ARTICLE

CrossMark

Farseer-NMR: automatic treatment, analysis and plotting of large, multi-variable NMR data

João M. C. Teixeira[1] · Simon P. Skinner[2] · Miguel Arbesú[1] · Alexander L. Breeze[2] · Miquel Pons[1]

Abstract

We present Farseer-NMR (https://git.io/vAueU), a software package to treat, evaluate and combine NMR spectroscopic data from sets of protein-derived peaklists covering a range of experimental conditions. The combined advances in NMR and molecular biology enable the study of complex biomolecular systems such as flexible proteins or large multibody complexes, which display a strong and functionally relevant response to their environmental conditions, e.g. the presence of ligands, site-directed mutations, post translational modifications, molecular crowders or the chemical composition of the solution. These advances have created a growing need to analyse those systems' responses to multiple variables. The combined analysis of NMR peaklists from large and multivariable datasets has become a new bottleneck in the NMR analysis pipeline, whereby information-rich NMR-derived parameters have to be manually generated, which can be tedious, repetitive and prone to human error, or even unfeasible for very large datasets. There is a persistent gap in the development and distribution of software focused on peaklist treatment, analysis and representation, and specifically able to handle large multivariable datasets, which are becoming more commonplace. In this regard, Farseer-NMR aims to close this longstanding gap in the automated NMR user pipeline and, altogether, reduce the time burden of analysis of large sets of peaklists from days/weeks to seconds/minutes. We have implemented some of the most common, as well as new, routines for calculation of NMR parameters and several publication-quality plotting templates to improve NMR data representation. Farseer-NMR has been written entirely in Python and its modular code base enables facile extension.

Keywords NMR spectroscopy · Data analysis · Intrinsically disordered proteins · Paramagnetic-NMR · Proteins · Chemical shift perturbations

✉ João M. C. Teixeira
joaomcteixeira@gmail.com

Simon P. Skinner
s.p.skinner@leeds.ac.uk

Miguel Arbesú
miguelarbesu@ub.edu

Alexander L. Breeze
a.l.breeze@leeds.ac.uk

Miquel Pons
mpons@ub.edu

[1] BioNMR Group, Inorganic and Organic Chemistry Department, University of Barcelona, Barcelona, Spain

[2] Astbury Centre for Structural Molecular Biology, Faculty of Biological Sciences, University of Leeds, Leeds, UK

Introduction

Flexible proteins, including fully disordered proteins, short disordered segments or interdomain linkers, as well as multibody macromolecular complexes, play a fundamental role in many key regulatory processes by which environmental information is sensed, processed or used to modulate complex responses. The reaction of these systems to environmental changes, which provide fundamental information on their function, can be experimentally investigated by exploring them under a large set of conditions and *evaluating* their multivariable response. In this regard, Nuclear magnetic resonance (NMR) generates atomic level data bearing structural and dynamic information. Over the last two decades, the information that can be extracted from NMR observables has continuously increased owing to the development of sophisticated NMR methods combined with molecular biology protocols for sample preparation, which together have

considerably enriched the NMR spectroscopist's toolkit. The advent of these advances places NMR in the front line as one of the most valuable techniques to investigate complex protein systems under a multitude of conditions (Kay 2016; Sormanni et al. 2017), which include the mapping of ligand binding sites and their affinities (Arai et al. 2012; Teilum et al. 2017), the presence of molecular crowders (Diniz et al. 2017), different chemical compositions of the solution (Sengupta et al. 2017) as well as site-directed mutations (Arbesú et al. 2017) and local post-translational modifications (Theillet et al. 2012) whose effects expand to the whole molecule. Such experimental design generates large and multivariable sets of NMR data. Complex datasets are also generated in the context of time-resolved NMR studies (Theillet et al. 2013; Mylona et al. 2016), and structural/dynamic analysis by paramagnetic observables (Mahawaththa et al. 2017) from different sources (Carlon et al. 2016; Schilder et al. 2016). Nonetheless, exploiting a combined use of such available NMR tools is not yet routine, mostly due to a lack of platforms that close the gap between the expertise of developers and the users' needs, both at the acquisition and the analysis level.

Numerous NMR data analysis packages have been developed over the past 25 years (Maciejewski et al. 2017); this includes packages for specific applications e.g. relaxation analysis (Mandel et al. 1995; Orekhov et al. 1995; d'Auvergne and Gooley 2008), structure solution (Güntert et al. 1997; Herrmann et al. 2002; Rieping et al. 2007; Güntert 2009), paramagnetic NMR analysis (Pintacuda et al. 2004; Schmitz et al. 2006, 2008; John et al. 2007; Skinner et al. 2013), resonance assignment (Zimmerman et al. 1997; Jung and Zweckstetter 2004a, b; Narayanan et al. 2010), and comprehensive packages to perform a multitude of different analyses mainly focused on spectral processing and data representation, whilst also taking care of the associated bookkeeping (Johnson and Blevins 1994; Bartels et al. 1995; Delaglio et al. 1995; Keller 2003; Vranken et al. 2005; Lee et al. 2015; Skinner et al. 2016). In spite of the plethora of analysis that these packages can perform, which unarguably constitute milestone achievements in the field, we persistently find a gap on software availability dedicated to conversion of NMR observables in large sets of peaklists into correlated information-rich parameters from which biologically relevant answers can be extracted and comfortably presented to the community. This gap expands as the complexity of the systems under the scope of biomolecular NMR projects increases and the number of variables needed to characterise them also grows. Hence, it is often necessary to acquire a multitude of NMR spectra covering this whole range of variables—we term these datasets multivariable. Upon spectral analysis, the researcher faces the task of extracting structural information from correlated series of, maybe hundreds,

NMR-peaklist text files which contain the NMR observables. Moreover, there are multiple ways of combining the variable-dependent information, increasing the complexity of the analysis and leading to an overload of repetitive tasks if peaklists are treated manually.

In addition to the above arguments, the nature of NMR users, and consequently NMR software users, has also evolved over the past two decades, from specialist NMR spectroscopists to structural biologists, who are not necessarily NMR experts, but use the technique routinely for data acquisition and analysis. As a result of the evolution of the biomolecular NMR user base, there is an increasing need for the development of a user-directed NMR pipeline, which focuses on the needs of the community as a whole. The interplay between NMR and other techniques for structural biology is accelerating this paradigm shift (Wassenaar et al. 2012, http://www.inext-eu.org). Moreover, we believe that such a pipeline should be open, easily extensible and accessible to anyone who wants to use and contribute to it, in compliance with the most up-to-date Open Access policies of the European Community (http://ec.europa.eu/research/openscience/index.cfm), and not closed to all except the software's developers.

Taken together, the above arguments have motivated us to develop Farseer-NMR, a user-directed, open source and modular toolbox into which users can input and analyse large datasets of multivariable NMR curated peaklists in an efficient, reproducible and organized manner and without a steep learning curve. The current version has been specifically designed for the analysis of protein-related data. Farseer-NMR transforms the analysis of numerous peaklists from weeks of work to a few minutes of automatic execution, as is represented in Fig. 1 of Farseer-NMR Documentation provided in Supplementary Material, something which has been lacking in the NMR analysis pipeline. The software is coded using the Python programming language. We propose Farseer-NMR also as a nucleation point for the various NMR software packages/routines built to treat NMR-derived parameters, in particular those without a graphical user interface (GUI). We have endeavoured to facilitate this by providing a platform upon which additional routines can be developed and launched. Farseer-NMR has been developed around the concept of a "transparent box", rather than a "black box", or just another Python library for programmers to use. In this respect, we distribute Farseer-NMR with a complete documentation record which, together with inline code commentaries, explain the program structure and architecture, provide examples of new implementations, both for core and for the user interface, and describe the setup and execution of a Farseer-NMR calculation. All settings required for a Farseer-NMR calculation are user-settable, either using a configuration file, or via the GUI. This report describes the concepts of Farseer-NMR, its data structure
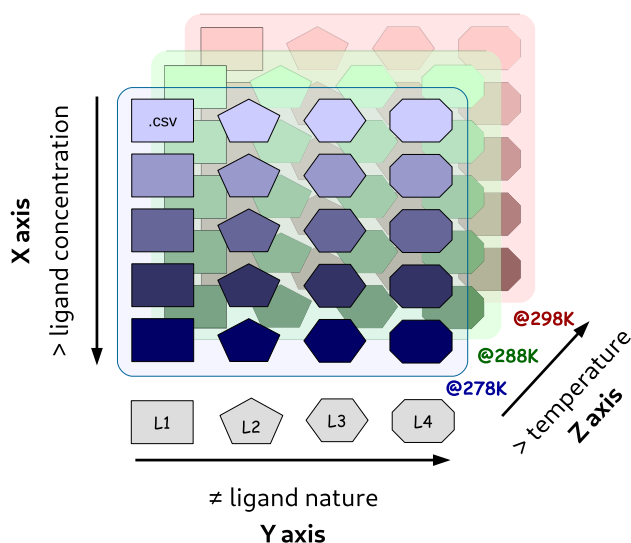
Fig. 1 A schematic representation of the Farseer-NMR Cube. Examples of different variables are given of each Cube's axis. Each coloured rectangle, the first one highlighted with '.csv' for representation is a 2D-NMR peaklist

and implementation of its functionalities, all of which do not require any programming knowledge.

Methods

The Farseer-NMR Project is hosted at GitHub (https://github.com) and distributed under the GPL-3.0 license (https://www.fsf.org/); the following link directs to the project front page: https://git.io/vAueU. Farseer-NMR was written using version 3.6 of the Python programming language, and various Python libraries were used for data management and analysis. Peaklist treatment and management, parameter calculation and data fitting are performed using a combination of routines from Numpy and Scipy (van der Walt et al. 2011) and Pandas (Wes 2010), and plotting templates were developed using Matplotlib (Hunter 2007). The GUI was developed using PyQt v5.7. Version control of the code base was achieved using a git repository hosted by GitHub (http://www.github.com). Farseer-NMR was developed and tested using the GNU/Linux operating system and Anaconda Python (http://www.anaconda.com) distribution, additionally we provide Anaconda environment files for users to setup a running environment if necessary. We have also produced complete documentation containing a plethora of technical information about Farseer-NMR, its concepts and its functionalities. A maintained version of the Farseer-NMR Documentation is available online from the home page repository, the version corresponding to the current release has been supplied as Supplementary Material. At the time of this manuscript, the version of Farseer-NMR is that of an

alpha-release, it is fully operational, but very much open to improvements that will be directed by the interaction with the community. The artwork in this manuscript was prepared using the Matplotlib functions developed for Farseer-NMR, the LibreOffice Suite (https://www.libreoffice.org/), Adobe Photoshop CS2 (http://www.adobe.com) and assembled with GIMP (https://www.gimp.org/).

The Farseer-NMR structure

Farseer-NMR has been developed to enable analysis of multivariable sets of protein-related NMR data in a concerted, correlated and automated manner; variables can be continuous e.g. concentration of a titrated ligand, or discreet e.g. different protein constructs, ligands, environment. User curated peaklists from two-dimensional NMR spectra are imported into the Farseer-NMR data structure, and from here, experimental datasets can be freely navigated and exploited for analysis without any restrictions of the acquisition schedule. Up to three experimental variables can be simultaneously analysed by Farseer-NMR and, therefore, we refer to the data structure as the "Farseer-NMR cube" (Fig. 1 and Sect. II-a of the Documentation PDF), on which each dimension/axis, *x, y* or *z*, of the cube represents a different experimental variable and each cube data point, a coordinate within these axes, is an NMR peaklist table composed of rows that represent the protein residues and columns that contain all residue-related information extracted from the NMR spectra alongside user annotations. This configuration yields a final 5D matrix-like structure stored in a single digital object. The Farseer-NMR Cube can be traversed along the different dimensions describing the dependency of the data on the different defined variables.

For example, if a data set contains diamagnetic and paramagnetic spectra of two different protein constructs, into which a ligand concentration range was titrated, Farseer-NMR can analyse the spectral variations (for example, chemical shifts, intensities, linewidths, couplings) with respect to ligand concentration and for each construct separately (extracting information that can be used to calculate binding affinities, *x* axis), *calculate* the changes caused by a given ligand concentration between the two constructs (*y* axis), and extract paramagnetic relaxation enhancements (PRE) for all spectra by comparing the paramagnetic series with the diamagnetic one (*z* axis), from which distance dependent information can be extracted. Additionally, data calculated along different axes can be parsed and represented together along the other axes (*comparative analysis, see Case Study*). The selection of different dimensions of the cube for analysis and comparison across dimensions are Boolean flags in the setup of a Farseer-NMR calculation (*vide infra*). Axes can

have any number of data points (peaklists) without compromising the organization and representation of results.

Any type of data can be specified along any of the different *x, y*, and *z* axes; however, for the sake of simplicity and clarity, some analysis routines are currently restricted to specific axes, e.g. paramagnetic analysis can only be performed along the *z* axis, inclusion of different protein sequences can only be performed along the *y* axis, and fitting of continuous data can only be performed along the *x* axis. Complete descriptions of axis characteristics can be found in Sect. V-d of the Farseer-NMR Documentation.

The Farseer-NMR workflow

One of the great advantages of having all experimental data in one computational object, the Farseer-NMR Cube, is the ability to slice the cube in any direction and interrogate the data to investigate specific questions or phenomena which are not limited to the acquisition schedule. The first step of any Farseer-NMR calculation is to load the user curated peaklists as digital tables. Farseer-NMR accepts as input peaklists from a variety of formats, namely, Ansig, CCPN Analysis version 2 (Vranken et al. 2005), NMRPipe (Delaglio et al. 1995), NmrView (Johnson and Blevins 1994) and Sparky (Lee et al. 2015), taking a step towards the integration of the NMR community as a whole. Once imported, the peaklists are converted into an enhanced format initially derived from CCPN Analysis version 2 and are stored in a variable-defined hierarchical directory structure created by Farseer-NMR. The peaklists dataset is loaded to memory as a nested dictionary. At this stage, peaklists are scanned for identification of *missing* and/or *unassigned* residues along the first variable (*x* axis) and new rows representing these residues are added to the peaklists. This serves two purposes: identification of those residues and normalization of the peaklists to the same number of rows; these procedures are described in detail in Sect. II-c of the Farseer-NMR Documentation. Amide sidechain entries are also identified at this stage, if required. Following peaklist treatment, the Farseer-NMR Cube is generated.

All settings required by Farseer-NMR for a calculation: selection of dimensions for analysis, types of analysis, types of output plots and settings for the individual plot types, are specified in a JSON file (http://www.json.org/), which can be prepared manually or via the GUI (*vide infra*). Once executed, Farseer-NMR sequentially generates series of peaklists from the Farseer-NMR cube, which represent the system's evolution along an axis/variable at fixed data points of the other two axes and walks through all the possible combinations of x, y and z variables, (Sect. II-b of the Farseer-NMR Documentation) and analyses these series according to the user-specified settings.

Farseer-NMR analysis routines

Each series extracted from the Farseer-NMR cube can be analysed using a number of analysis routines, which are user-configurable. Some commonly used calculation and plotting routines for NMR analysis of macromolecular data have been implemented in the Farseer-NMR code base; we have also included novel methods/templates among which is the recently developed ΔPRE method (Arbesú et al. 2017). In addition, and more importantly, the core of Farseer-NMR has been designed such that virtually any calculation or plotting routine can be implemented. Farseer-NMR analysis routines consist of three steps, namely, (i) calculation of NMR parameters from raw data, (ii) data and values export and (iii) plotting. The current version of Farseer-NMR can calculate nucleus-specific and combined chemical shift perturbations (CSPs), according to Williamson (2013), height and volume ratios, and ΔPRE values according to Arbesú et al. (2017). Elaborated NMR parameters are calculated by *evaluating* each peaklist in an experimental series with the appropriate reference experiment, which is the first in the series, and the calculated values are stored in the corresponding peaklist in newly added columns. Farseer-NMR also contains a restraint fitting platform, whereby continuous data can be fit, currently restricted to data contained along the *x* axis. It is not the aim of Farseer-NMR to provide, presently, a complete suite of routines for fitting the most diverse natures of continuous NMR data, instead, to provide a platform for which users/developers can implement their own fitting routines that can plug into the core Farseer-NMR routines, and, in this way, open the door for an easy distribution among the whole NMR community.

During the calculation run, a set of folders, organised hierarchically according to the defined variables, is created to store all Farseer-NMR output, so that the user can readily access the generated results. Within these folders, formatted, treated and parsed peaklists are exported as comma-separated files, which contain the newly calculated NMR parameters in addition to the original data. Moreover, the user-specified plots are created alongside parsed text tables containing only the represented data. UCSF Chimera (Pettersen et al. 2004) attribute files are generated to allow straightforward representation of the calculated restraints in 3D molecular structures; facile extension of this feature to meet other molecular representation software requirements is possible. At present, Farseer-NMR contains eight plotting templates (three bar plots, two scatter plots, a heat map and two continuous data plots), which have been prepared specifically for NMR-derived data representation. These represent some of the most common plotting styles in NMR literature,

along with novel methods of presenting NMR-derived data. The plots generated by Farseer-NMR are composed of subplots that either represent data for all residues in a given series, or residue-specific variations for a given variable (where appropriate). This enables facile identification of significant results from a single figure, either by condition or by residue. Default configurations for all plots have been implemented, which yield publication-quality figures, although templates can be fully configured by the user. Figure 2 shows examples of three of the plotting templates while full information regarding Farseer-NMR plotting features is available in Sect. VI-d of the Farseer-NMR Documentation. If continuous data fitting is performed as part of a Farseer-NMR calculation run, the values obtained from fitting can be represented as a plot of restraint evolution per residue. Farseer-NMR has already been used to generate figures in published work (Bijlmakers et al. 2016; Marimon et al. 2016; Arbesú et al. 2017).

The graphical user interface (GUI)

To facilitate easy preparation and execution of Farseer-NMR calculation runs, a user-friendly GUI has been developed; full description can be found in Sect. III of the Farseer-NMR Documentation. The current GUI is divided into two tabs, namely, Peaklist Selection (Fig. 3) and Settings (Sect. III-c of the Documentation PDF). In similarity with the core code, the GUI was developed to be easily expansible and designed so that new features can be added as new tabs with their respective options and functions. We envisage user-developer contributions playing a significant role in this evolution of the interface. The Peaklist Selection tab consists of three distinct areas: a sidebar for data import and selection, an area for specifying experimental conditions, and an area for construction of the Farseer-NMR data structure. Peaklists are imported by dropping them on to the side bar, which will detect the peaklist format and import the data if the peaklist format is recognised (*vide supra*). Setting up the Farseer-NMR data structure consists of two stages: (1) specifying the numbers of points and labels for each of the axes of the Farseer-NMR cube, that is, experimental variables and number
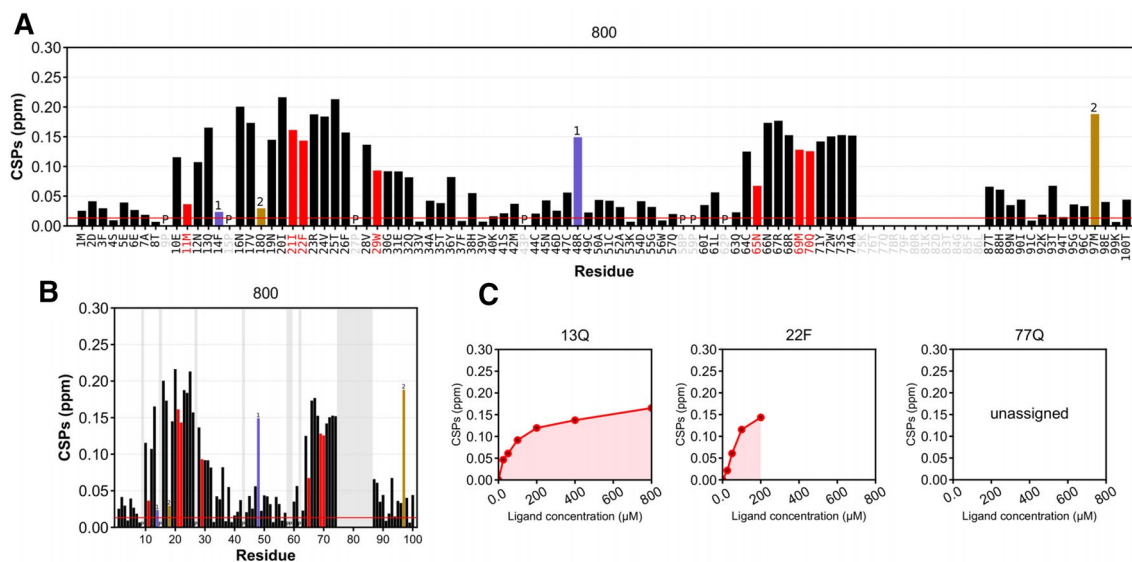

Fig. 2 Three examples of the implemented Farseer-NMR plotting templates. ^1H and ^{15}N chemical shifts were generated from a synthetic data set simulating seven points of a ligand titration of a 100 residues protein. **a** Extended bar plot template and **b** compacted bar plot template representing the combined chemical shift perturbations (CSP), calculated according to Williamson (2013), for the last experiment in the series (with 800 μM ligand) versus the reference experiment (ligand free). Black bars represent residues measured in that spectrum. Unassigned residues are represented in grey **a** × ticks or **b** background shade. Red bars represent missing residues, i.e. residues that have been observed previously in the series but have disappeared at a given ligand concentration; the last measured value is kept. Pro-

lines are identified by character "P". Blue and gold bars (labelled 1 and 2, respectively) represent two pairs of residues with alternative assignment, that after plot analysis can be swapped with confidence (14F ↔ 48R, 18Q ↔ 97M), colours and labels are representations of the user annotations present in the original peaklists. A significance threshold is represented as a red line. **c** Progression of the CSP parameter represented individually for each residue. Peaks that have disappeared along the series are easily identified (22F), as well as unassigned peaks. The represented plots are crops of the full pictures generated by Farseer-NMR that represent the whole series. All colours, sizes, shapes, and labels are user-configurable
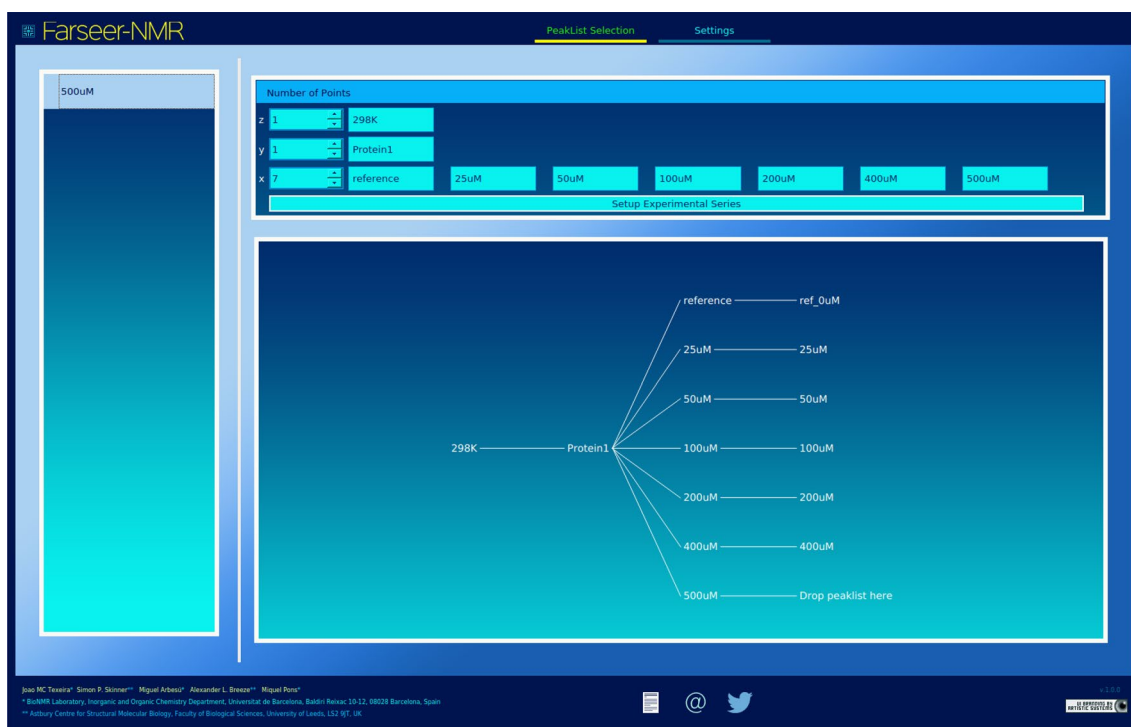
Fig. 3 A screenshot of the Farseer-NMR user interface. The Peaklist selection tab is shown loaded with an artificial dataset as an example. The Farseer-NMR Cube variables are set up and the corresponding data points are populated with peaklists

of measured data points, respectively, and (2) selection of the appropriate peaklist for each combination of variables. Once the numbers of points, and labels for each axis point have been specified, a hierarchical tree diagram representing all specified conditions is drawn with "Drop peaklist here", written at the end of each branch. From here, peaklists corresponding to the correct combination of variables can be dragged and dropped onto the end of each branch of the tree, thereby creating the Farseer-NMR data structure.

The Settings tab of the GUI contains input fields for all the parameters needed to analyse the data using Farseer-NMR. There are two text input boxes, one for setting the *Peaklist Dataset Folder* from which peaklists can be imported aside from the drag-and-drop area, and a second for setting the Calculation Output Path to which all treated peaklists and plots will be written. Selection of axes along which to perform analysis, and options to compare across dimensions, consider side chain amide peaks *and identification* of missing and unassigned residues, are specified by checkboxes in this tab. Settings for chemical shift normalisation, NMR observable selection and parameter calculation, plotting template selection, figure format and file dimension and resolution are also configurable from this tab. A series of popups accessible from this tab contain all the settings for all the implemented plotting templates. The GUI has been forged to facilitate, and invite, advanced users to implement new calculation routines and design the appropriate

setting menus under a new tab. Default settings for all of the fields, are set at start-up, and configuration files can also be loaded and saved from this tab. The footer of the GUI enables instant access to Farseer-NMR documentation, to the Farseer-NMR Twitter account, and to contact the developers via e-mail to submit feedback, request features and report any bugs.

Case study: ΔPRE analysis

An example of the versatility of Farseer-NMR is the implementation of a new procedure for the analysis of intramolecular PRE induced by a paramagnetic probe on disordered protein fragments of a series of proteins harboring mutations that may affect their dynamics and conformational properties (Arbesú et al 2017). The data set was formed of 10 sequence variants that probed the effect of four point mutations at conserved aromatic residues and the presence of the folded domain adjacent to the disordered region by measuring the relaxation induced by paramagnetic centre incorporated in three alternative sites along the sequence. Experimental data consisted of HSQC spectra of the paramagnetic proteins and the corresponding diamagnetic control (20 experiments). The observed PREs were compared with the values expected from a random coil model for the disordered region, independently calculated using Flexible-Meccano (Ozenne et al.
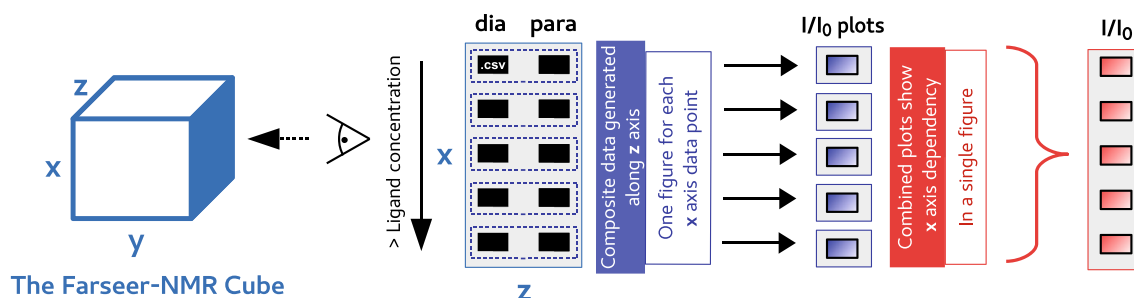
Fig. 4 A schematic representation of the *Comparative/Stacking* Analysis workflow. Farseer-NMR can easily generate combined data from experiments (e.g. Intensity ratios for each residue between paramagnetic and diamagnetic samples in PRE experiments—generated along the z axis—and compare the derived results as a function of an external variable (e.g. ligand concentration, corresponding to the *x* axis of the Farseer-NMR's Cube) as a single combined plot or as a PRE versus ligand concentration curves for individual residues

2012) but incorporated into Farseer-NMR as additional (synthetic) data. Farseer-NMR generated plots showed a direct comparison of predicted and measured PRE, smoothed difference profiles obtained using a running Gaussian convolution filter, heat-map representation of the ΔPRE data, side by side comparisons, and per-residue chemical shift perturbation mapping of the whole constructs. This analysis clearly identified departures from the random coil model and the role of the conserved aromatic residues in the formation of a fuzzy complex in the N-terminal region of the Src protein.

Technically, analysis of paramagnetic NMR data consists in representing the evolution of a paramagnetically generated NMR parameter, in this case the ΔPRE calculated along the *z* axis, along another axis that describes the system as a function of an external variable (sequence variants, point mutations, probe position). Farseer-NMR implements this algorithm under the name "*Comparative/Stacking Analysis*". It serves as the basis for analysing paramagnetic NMR data as well as a parsing routine to regroup results for improved visualization. Figure 4 schematises the workflow of the Comparative Analysis, described in detail in Sect. II-d of the Farseer-NMR Documentation.

Concluding remarks

Farseer-NMR is a software package to streamline and standardize the analysis of NMR data obtained from sets of samples exploring multiple variations in terms of protein sequence or experimental conditions altering the chemical nature of the sample, including ligand binding, or its spectroscopic properties (e.g. paramagnetic and diamagnetic forms or isotropic *versu*s oriented samples). Farseer-NMR facilitates the analysis of multivariable data by allowing the extraction of the individual contribution of the variables, and the identification and characterization of correlated effects and directly produces publication-quality plots, reducing weeks of tedious and

error-prone manual analysis to a few minutes. Farseer-NMR closes a longstanding gap in the quest for creating an automated BioNMR pipeline. It is important to consider also that BioNMR-related peaklists are often analysed iteratively, that is, after restraint calculation and representation, further inspection of the spectra is often necessary to correct for interpretational errors, which may only be evident after considerable analysis. Farseer-NMR drastically reduces the time required in this process by enabling iterative analysis of a complete dataset with a single mouse "click", using previously saved settings, which can be reloaded *a posteriori*. Farseer-NMR was developed focusing on a "box"-like concept, that is, *input → routine → output*, where routines can be as many as imaginable. Rather than being a "black box", Farseer-NMR is an open-source "transparent box" where routines can be inspected, improved and implemented by the users themselves, enthusiast developers or via user-developer interactions. In this vein, Farseer-NMR has been designed to be user-friendly to NMR users with biological background, but also as flexible as possible to engage advanced NMR users. The software is written in Python 3.6 related libraries and is stored in a GitHub repository, which is open-source and free-to-use according to the terms of the GPL-3.0 license. We invite users to fork the repository and create pull requests to contribute to the Farseer-NMR project. Indeed, we have developed Farseer-NMR with the scope of creating a nucleation point for specialized NMR analysis routines that exist throughout the NMR community so that those can become available to the community as a whole in a easy and convenient manner.

during the development of this first release and to Jamie Ferrar (artisticsystems.uk) for providing the UI branding.

References

Arai M, Ferreon JC, Wright PE (2012) Quantitative analysis of multisite protein–ligand interactions by NMR: binding of intrinsically disordered p53 transactivation subdomains with the TAZ2 domain of CBP. J Am Chem Soc 134:3792–3803. https://doi.org/10.1021/ja209936u

Arbesú M, Maffei M, Cordeiro TN, Teixeira JMC, Pérez Y, Bernadó P, Roche S, Pons M (2017) The unique domain forms a fuzzy intramolecular complex in Src family kinases. Structure 25:630–640.e4. https://doi.org/10.1016/j.str.2017.02.011

Bartels C, Xia TH, Billeter M, Güntert P, Wüthrich K (1995) The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. J Biomol NMR 6:1–10. https://doi.org/10.1007/BF00417486

Bijlmakers M-J, Teixeira JMC, Boer R, Mayzel M, Puig-Sàrries P, Karlsson G, Coll M, Pons M, Crosas B (2016) A C2HC zinc finger is essential for the RING-E2 interaction of the ubiquitin ligase RNF125. Sci Rep 6:29232. https://doi.org/10.1038/srep29232

Carlon A, Ravera E, Andrałojć W, Parigi G, Murshudov GN, Luchinat C (2016) How to tackle protein structural data from solution and solid state: an integrated approach. Prog Nucl Magn Reson Spectrosc 92–93:54–70. https://doi.org/10.1016/j.pnmrs.2016.01.001

d'Auvergne EJ, Gooley PR (2008) Optimisation of NMR dynamic models I. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces. J Biomol NMR 40:107–119. https://doi.org/10.1007/s10858-007-9214-2

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6:277–293

Diniz A, Dias JS, Jiménez-Barbero J, Marcelo F, Cabrita EJ (2017) Protein–glycan quinary interactions in crowding environment unveiled by NMR spectroscopy. Chem - A Eur J 23:13213–13220. https://doi.org/10.1002/chem.201702800

Güntert P (2009) Automated structure determination from NMR spectra. Eur Biophys J 38:129–143. https://doi.org/10.1007/s00249-008-0367-z

Güntert P, Mumenthaler C, Wüthrich K (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol 273:283–298. https://doi.org/10.1006/jmbi.1997.1284

Herrmann T, Güntert P, Wüthrich K (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol 319:209–227. https://doi.org/10.1016/S0022-2836(02)00241-3

Hunter JD (2007) Matplotlib: a 2D graphics environment. Comput Sci Eng 9:90–95. https://doi.org/10.1109/MCSE.2007.55

John M, Schmitz C, Park AY, Dixon NE, Huber T, Otting G (2007) Sequence-specific and stereospecific assignment of methyl groups using paramagnetic lanthanides. J Am Chem Soc 129:13749–13757. https://doi.org/10.1021/ja0744753

Johnson BA, Blevins RA (1994) NMR view: a computer program for the visualization and analysis of NMR data. J Biomol NMR 4:603–614. https://doi.org/10.1007/BF00404272

Jung Y-S, Zweckstetter M (2004a) Backbone assignment of proteins with known structure using residual dipolar couplings. J Biomol NMR 30:25–35. https://doi.org/10.1023/B:JNMR.0000042955.14647.77

Jung YS, Zweckstetter M (2004b) Mars: robust automatic backbone assignment of proteins. J Biomol NMR 30:11–23. https://doi.org/10.1023/B:JNMR.0000042954.99056.ad

Kay LE (2016) New views of functionally dynamic proteins by solution NMR spectroscopy. J Mol Biol 428:323–331. https://doi.org/10.1016/j.jmb.2015.11.028

Keller R (2003) The CARA/Lua programmers manual. DOTANAL AG

Lee W, Tonelli M, Markley JL (2015) NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. Bioinformatics 31:1325–1327. https://doi.org/10.1093/bioinformatics/btu830

Maciejewski MW, Schuyler AD, Gryk MR, Moraru II, Romero PR, Ulrich EL, Eghbalnia HR, Livny M, Delaglio F, Hoch JC (2017) NMRbox: a resource for biomolecular NMR computation. Biophys J 112:1529–1534. https://doi.org/10.1016/j.bpj.2017.03.011

Mahawaththa MC, Pearce BJG, Szabo M, Graham B, Klein CD, Nitsche C, Otting G (2017) Solution conformations of a linked construct of the Zika virus NS2B-NS3 protease. Antiviral Res 142:141–147. https://doi.org/10.1016/j.antiviral.2017.03.011

Mandel AM, Akke M, Palmer AG III (1995) Backbone dynamics of *Escherichia coli* ribonuclease HI: correlations with structure and function in an active enzyme. J Mol Biol 246:144–163. https://doi.org/10.1006/jmbi.1994.0073

Marimon O, Teixeira JMC, Cordeiro TN, Soo VWC, Wood TL, Mayzel M, Amata I, García J, Morera A, Gay M, Vilaseca M, Orekhov VY, Wood TK, Pons M (2016) An oxygen-sensitive toxin–antitoxin system. Nat Commun 7:13634. https://doi.org/10.1038/ncomms13634

McKinney W (2010) Data structures for statistical computing in python. In: Proceedings of the 9th Python in Science Conference, pp 51–56

Mylona A, Theillet F-X, Foster C, Cheng TM, Miralles F, Bates PA, Selenko P, Treisman R (2016) Opposing effects of Elk-1 multisite phosphorylation shape its response to ERK activation. Science 354:233–237. https://doi.org/10.1126/science.aad1872

Narayanan RL, Dürr UHN, Bibow S, Biernat J, Mandelkow E, Zweckstetter M (2010) Automatic assignment of the intrinsically disordered protein Tau with 441-residues. J Am Chem Soc 132:11906–11907. https://doi.org/10.1021/ja105657f

Orekhov VY, Nolde DE, Golovanov AP, Korzhnev DM, Arseniev AS (1995) Processing of heteronuclear NMR relaxation data with the new software DASHA. Appl Magn Reson 9:581–588. https://doi.org/10.1007/BF03162365

Ozenne V, Bauer F, Salmon L, Huang J-R, Jensen MR, Segard S, Bernadó P, Charavay C, Blackledge M (2012) Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. Bioinformatics 28:1463–1470. https://doi.org/10.1093/bioinformatics/bts172

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera: a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612. https://doi.org/10.1002/jcc.20084

Pintacuda G, Kaikkonen A, Otting G (2004) Modulation of the distance dependence of paramagnetic relaxation enhancements by CSA × DSA cross-correlation. J Magn Reson 171:233–243. https://doi.org/10.1016/j.jmr.2004.08.019

Rieping W, Habeck M, Bardiaux B, Bernard A, Malliavin TE, Nilges M (2007) ARIA2: automated NOE assignment and data integration

in NMR structure calculation. Bioinformatics 23:381–382. https://doi.org/10.1093/bioinformatics/btl589

Schilder J, Liu W-M, Kumar P, Overhand M, Huber M, Ubbink M (2016) Protein docking using an ensemble of spin labels optimized by intra-molecular paramagnetic relaxation enhancement. Phys Chem Chem Phys. https://doi.org/10.1039/C5CP03781F

Schmitz C, John M, Park AY, Dixon NE, Otting G, Pintacuda G, Huber T (2006) Efficient chi-tensor determination and NH assignment of paramagnetic proteins. J Biomol NMR 35:79–87. https://doi.org/10.1007/s10858-006-9002-4

Schmitz C, Stanton-Cook MJ, Su X-C, Otting G, Huber T (2008) Numbat: an interactive software tool for fitting Deltachi-tensors to molecular coordinates using pseudocontact shifts. J Biomol NMR 41:179–189. https://doi.org/10.1007/s10858-008-9249-z

Sengupta I, Bhate SH, Das R, Udgaonkar JB (2017) Salt-mediated oligomerization of the mouse prion protein monitored by real-time NMR. J Mol Biol 429:1852–1872. https://doi.org/10.1016/j.jmb.2017.05.006

Skinner SP, Moshev M, Hass MAS, Keizers PHJ, Ubbink M (2013) PARAssign—paramagnetic NMR assignments of protein nuclei on the basis of pseudocontact shifts. J Biomol NMR 55:379–389. https://doi.org/10.1007/s10858-013-9722-1

Skinner SP, Fogh RH, Boucher W, Ragan TJ, Mureddu LG, Vuister GW (2016) CcpNmr AnalysisAssign: a flexible platform for integrated NMR analysis. J Biomol NMR 66:111–124. https://doi.org/10.1007/s10858-016-0060-y

Sormanni P, Piovesan D, Heller GT, Bonomi M, Kukic P, Camilloni C, Fuxreiter M, Dosztanyi Z, Pappu RV, Babu MM, Longhi S, Tompa P, Dunker AK, Uversky VN, Tosatto SCE, Vendruscolo M (2017) Simultaneous quantification of protein order and disorder. Nat Chem Biol 13:339–342. https://doi.org/10.1038/nchembio.2331

Teilum K, Kunze MBA, Erlendsson S, Kragelund BB (2017) (S)Pinning down protein interactions by NMR. Protein Sci 26:436–451. https://doi.org/10.1002/pro.3105

Theillet F-X, Smet-Nocca C, Liokatis S, Thongwichian R, Kosten J, Yoon M-K, Kriwacki RW, Landrieu I, Lippens G, Selenko P (2012) Cell signaling, post-translational protein modifications and NMR spectroscopy. J Biomol NMR 54:217–236. https://doi.org/10.1007/s10858-012-9674-x

Theillet F-X, Rose HM, Liokatis S, Binolfi A, Thongwichian R, Stuiver M, Selenko P (2013) Site-specific NMR mapping and time-resolved monitoring of serine and threonine phosphorylation in reconstituted kinase reactions and mammalian cell extracts. Nat Protoc 8:1416–1432. https://doi.org/10.1038/nprot.2013.083

van der Walt S, Colbert SC, Varoquaux G (2011) The NumPy array: a structure for efficient numerical computation. Comput Sci Eng 13:22–30. https://doi.org/10.1109/MCSE.2011.37

Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins Struct Funct Genet 59:687–696. https://doi.org/10.1002/prot.20449

Wassenaar T, Dijk M, Loureiro-Ferreira N, Schot G, Vries SJ, Schmitz C, Zwan J, Boelens R, Giachetti A, Ferella L, Rosato A, Bertini I, Herrmann T, Jonker HR, Bagaria A, Jaravine V, Güntert P, Schwalbe H, Vranken WF, Doreleijers JF, Vriend G, Vuister GW, Franke D, Kikhney A, Svergun DI, Fogh RH, Ionides J, Laue ED, Spronk C, Jurkša S, Verlato M, Badoer S, Pra SD, Mazzucato M, Frizziero E, Bonvin AMJJ. (2012) WeNMR: structural biology on the grid. J Grid Comput 10:743–767. https://doi.org/10.1007/s10723-012-9246-z

Williamson MP (2013) Using chemical shift perturbation to characterise ligand binding. Prog Nucl Magn Reson Spectrosc 73:1–16

Zimmerman DE, Kulikowski CA, Huang Y, Feng W, Tashiro M, Shimotakahara S, Chien C, Powers R, Montelione GT (1997) Automated analysis of protein NMR assignments using methods from artificial intelligence. J Mol Biol 269:592–610. https://doi.org/10.1006/jmbi.1997.1052