to the global NMR field.

For these reasons, I believe this paper deserves acceptance in the relevant international journal.

Minor suggestions for revision:

ShiftX+ blends the older ShiftX method (hyperdimensional statistical analysis) with ShiftY (sequence similarity-based). Did the authors test performance with ShiftY turned off? It might reveal even higher predictive power of NMRFx.

The RNA analysis support functions are highly attractive. If possible, releasing tutorial datasets including basic sets of FIDs would draw many users' interest. Publishing instructional videos on platforms like YouTube would also be beneficial.

Please unify the notation of atom names, e.g., Co, Ca, Cb → 13C′, 13Cb, 13Cb (a, b in Greek) (main-text page 10, line 368).

Please unify the author name format in the cited references, e.g., Andrea Muhlbauer, Stephan Seip… → A. Muhlbauer, S. Seip… (main-text page 35, line 847). (also please unify the format items in the references such as pages, volume, year)

Figure 2 (page 18) currently lacks error bars. Ideally, mean, maximum, and minimum error values should be shown, or alternatively a data point distribution plot (dots) provided.

Figure 6 (page 22) shows only 15N R1 and R2 relaxation data. It would be interesting to also see GUI support for heteronuclear NOE (2D 15N–1H NOE). Recently, 2D heteronuclear NOE measurements, sensitive to fast dynamics (ns–ps), have been favored for assessing local fast motions, rather than relying solely on model-free or spectral density analysis using full 15N relaxation data sets. Many users may also wish to display per-residue bar graphs of S^2 values estimated from these data and chemical shifts.

Reviewer #3

(Remarks to the Author)
This manuscript provides a general introduction to the integrated NMR software NMRFx, released under the GPL license.

The manuscript is well written and informative. I recommend publication after addressing a few minor concerns.

Could the author clarify the level of automation in NMRFx's time-domain data processing? For example, is it more like NMRPipe (fully manual) or MestReNova (fully automated)? Can it read acquisition parameters automatically, perform automatic phase correction, and/or baseline correction? If possible, could the author provide more information about how NMRFx handles Bruker digital filtering?

It is commendable that NMRFx supports plugins. Could the author confirm whether the technology standard is published and documented in sufficient detail so that other developers can create third-party plugins?

The author mentioned embedding TensorFlow, which is frequently updated and often sensitive to Python versions, Nvidia drivers, and CUDA libraries. How does the author ensure smooth installation of the embedded TensorFlow?

Response to Reviewer Comments

NMR Data Processing, Visualization, Analysis and Structure Calculation with NMRFx

Ellen Koag[1], Simon G. Hulse[1], Gregory L. Helms[1], Kevin M. Call[1], Michael F. Summers[2,3], Jan Marchant[2], and Bruce A. Johnson[1,*]

[1]Structural Biology Initiative, Advanced Science Research Center at the CUNY Graduate Center, 85 St. Nicholas Terrace, New York, NY, 10031, USA
[2]Department of Chemistry and Biochemistry, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD, 21250, USA
[3]Howard Hughes Medical Institute, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD, 21250, USA
[*]bjohnson@gc.cuny.edu

November 7, 2025

Reviewer 1

I think the manuscript is well written and the software seems useful. What I would ask for is a clearer distinction if things apply for small molecules or macro molecules. I understand that there are some small molecule aspects as well (e.g. the introduction gives "Assignment of the natural product taccalonolide E" as a use case), but later on it is not always clear what things refer to. For example, 2.10 Prediciton - is that for macro molecules only, or for small molecules, or both? So overall, a clear statement of the application domain would be good. It's probably obvious for the authors, but not necessarily for the reader.

Although it was developed with biomolecular NMR as the primary use case, NMRFx provides a rich set of features which make it well-suited to small molecule NMR analysis. Its support to visualize multiple 2D spectra with correlated crosshair cursors (§2.7) along with sophisticated assignment tools puts NMRFx in a position to be a leader in the small molecule NMR space.

The only features of NMRFx which are not suitable for small molecule projects are those which have defined use cases, including the protein and RNA chemical shift prediction models and RunAbout, for the assignment of protein backbone shifts. As mentioned in §2.10.3, chemical shift prediction is supported for small molecules, with further development expected in this area in the future. Furthermore, tools exist that aid spectral assignment (§2.9.1) and three-dimensional structure determination (§2.11.2). The case study in §3.2 gives an example of the full assignment of a complex natural product.

Furthermore, we now (as mentioned in response to reviewer 2) provide additional tutorial information at the nmrfx.org website. This includes an NMRFx project with the full data used in the Taccalonlide E assignment. The project can be accessed (with our Git project manager) at various stages in the assignment process.

Reviewer 2

In this paper, the authors present a software package equipped with a wide variety of functions for NMR measurements, reaching a high level of completeness.

Modern NMR spectroscopy is already a highly established, state-of-the-art methodology for signal analysis and structural modeling of biomolecules such as proteins and nucleic acids (DNA, RNA). By leveraging isotope labeling, diverse pulse sequence developments, and magnetic alignment techniques such as RDC and PCS, as well as long-range restraints from PCS (pseudo-contact shifts) and PRE (paramagnetic relaxation effects), NMR provides an unusually rich variety of structural information compared to other spectroscopic methods. At the same time, objective measurement, analysis, and interpretation require deep theoretical background knowledge—including organic chemistry, elementary quantum mechanics, and thermodynamics—as well as computational sciences such as mathematical optimization, molecular dynamics, statistics, and information theory.

Unlike X-ray crystallography or cryo-EM, NMR structural modeling proceeds through intermediate steps of signal assignment and structural information extraction, followed by computational model construction. Whereas X-ray or cryo-EM directly provide structural maps, NMR inherently leaves more room for human interpretation. Depending on the case, analyses may stop at signal assignment for ligand interaction or dynamics studies, or proceed directly to full 3D structural modeling.

For these reasons, full automation of NMR analysis systems is realistically very difficult, despite many attempts. In practice, users require not only semi-automated analysis but also flexible functions that allow visual verification of data.

The software includes data processing functions (1D, 2D, 3D, and 4D FFT), even for the highly labor-intensive NUS (Non-Uniform Sampling) method, implemented within the same package. This allows users to begin with standard NMR data sets (FIDs and measurement parameters), then seamlessly analyze spectral data sets for all NMR analyses. Importantly, it also allows users to easily return to the processing stage if adjustments are required—for example, to correct poor phasing, slight offset mismatches in individual dimensions, or inverted data increments in 3D spectra acquired using States-TPPI or Echo/Anti-Echo methods. These issues frequently arise even in routine processing.

Moreover, the system enables automatic NMR signal detection (or import of peak tables generated by external software such as DEEP-Picker), automatic signal assignment, and consistency analysis between experimentally observed data and computationally modeled NMR structural coordinates using RDCs.

A particularly noteworthy feature is that the software also implements effective functions for RNA NMR analysis, which has long been a challenge in structural biology.

Although recent advances in deep learning have improved methods for predicting chemical shifts from biomolecular structural coordinates, this remains an extremely difficult task. Nonetheless, the approach reported in this paper for predicting RNA NMR chemical shifts—through sequence dependency and statistical refinements—achieves good predictive performance using strategies distinct from those applied to proteins.

The tool developed by this group is especially valuable in that it can visually represent sequence-specific assignment information, making it easier to identify problems, while cleverly avoiding a fully "black-box" design. This enables safer and more reliable RNA NMR data interpretation and structural modeling, making the tool highly practical for signal assignment and model building.

Another strength of this software is its ease of installation and wide compatibility across operating systems (Windows, macOS, Linux).

The use of Git for recording analysis histories is also highly effective, enabling review of workflows and improving efficiency.

Although CCP-NMR is a well-known comparable software, the present tool is more refined, particularly in its powerful support for RNA NMR signal assignment and structural analysis in solution—a feature of great current interest.

While artificial intelligence systems such as AlphaFold2 have succeeded in accurately predicting protein structures from sequence data, their training on crystal structures limits them to predicting only the most stable conformations. They provide no information on dynamic states or folding stability, and structural biologists should pay greater attention to this limitation. NMR remains the only technique capable of directly providing atomic-level information on motion and stability. This software, through its diverse functions—signal linearity, chemical shift prediction, and RDC consistency—can contribute to identifying data suggestive of dynamics.

In the future, many AI-assisted NMR data analysis tools will surely appear. However, the fact that this software achieves robust automation without turning the crucial task of signal assignment into a complete black box, while still allowing easy visual confirmation, is a highly commendable strength.

The software pursues modern usability and stability, supports a wide range of formats, and is filled with functions designed to assist many users. It can also be expected to flexibly incorporate future technical advances.

Making such a large-scale software, which must have required enormous effort and time, freely available to the academic community represents a highly significant contribution to the global NMR field.

For these reasons, I believe this paper deserves acceptance in the relevant international journal.

Minor suggestions for revision:

1. ShiftX+ blends the older ShiftX method (hyperdimensional statistical analysis) with ShiftY (sequence similarity-based). Did the authors test performance with ShiftY turned off? It might reveal even higher predictive power of NMRFx.

ShiftX2 consists of ShiftX+, the structure-based predictor, and ShiftY+, the sequence-based predictor. This means that ShiftX+ *is* the program most similar to NMRFx's predictor. The same test proteins considered in Table S9 of the ShiftX2 publication (Han et al., *J Biomol NMR* **50**, 43–57 (2011)) were considered, with the SHIFTX+ RMSD values taken from said table. Although future work is planned in this area, the primary intention of the figure is to show that NMRFx's current predictor, which has the advantage of convenient integration into NMRFx rather than involving an upload/download to some web server, has an efficacy which is comparable to SHIFTX+.

2. The RNA analysis support functions are highly attractive. If possible, releasing tutorial datasets including basic sets of FIDs would draw many users' interest. Publishing instructional videos on platforms like YouTube would also be beneficial.

We are committed to a full series of tutorials on the use of NMRFx. There are currently seven video tutorials on the NMRFx web site. We did have a tutorial on RNA assignment, but we have removed it as it used an earlier version of NMRFx. We expect to have a new one available relatively soon (certainly prior to publication). We are making available the full NMRFx project (with Git history) and raw data used in the RNA assignment case study. Additionally we have data and a written step-by-step tutorial for a metabolomics project examining freeze tolerance in wheat. This uses the Scanner tool described in the case-study on algal metabolism. We also have a tutorial for the natural product assignment (taccalonolide E) described in one of the article's case studies. This includes the raw and processed data and an NMRFx project saved at various points in the assignment process.

3. Please unify the notation of atom names, e.g., Co, Ca, Cb → 13C', 13Cb, 13Cb (a, b in Greek) (main-text page 10, line 368).

The atoms names have been updated to be consistent as requested by the reviewer.

4. Please unify the author name format in the cited references, e.g., Andrea Muhlbauer, Stephan Seip... → A. Muhlbauer, S. Seip... (main-text page 35, line 847). (also please unify the format items in the references such as pages, volume, year)

The references have been updated so that they are in close alignment with the Nature citation style. The references are stored in a `.bib` file, so that when they are processed by the journal after acceptance they should be formatted appropriately.

5. Figure 2 (page 18) currently lacks error bars. Ideally, mean, maximum, and minimum error values should be shown, or alternatively a data point distribution plot (dots) provided.

As noted above the RMS deviations for ShiftX+ data are taken from the supplementary material of the ShiftX2 paper, so the variance of these deviations are not accessible. We now provide, in the supplementary material, scatter plots of experimental vs. predicted shifts (Figure S1), and violin plots of the deviations between the experimental and predicted shifts (Figure S2).

6. Figure 6 (page 22) shows only ^{15}N R_1 and R_2 relaxation data. It would be interesting to also see GUI support for heteronuclear NOE (2D ^{15}N–^{1}H NOE). Recently, 2D heteronuclear NOE measurements, sensitive to fast dynamics (ns–ps), have been favored for assessing local fast motions, rather than relying solely on model-free or spectral density analysis using full ^{15}N relaxation data sets. Many users may also wish to display per-residue bar graphs of S^2 values estimated from these data and

chemical shifts.

We've updated Figure 6 to show not only the R_1 and R_2 values, but also the heteronuclear NOE values, and the order parameter S^2 derived from a modelfree analysis of the three NMR parameters for each residue. More complete description of the analyses that can be done with the RING NMR Dynamics program (standalone or operating as a plugin as shown here) will be included in a separate publication on a new release of RING.

Reviewer 3

This manuscript provides a general introduction to the integrated NMR software NMRFx, released under the GPL license. The manuscript is well written and informative. I recommend publication after addressing a few minor concerns.

1. Could the author clarify the level of automation in NMRFx's time-domain data processing? For example, is it more like NMRPipe (fully manual) or MestReNova (fully automated)? Can it read acquisition parameters automatically, perform automatic phase correction, and/or baseline correction? If possible, could the author provide more information about how NMRFx handles Bruker digital filtering?

Time-domain processing is generally automatic. NMRFx parses the vendor parameter files and automatically generates and applies a processing script for 1D and 2D datasets. Scripts can also be generated for higher dimension datasets, but this is not automatically done at the point of loading the FID. The user can set preferences which dictate whether phases are loaded from a vendor file, are automatically calculated, or should be set manually.

A couple of methods are provided to handle Bruker digital filtering. The first simply involves FT, a first order phase correction based on the group delay, and IFT. A more complex protocol, based on phasing the spectra to accound for the group delay, discarding the imaginary data and then recovering it with a Hilbert transformation, is also provided and is used by default. The user can select the mode with a checkbox and choose, based on the quality of the baseline, which method is preferable for their data. These methods happen as the Bruker FID is loaded. Alternatively, one can insert an explicit processing operation (the BZ op) that has these and other choices for the calculation. Since we provide complete source code for NMRFx all these operations can be examined by interested users.

2. It is commendable that NMRFx supports plugins. Could the author confirm whether the technology standard is published and documented in sufficient detail so that other developers can create third-party plugins?

We thank the reviewer for their enthusiasm about supporting plugins and encouraging us to document the protocol. We've now added documentation on how to install the necessary tools, download NMRFx, and compile the full application, and documentation on how to build a Plugin. The GUI code of a plugin can currently appear in nine different locations (such as various menus or tool sections). For each of the nine locations we provide examples of the code to build a tool that displays at that location. The examples are located in an nmrfx-example-plugins module in the main source repository and their use is described in the new documentation.

3. The author mentioned embedding TensorFlow, which is frequently updated and often sensitive to Python versions, Nvidia drivers, and CUDA libraries. How does the author ensure smooth installation of the embedded TensorFlow?

The platform-appropriate library for Tensorflow is included in the installation. The CPU version is provided, since model inference tends to perform adequately without the need for GPUs. This also

means that issues with Nvidia drivers and CUDA libraries are avoided. Despite this, we are currently exploring ways of providing NMRFx with direct access to GPUs, which could enhance the performance for model inference and other operations like NUS processing.