

NMR Constraints Analyser: a web-server for the graphical analysis of NMR experimental constraints

Davide Martin Heller and Alejandro Giorgetti*

Department of Biotechnology, University of Verona, Strada 'Le Grazie' 15, Ca' Vignal 1, 37134 Verona, Italy

Received January 29, 2010; Revised April 22, 2010; Accepted May 14, 2010

ABSTRACT

Nuclear magnetic resonance (NMR) spectroscopy together with X-ray crystallography, are the main techniques used for the determination of high-resolution 3D structures of biological molecules. The output of an NMR experiment includes a set of lower and upper limits for the distances (constraints) between pairs of atoms. If the number of constraints is high enough, there will be a finite number of possible conformations (models) of the macromolecule satisfying the data. Thus, the more constraints are measured, the better defined these structures will be. The availability of a user-friendly tool able to help in the analysis and interpretation of the number of experimental constraints per residue, is thus of valuable importance when assessing the levels of structure definition of NMR solved biological macromolecules, in particular, when high-quality structures are needed in techniques such as, computational biology approaches, site-directed mutagenesis experiments and/or drug design. Here, we present a free publicly available web-server, i.e. NMR Constraints Analyser, which is aimed at providing an automatic graphical analysis of the NMR experimental constraints atom by atom. The NMR Constraints Analyser server is available from the web-page <http://molsim.sci.univr.it/constraint>

INTRODUCTION

Nuclear magnetic resonance (NMR) spectroscopy is, together with X-ray crystallography, a major technique for the determination of high-resolution 3D structures of macromolecules. To date, nearly 8200 peptide and protein structures have been determined by NMR and deposited into the RCSB protein data bank (PDB; 1). This technique is based on the observation that several nuclei (e.g. ^1H , ^{13}C and ^{15}N) have an intrinsic magnetic moment of spin and that the spin energy levels are non-degenerate in the

presence of an applied magnetic field. After placing a concentrated homogeneous solution of biological molecules inside a powerful magnet it becomes possible to induce transitions of nuclear spins between energy levels by the application of radio-waves to the sample, and to measure the frequency of the corresponding energetic jump. Each atom produces an NMR signal at a characteristic frequency (chemical shift) which depends on the chemical environment and the molecular structure. Small frequency differences displayed by the same atoms in different environments make it possible to assign different signals (resonances) to specific atoms in the biomolecule.

The standard protocol for the determination of the structure of biological molecules by NMR involves three basic steps: (i) extensive chemical shift assignments of spectral frequencies to the atoms of a target biological molecule; (ii) determination of a set of empirical structural parameters, such as inter- and intra-residue distance restraints based on ^1H – ^1H Nuclear Overhauser Enhancements Spectroscopy (NOESY) experiments; (iii) introduction of the distance restraints in simulated annealing or distance geometry calculations to generate the 3D structure(s) of the biomolecule (2). These calculations are currently performed by using a variety of stand-alone programs such as, CYANA (3), XPLOR-NIH (4), CNS (5) also incorporating or in combination with automatic assignment software CANDID (6), ARIA (7) and others.

The information content of an NMR NOESY experiment is given by the peak frequencies, which determine the assignment of the NOESY peak to the corresponding pair of atoms, and by the peak intensities (volumes), which are approximately proportional to the inverse of the sixth power of the distance between atoms that are sufficiently close in space ($<6\text{ \AA}$). The peak intensities are then converted into a set of lower and upper limits for the distances between pairs of atoms (distances constraints). If the number of constraints is high enough, there will be a finite number of possible conformations (models) of the protein satisfying the constraints, thus generating a family (bundle) of structures that minimize the violations to the spatial constraints. Hence, the more the constraints are

*To whom correspondence should be addressed. Tel: +0039 045 7801; Fax: +0039 045 7929; Email: alejandrogiorgetti@univr.it

satisfied, the lower the root-mean-squared-deviation (RMSD) between the members of the bundle will be. It has to be stressed that along a protein structure, there might be regions better defined than others or residues for which the restraints may suffer from a number of violations along the ensemble of structures.

In addition, the deposition of structural restraints to the PDB (1) and to a specialized database, i.e. Biological Magnetic Resonance Data Bank (BMRB; 8), is mandatory since February 2008 (9). Thus, the availability of a user-friendly tool to analyse and interpret the number of constraints (and their violations) per residue used for the determination of the final structures, may be of valuable importance when assessing the accuracy of the structure, in particular, when high-quality structures are needed in, among others, computational biology approaches and/or drug design techniques. Albeit the importance of performing such analysis and the broad spectra of potential users, there is, at present, no interactive web-server able to provide a quick visualization of the experimental constraints distribution. So far, several scripts have been developed and made available to the community to be used only through locally installed versions of the most frequently used molecular visualization programs such as Pymol (10) and MolMol (<http://www.marcsaric.de/index.php/Molmol>). Here, we present a free publicly available web-server, i.e. NMR Constraints Analyser, aimed at an automatic graphical analysis of the NMR experimental constraints distribution on biological macromolecules, including the analysis of the violated constraints

and the distinction between ambiguous/non-ambiguous restraints.

MATERIALS AND METHODS

The system

The server is divided in to three main modules. The first one constitutes the query module where the user can upload the coordinates and constraints files, respectively. In addition, these files can also be automatically retrieved from a locally installed version of the manually curated database NRG-FRED (11,12; see the 'Discussion' section for details) just by inserting the corresponding PDB accession code. To illustrate the functionalities of the server, three different examples have been included (available from the home-page), together with a direct link to an extensive help/tutorial page. The examples can be reached directly just by choosing among them, and by clicking on the 'Visualise Constraints' button, without the need of filling in the forms.

The core of the server (Figure 1) is constituted by the second module that, upon uploading of both the NMR constraints and pdb-formatted coordinates files, generates a click-able bar chart image showing: (i) the number of constraints per residue (white dots if there are no torsional restraints, otherwise green, in Figure 1A), (ii) the number of violated restraints (red triangles) per residue as extracted from the NRG-FRED database and (iii) the number of ambiguous constraints (magenta dots) involving each residue. Beside the plot, the 3D structure

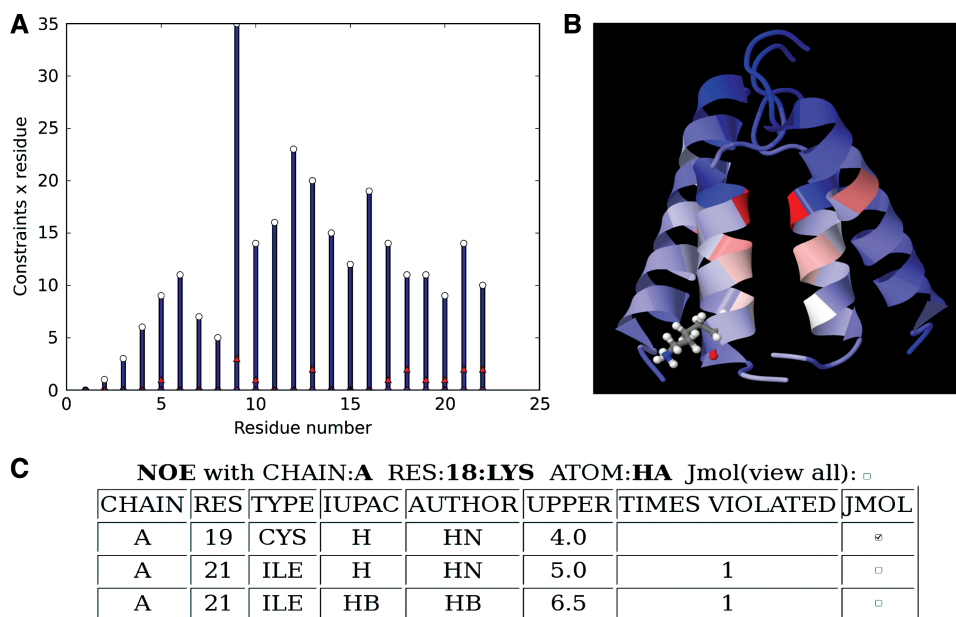


Figure 1. Results page of the NMR analyser server using the file 1XKM (15). (A) Distribution of the experimental constraints per amino acid for chain A. While white dots indicate the number of restraints per residue, red triangles indicate the number of restraint violations. (B) Jmol embedded applet showing the protein and a selected amino acid, i.e. Lys 18 of chain A. The structure is coloured according to the number of constraints per residue. The colour scale expands from blue to red indicating lower or higher number of constraints, respectively. (C) Tabular formatted output listing all distance (torsional, if present) constraints for atoms HA of residue Lys18. First, second and third columns indicate the chain, number and residue of the interacting atoms, respectively; columns fourth and fifth indicate the IUPAC and author atom nomenclatures, respectively; the sixth column indicates the upper limit of the distance constraint, and the seventh column indicates the amount of models of the ensemble in which the restraint is violated. The last column allows the visualization of the constraint in Jmol.

of the molecule is automatically visualized through a web-page embedded Jmol applet (Figure 1B), using a cartoon-type representation. The server highlights the number of constraints per residue with a specific-defined colour scale.

The users can select the residues of interest by a simple graph clicking. Upon selection, while the residue is visualized in ball-and-stick representation and the protein is automatically centered on that residue, the constraints can be visualized as dashed lines, indicating the actual distance between both atoms. Together, these graphical tools offer the user a very simple way of selecting, visualizing and analyzing the residue of interest, atom by atom. Furthermore, a tabular formatted output (Figure 1C) for each residue, detailing the number and name (both IUPAC and author names are considered) of interacting atoms, together with the distances between them, can be generated from a selection-form that can be found below the plot. Finally, the number of models in which the restraint is violated is also detailed in the table.

The third module comprises the automatic generation of two downloadable files, respectively a file containing the hydrogen bond (HB) distribution of the entire biological molecule, calculated by using the HBplus program (13), and an edited pdb-formatted coordinates file with the constraints per residue written in correspondence of the X-ray- β -factor column. The latter utility is based on the fact that, while X-ray crystallography uses the 11th column of the pdb-formatted coordinates files for writing the atom β -factors, NMR-generated coordinates files generally fill the corresponding column with zeroes. We have thus decided to use this 'free' column to add such important information.

Programs and modules

The server was implemented by using Python programming language and returns a HTML-JavaScript output through CGI (Common Gateway Interface) standard protocol. The distribution of experimental constraints from the selected biological molecule is visualized through a click-able graph, generated by the use of Matplotlib (<http://matplotlib.sourceforge.net/index.html>), a Python module for the production of high-quality graphs in two or three dimensions. It provides a wide variety of plot types (lines, bars, pie charts, histograms and many more) in several output formats (PNG, PS and many others) with high degree of customization and flexibility. In the present article, it was chosen for its ability to retrieve image relative coordinates of the plotted information which is needed for the html-imagemap and the relative JavaScript events.

The biological molecule is automatically visualised by the use of Jmol (<http://www.jmol.org/>), an open-source Java viewer for chemical structures in 3D. A very important Jmol feature is the possibility to visualize the molecule through a web-applet built right inside the web-page. To present the retrieved information, we choose a default visualisation in cartoon-style. The colour scale automatically generated by our server, reflects the number of

distance constraints per residue. In addition Jmol gives also the possibility to extend the web-site functionalities by accessing to the Jmol menu upon right-mouse clicking.

The hydrogen bond (H-bonds) calculations are carried out automatically using a server-side version of the program HBplus (13), one of the most frequently used programs for H-bond detection in biological molecules. The output is a text file in which the corresponding H-bond donors and acceptors are listed. The parsing of this file is very simple, because it is tabular-formatted, making it easy to extract information of H-bonds formation between main-chain/main-chain atoms (MM in the file), side-chain/main-chain atoms (SM) and side-chain/side-chain atoms (SS). This file can be downloaded from the results page of our server.

RESULTS

Server testing

The server has been extensively tested by analyzing more than 250 structures taken directly from the PDB and from the locally installed version of the NRG-FRED database. The server is, at the moment, able to deal with the following formats of constraints files: NMR-STAR (uploaded from the NRG-FRED database), CYANA, XPLOR/CNS and several variations within them. The format variations are found specially when considering multi-chain biological macromolecules.

Studied case

Among the analyzed structures we have chosen one of them containing interesting features that prompted us to introduce modifications in the web-server. The structure for which the distance constraints distribution and coordinates are shown in Figure 1 (PDB accession code 1XKM; 15), corresponds to a multi-chain protein present in both the PDB and the NRG-FRED databases. This entry is very interesting because the structure presents a dimer of dimers quaternary fold (Figure 1B). In the plot, together with the indication of the number of restraints per residue (green dots if there are torsional constraints, otherwise white), the number of violated restraints involving the same residue, as extracted from the NRG database, is shown as a red triangle. These features are detailed in the tabular output (Figure 1C). The latter offers, for each atom of interest, the following data: (i) chain of the interacting atom; (ii) residue number; (iii) residue type; (iv) IUPAC atom name; (v) author defined atom name; (vi) upper allowed distance for the constraint; (vii) the number of the models in the ensemble of structures in which the restraint is violated and (viii) a clickable box that allows restraints visualization. By performing four server runs, the different distributions of constraints defining the structure and the violations to the latter on each monomer can be calculated. The Jmol visualization with a selected amino acid, i.e. Lys 18, was plotted in Figure 1 B. In the latter case, the colour-scale expands from dark blue to red, thus indicating the presence of a low and a high number of constraints, respectively.

DISCUSSION

The NMR Constraints Analyser is publicly free available server that allows, with a few mouse-clicks, to perform a rapid analysis of the NMR experimental constraints from which the 3D structure of the biological macromolecule was determined. As discussed previously, another important parameter commonly used for assessing the definition of different regions of NMR structures is the RMSD per residue of the bundle of structures produced by NMR spectroscopy. Differently from what can be expected, it was recently shown (14) that the correlation between the number of constraints and the RMSD along the 3D structure of the biological molecule sometimes is weak, indicating that the study of the constraints may offer a more detailed measurement of the accuracy of the structure atom by atom. In this regard, and considering that the RMSD of a bundle of NMR structures can be visualized with almost all the existing visualization programs, we have chosen: (i) to plot the distribution of constraints (ambiguous/non-ambiguous, distance and torsional) per residue; (ii) to plot the restraints violations and (iii) to write the number of constraints per residue in the 11th column of the pdb-formatted coordinates file (X-ray- β -factor column) instead of the RMSD, so the user, if interested, shall download the modified file and use any protein visualization program locally installed to perform further analysis. Our principal aim was to develop a tool able to complement the results obtained by the study of the RMSD with a quantity directly derived from the experiments, i.e. the number of constraints, providing in this way a more complete and accurate picture of the structure of interest.

One of the main problems encountered during the creation of the present web-server was due to the high heterogeneity of the PDB deposited NMR experimental data, making the different formats very difficult to be parsed with a single program. In this sense, very recently a large-scale international collaborative effort (11,12) took an extensive approach toward the filtering and format conversion of nearly all experimental NMR data files (around 5300 structures with submitted experimental data) present in the PDB databank (1). The filtered and converted data (FRED; 11), including 'cleaned' NMR pdb-formatted coordinates (DOCR), the distance and torsional constraints files in the STAR format and the calculation of the restraints violations can be found in the NMR Restraints Grid (NRG) FRED accessible from the BMRB database. The final version of this project included data from more than 5266 entries (12), and it is continuously growing. For this reason, we suggest to the user to search first the desired pdb-entry in our locally installed version of the NRG-FRED, offering automatic upload and maximum compatibility. This option does not mean that the present server is intended to act as a graphical interface of the NRG-FRED database, but may help avoiding future tedious and not always easy format conversions, especially for researchers not used in dealing with this kind of formats. Still, if a particular file of interest has not yet been processed and/or included in NRG-FRED, or if the user just needs to visualise the

constraints of his/her structure, the pdb-formatted coordinates and constraints files can always be uploaded directly into our server.

The present server is intended not only to be used by NMR spectroscopists, but it is also meant for the broad community of scientists who may necessitate to assess the different levels of definition of the residues along NMR-solved structures for different scopes, including the generation of homology models, drug design, molecular dynamics simulations, quantum mechanical studies and site-directed mutagenesis experiments.

ACKNOWLEDGEMENTS

R. Laskowski for allowing us to use the HBplus program. We acknowledge Prof. H. Molinari, Dr M. Assfalg, Dr M. D'Onofrio, S. Zanzoni and the entire NMR-group of the University of Verona for very useful and illuminating discussions.

FUNDING

Funding for open access charge: Starting grant from the University of Verona, Department of Biotechnology.

Conflict of interest statement. None declared.

REFERENCES

- Berman, H., Henrick, K., Nakamura, H. and Markley, J.L. (2007) The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res.*, **35**, D301–D303.
- Wuthrich, K. (1995) NMR- this other method for protein and nucleic acid structure determination *Acta Crystallogr. D. Biol. Crystallogr.*, **51**, 249–270.
- Guntert, P. (2004) Automated NMR structure calculation with CYANA. *Methods Mol. Biol.*, **278**, 353–378.
- Schwieters, C.D., Kuszewski, J.J., Tjandra, N. and Clore, G.M. (2003) The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.*, **160**, 65–73.
- Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S. *et al.* (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D. Biol. Crystallogr.*, **54**, 905–921.
- Herrmann, T., Guntert, P. and Wuthrich, 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.
- Rieping, W., Habeck, M., Bardiaux, B., Bernard, A., Malliavin, T.E. and Nilges, M. (2007) ARIA2: automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics*, **23**, 381–382.
- Doreleijers, J.F., Mading, S., Maziuk, D., Sojourner, K., Yin, L., Zhu, J., Markley, J.L. and Ulrich, E.L. (2003) BioMagResBank database with sets of experimental NMR constraints corresponding to the structures of over 1400 biomolecules deposited in the Protein Data Bank. *J. Biomol. NMR*, **26**, 139–146.
- Markley, J.L., Ulrich, E.L., Berman, H.M., Henrick, K., Nakamura, H. and Akutsu, H. (2008) BioMagResBank (BMRB) as a partner in the Worldwide Protein Data Bank (wwPDB): new policies affecting biomolecular NMR depositions. *J. Biomol. NMR*, **40**, 153–155.
- Papadopoulos, E., Oglecka, K., Maler, L., Jarvet, J., Wright, P.E., Dyson, H.J. and Graslund, A. (2006) NMR solution structure of

- the peptide fragment 1-30, derived from unprocessed mouse Doppel protein, in DHPC micelles. *Biochemistry*, **45**, 159–166.
11. Nederveen, A.J., Doreleijers, J.F., Vranken, W., Miller, Z., Spronk, C.A., Nabuurs, S.B., Guntert, P., Livny, M., Markley, J.L., Nilges, M. *et al.* (2005) RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins*, **59**, 662–672.
 12. Doreleijers, J.F., Vranken, W.F., Schulte, C., Lin, J., Wedell, J.R., Penkett, C.J., Vuister, G.W., Vriend, G., Markley, J.L. and Ulrich, E.L. (2009) The NMR restraints grid at BMRB for 5,266 protein and nucleic acid PDB entries. *J. Biomol. NMR*, **45**, 389–396.
 13. McDonald, I.K. and Thornton, J.M. (1994) Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.*, **238**, 777–793.
 14. Snyder, D.A., Bhattacharya, A., Huang, Y.J. and Montelione, G.T. (2005) Assessing precision and accuracy of protein structures derived from NMR data. *Proteins*, **59**, 655–661.
 15. Raimondo, D., Andreotti, G., Saint, N., Amodeo, P., Renzone, G., Sanseverino, M., Zocchi, I., Molle, G., Motta, A. and Scaloni, A. (2005) A folding-dependent mechanism of antimicrobial peptide resistance to degradation unveiled by solution structure of distinction. *Proc. Natl Acad. Sci. USA*, **102**, 6309–6314.