



## ARIAweb: a server for automated NMR structure calculation

Fabrice Allain, Fabien Mareuil, Hervé Ménager, Michael Nilges, Benjamin Bardiaux

### ► To cite this version:

Fabrice Allain, Fabien Mareuil, Hervé Ménager, Michael Nilges, Benjamin Bardiaux. ARIAweb: a server for automated NMR structure calculation. Nucleic Acids Research, 2020, 10.1093/nar/gkaa362 . pasteur-02569631

**HAL Id: pasteur-02569631**

**<https://pasteur.hal.science/pasteur-02569631>**

Submitted on 11 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

# ARIAweb: a server for automated NMR structure calculation

Fabrice Allain<sup>1,†</sup>, Fabien Mareuil<sup>2,†</sup>, Hervé Ménager<sup>2</sup>, Michael Nilges<sup>1</sup> and Benjamin Bardiaux<sup>1,\*</sup>

<sup>1</sup>Structural Bioinformatics Unit, Department of Structural Biology and Chemistry, CNRS UMR 3528, Institut Pasteur, Paris, 75015, France and <sup>2</sup>Bioinformatics and Biostatistics Hub, Department of Computational Biology, CNRS USR 3756, Institut Pasteur, Paris, 75015, France

Received March 10, 2020; Revised April 14, 2020; Editorial Decision April 27, 2020; Accepted April 28, 2020

## ABSTRACT

Nuclear magnetic resonance (NMR) spectroscopy is a method of choice to study the dynamics and determine the atomic structure of macromolecules in solution. The standalone program ARIA (Ambiguous Restraints for Iterative Assignment) for automated assignment of nuclear Overhauser enhancement (NOE) data and structure calculation is well established in the NMR community. To ultimately provide a perfectly transparent and easy to use service, we designed an online user interface to ARIA with additional functionalities. Data conversion, structure calculation setup and execution, followed by interactive visualization of the generated 3D structures are all integrated in ARIAweb and freely accessible at <https://ariaweb.pasteur.fr>.

## INTRODUCTION

Protein structure determination is crucial for understanding protein function, as it paves the way to the discovery of new drugs and of new approaches to control pathological biological processes. The recent advances in structural biology now allow collecting structural information from a variety of techniques at various resolutions (1). For decades, X-ray crystallography was the method of choice to determine structures of macromolecules at atomic resolution. Owing to the latest technical revolution in cryo-electron microscopy (cryo-EM), structures of large complexes can now be readily obtained at very high resolution without the need to obtain crystalline samples. On top of these methods, nuclear magnetic resonance (NMR) spectroscopy is the only technique that allows to determine the structure and dynamics of macromolecules in solution (2). Yet, high-resolution NMR is still limited by the macromolecule size and solubility as well as the necessity to obtain isotopically

labeled samples in sufficient amount. Three-dimensional (3D) structure determination by NMR relies on efficient and reliable assignment strategies (3). Specifically, interpretation of nuclear Overhauser effects (NOEs) as structural data is often ambiguous due to chemical shift degeneracy and spectral overlap. Manually assigning NOE cross-peaks to the corresponding atoms in order to derive distance restraints is a tedious process that generally depends on the, possibly subjective, examination by a trained expert. To automate this process and improve objectivity, software packages such as ARIA (4), CYANA (5), UNIO (6) or ASDP (7) are being actively developed.

The ARIA (Ambiguous Restraints for Iterative Assignment) software package automates the treatment of NMR data and calculation of 3D structures (4,8). In an iterative manner, ARIA automatically assigns NOE cross-peaks and converts them into distance restraints in order to compute structure ensembles by molecular dynamics simulation (9). In the last round of the CASD-NMR (Critical Assessment of Automated Structure Determination by NMR) challenge (10), ARIA consistently produced accurate structures with a mean backbone RMSD of 1.1 Å for 10 targets (11).

Despite being well established in the structural NMR community, the usage of ARIA can be hindered by the complexity of its installation and the sophistication of the configuration and execution tasks. Although being parallelizable, the iterative protocol in ARIA requires access to a computer cluster to reduce execution time. To improve the user experience for ARIA and provide a larger access to automated NMR structure calculation procedures, we have developed the ARIAweb server, with the aim of guiding users (from novice to expert) through the different stages required for NMR structure determination and providing an online service for computationally intensive calculations. Various software packages are used by the NMR community to process and analyze NMR data prior to engage in structure calculation, each having their own proprietary file formats which needs to be converted into the internal ARIA

\*To whom correspondence should be addressed. Tel: +33 01 40 61 38 84; Fax: +33 01 45 68 87 19; Email: benjamin.bardiaux@pasteur.fr

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Present address: Fabrice Allain, Institut Curie, PSL Research University, Paris 75005, France.

format. To circumvent this issue, we propose an online data conversion tool for which there is no equivalent user interface in the standalone ARIA package. Finally, outputs from an ARIA calculation, such as 3D structures and NMR restraints statistics, are displayed with graphical and interactive representations to ease inspection and interpretation of the results in terms of reliability and quality.

## METHODS

### ARIA protocol

ARIAweb implements the latest release of ARIA version 2.3 (4,12) in combination with CNS version 1.21 (13,14) modified with dedicated ARIA routines. The ARIA protocol matches NOE cross-peaks and assigned NMR resonances to convert them into ambiguous distance restraints that are subsequently used to calculate structures by means of molecular dynamics simulated-annealing (9). Considering data and structure consistency, these ambiguous distance restraints can be further refined by iteratively detecting unsatisfied restraints and generating new conformers which better explain the input NMR data. This approach is fully automated in ARIA 2.3. In short, the molecular sequence, NMR cross-peak and chemical-shift lists and other known restraints are supplied to ARIA for automated NOE assignment and structure calculation (Figure 1). ARIA 2.3 uses an iterative protocol where NOE cross-peaks are partially assigned, calibrated to obtain distance estimations and analyzed for restraint violations from the structure ensemble calculated in the previous iteration. This produces a new distance restraint set for the next iteration. After all iterations (typically 9) are completed, the lowest energy conformers are refined in a shell of solvent molecules (15). The final ensemble of conformers is validated with PROCHECK (16), WHATCHECK (17) and MolProbity (18). ARIA version 2.3 integrates all latest developments for objective and reliable NMR structure determination: a bounds-free log-harmonic distance restraint potential with optimal weighting factor (19,20) and an adaptive violation tolerance estimation (11) along with restraint-combination and network-anchoring procedures (21,22). Additionally, ARIA 2.3 includes routines to calculate structures of symmetric oligomers (C2 to C5 and D2 symmetries) (22,23).

### SERVER FUNCTIONALITIES AND USAGE

For the sake of modularity, calculations performed on the ARIAweb server are organized as projects with three main sections corresponding to the provided services: *Data conversion*, *Structure calculation* and *Result visualization* (Figure 1).

#### Data conversion

The minimal input data to perform ARIA calculations are (i) the sequence of the molecular chain(s) composing the molecule for which a user wants to calculate the structure, (ii) one (or more) list(s) of NOE cross-peaks and (iii) one (or more) list(s) of assigned chemical shifts. The input data for molecular sequence and NMR data (NOE cross-peak and assigned chemical shift lists) consist in formatted text

files produced by popular NMR analysis software. Due to the variety of formats to store spectra and chemical shifts information, data has to be converted into the ARIA XML format before setting up a *Structure calculation*.

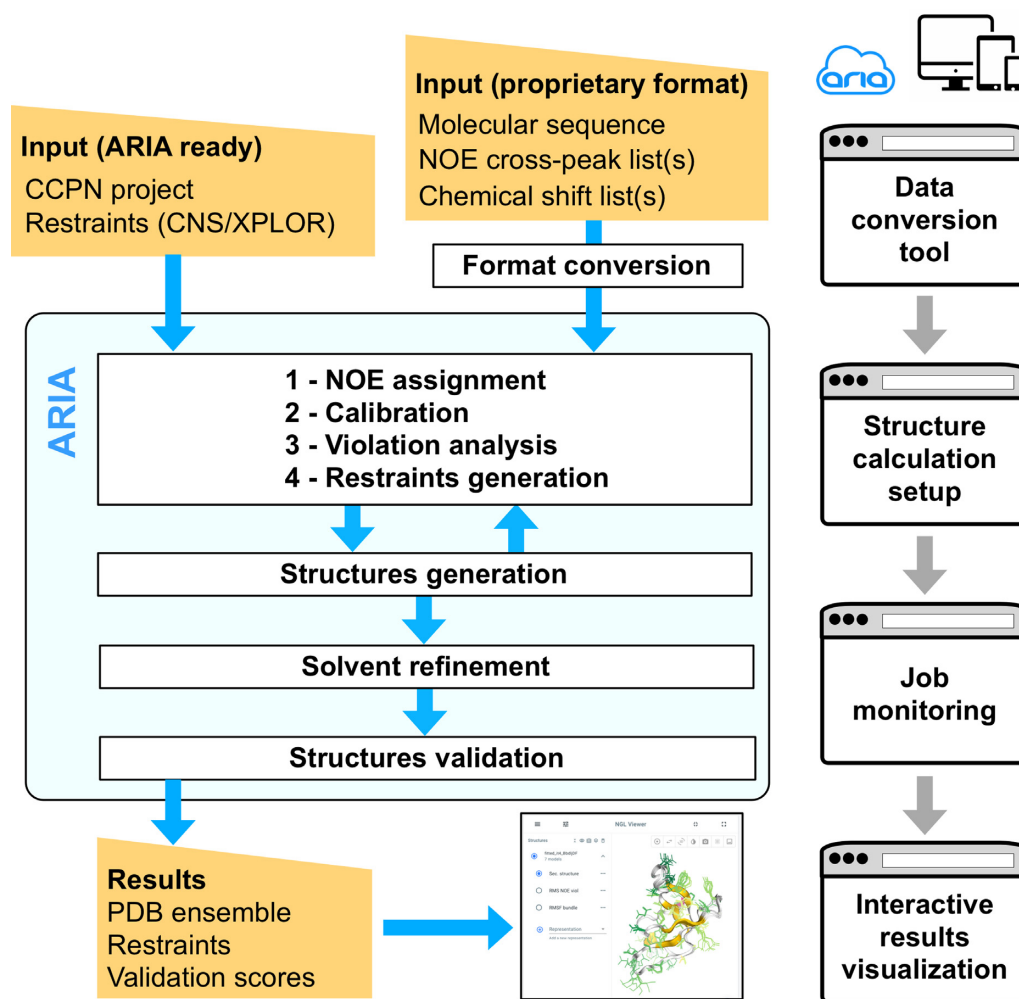
For a *Data conversion* project, supported formats for NOE and chemical shift lists are XEASY, SPARKY, NMRVIEW and ANSIG. Once all input data has been uploaded, the conversion can be processed online (Figure 2). Upon completion of a conversion job, users have the choice to download an archive with the converted data ready to be used with a standalone ARIA installation or to initialize a *Structure calculation* directly within the main project space. Alternatively, CCPN (Collaborative Computing Project for NMR) projects containing all required data in a format-free manner (24), archived as a single tar.gz file, can be directly accepted as input for *Structure calculation* without any conversion. Currently, ARIAweb only supports standard amino-acid residues or DNA/RNA bases definitions and incorporation of zinc ions in tetrahedral coordination. We recommend to use the standalone version of ARIA when modified residues or other organic ligands have to be included.

#### Structure calculation

Similarly to the graphical user interface (GUI) of the standalone ARIA program, the *Structure calculation* service allows users to specify parameters such as input data, NOE assignment criteria, number of conformers to generate or parameters for the solvent refinement. When initiating a *Structure calculation* from a *Data conversion* job, mandatory input data are pre-filled and shown in the corresponding forms (Figure 2). When using data from a CCPN project, a list of all compatible data entries are listed for each appropriate input data type. *Structure Calculation* on ARIAweb is designed to guide the user through the main categories of parameters that need to be checked (if default values are used) or specified (if a user wants to change with customized values). The four main categories are *Setup* (generic information of the project), *Data* (input data organized by their type), *Protocol* (parameters related to the ARIA iterative protocol) and *Structure generation* (generation of structures with CNS). After saving a *Structure calculation* project, users can download an XML file readily usable with a standalone ARIA installation, or submit directly the ARIA structure calculation job online.

#### Results visualization

Once a structure calculation job has been submitted and finished, users can download a ZIP archive containing the entire ARIA run directory with the ensemble of conformers (PDB format), NOE violation statistics, NOE assignments and structure validation scores (text format and graphical PDF) (25). If a CCPN project has been used as input, a new project is generated, notably including new assignments of NOE cross-peaks made by ARIA that can be analyzed further with the appropriate CcpNmr Analysis software (26). More importantly, results can be inspected online via an interactive visualization dashboard (Figure 2). Structure ensembles generated by the server will be displayed interactively for all ARIA iterations with various represen-



**Figure 1.** Schematic representation of the ARIAweb workflow. Input data can be converted to the ARIA XML format with the Data conversion tool. Data are then sent for structure calculation setup, where a user can define custom parameters. Once ready, the structure calculation job is started and enters the ARIA iterative procedure for NOE assignment and structure generation. After a validation step, structure calculation results are available for download and presented for interactive visualization.

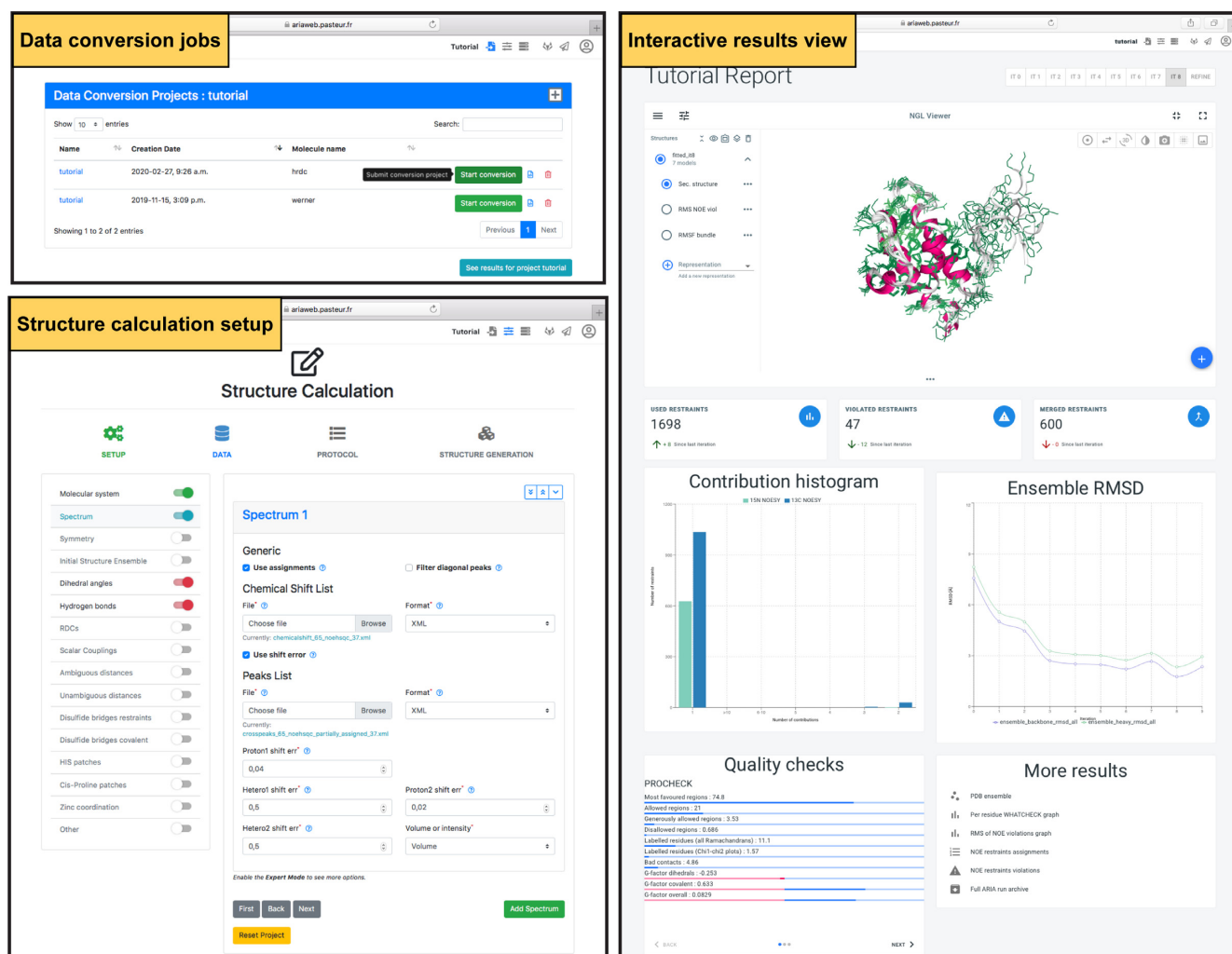
tations. Aside from the traditional ‘cartoon’ representation with colored secondary structures elements, users can directly visualize, on the three-dimensional structure, color-coded statistics such as the average RMS (root mean square) of NOE violations per residue or the RMSF (root mean square fluctuation) of the generated ensemble in ‘sausage’ representation. Such representations allow users to easily identify unconverged regions in the final structure ensemble or residues for which NOE-derived distance restraints are not consistent, which are indicators of problematic or unsatisfactory interpretation of some NOE data by ARIA and that need more in-depth inspection from the expert (27). Restraints statistics, ensemble RMSD or color-coded structure quality scores are also presented as graphs and charts (Figure 2). Additionally, a standardized table reporting various restraints and structural statistics is generated.

## SERVER IMPLEMENTATION

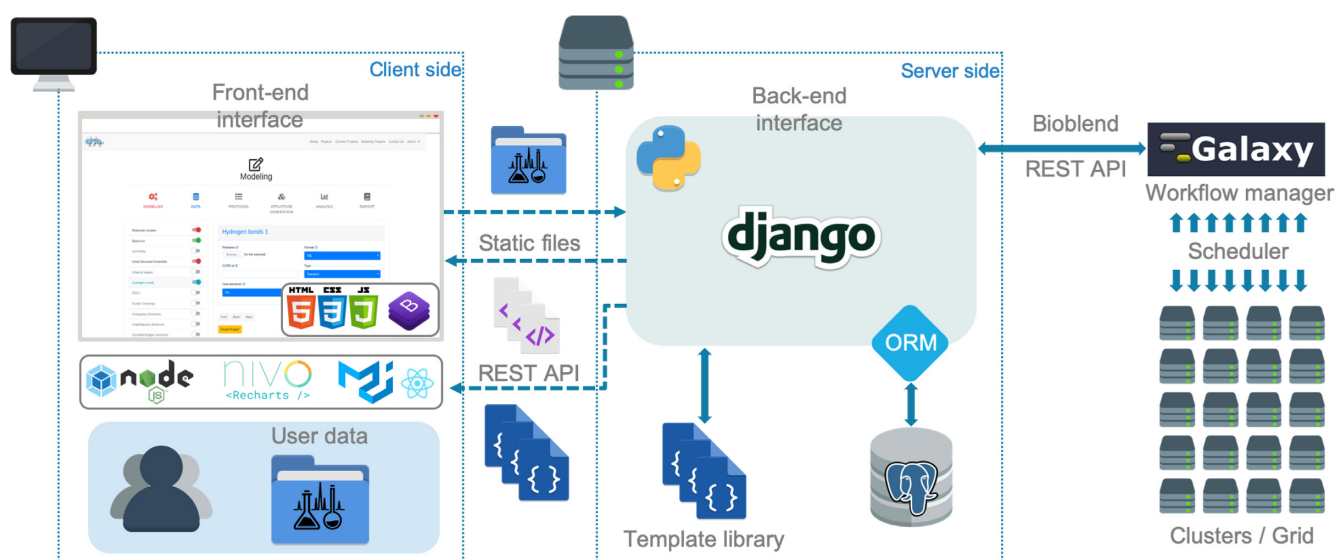
An overview of the ARIAweb server implementation and the technologies employed is shown in Figure 3. The web

application features dynamic forms such as *Data conversion* and *Structure calculation* forms formatted with Bootstrap (<https://getbootstrap.com>). The content displayed on the client-side comes from the templating system of a Django back-end server (<https://www.djangoproject.com>). Input data sent to the server are securely saved in a PostgreSQL database (<https://www.postgresql.org>) using the concept of Object Relational Mapping from the Django framework. To perform and monitor the execution of ARIA jobs on a HPC (high performance computing) cluster, the Django server communicates with the Pasteur Galaxy server (28) through the BioBlend REST API (29). Galaxy is a scientific workflow management system that provides means to build multi-step computational analysis akin to recipes (30). Job status is reported to the user in real-time with the following stages: Building, Pending, Running and Success. Finally, the results are displayed as a single page application written in ReactJS (<https://reactjs.org>) and connected to the back-end with the Django REST framework. This application makes use of the NGL JavaScript library for protein structure visualization (31,32).





**Figure 2.** Screenshots of the jobs management, structure calculation setup forms and the interactive results visualization page, which allows the user to view and analyze the results of a structure calculation job.



**Figure 3.** Diagram describing the ARIAweb implementation workflow. The front-end interface (in the web browser) basically corresponds to the visual aspect of the website and the logics related to graphical components. On the server side, back-end services are in charge of the interaction with the database, the computing cluster and generate the context for the front-end interface.

**Table 1.** Summary of input data and output statistics for the case studies. For each target, the corresponding reference PDB entry is given along with the length of the protein sequence and the number of input NOE cross-peaks used. The number of NOE restraints generated by ARIAweb at the last iteration is also shown. Ramachandran statistics, RMS of NOE restraints violations and bundle precision (mean pairwise RMSD) are presented as indicators for the quality of the generated structure ensembles

Target	PDB	Length.	# NOE cross-peaks (input)	# NOE restraints (output)	Ramachandran statistics (%) <sup>b</sup>	NOE RMS (Å)	Bundle RMSD <sup>a</sup> (Å)	Bundle Accuracy <sup>a</sup> (Å)
Tudor	1G5V	56	2172	1160	90.0/10.0/0.0/0.0	0.112	0.35 (0.08)	1.01 (0.11)
HRDC	1D8B	91	3155	1698	89.3/8.4/1.3/1.0	0.064	0.44 (0.07)	1.36 (0.12)
CsfB <sup>#</sup>	5N7Y	2 × 49	1658	2 × 709 <sup>c</sup>	77.8/18.9/0.0/3.3	0.143	0.33 (0.09)	1.12 (0.08)
RBM20	6SO9	112 <sup>d</sup>	6185	2982	86.7/10.1/2.6/0.6	0.142	0.13 (0.02)	1.57 (0.07)
MANEC	2MSX	114	5201	2349	84.0/15.2/0.2/0.6	0.288	0.58 (0.11)	1.41 (0.10)
HypA <sup>#</sup>	6G81	117	5336	2048	82.0/15.3/1.7/1.0	0.173	0.74 (0.24)	1.96 (0.32)
HR5460A	2LAH	160	17250	4807	88.5/11.2/0.3/0.0	0.175	0.38 (0.06)	1.25 (0.11)

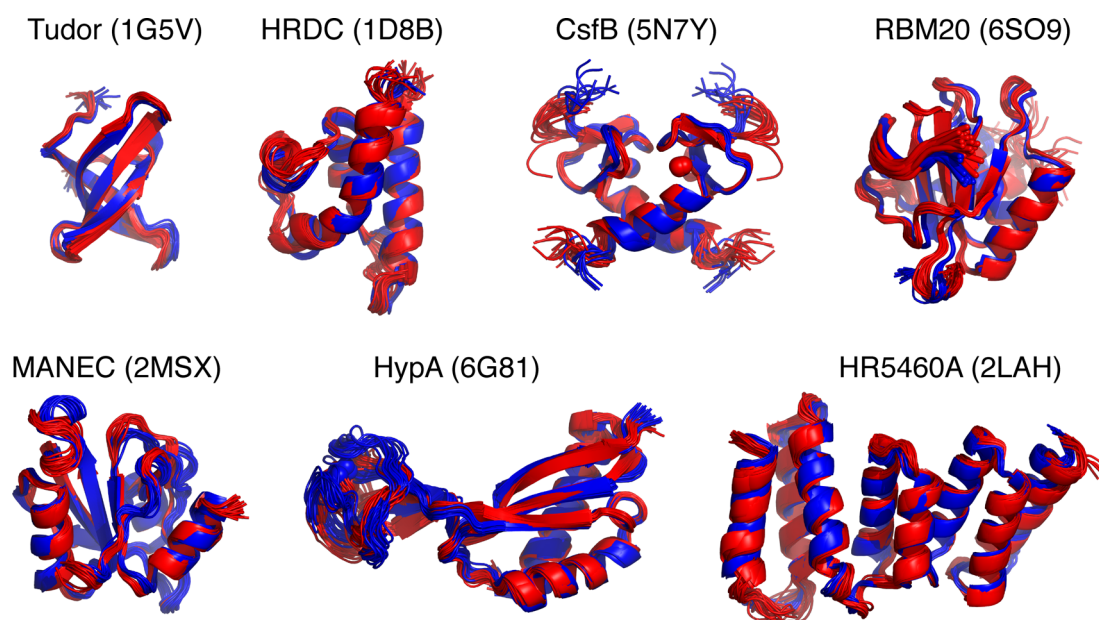
<sup>a</sup>Mean RMSD (sd) for backbone atoms of well-ordered residues.

<sup>b</sup>Percentage of residues of the full-length sequence in the most favored/allowed/generously allowed/disallowed regions of the Ramachandran plot.

<sup>c</sup>Restraint list is automatically copied for each chain in symmetric dimer.

<sup>d</sup>In complex with 6 bp RNA.

<sup>#</sup>protein containing Zn ion bound in tetrahedral coordination.



**Figure 4.** Overview of structure ensembles obtained with ARIAweb for the case studies using automated NOE assignment. For each entry, ARIAweb ensembles (in blue) are shown as cartoon and superimposed on the reference PDB ensembles (in red).

A detailed description of the server functionalities is provided in the online documentation along with example input files in various formats, a step-by-step tutorial and a set of recommendations to help users interpret the results. Additionally, all forms requiring an input file to be specified contains a link to a corresponding example file. Due to the complexity of the *Structure calculation* section, several customizable parameters are hidden by default. Advanced users can reveal and modify those additional parameters by triggering the ‘expert mode’ in their user profile.

Jobs submitted via ARIAweb are executed on the Institut Pasteur HPC cluster with 7000+ cores shared by all institutional services. Each *Structure calculation* job uses 12 cores simultaneously. Typical execution times for *Data conversion* and *Structure calculation* jobs are around 5 min and 90 min, respectively.

## CASE STUDIES

To validate the efficiency of the ARIAweb server, *Structure calculation* projects were set up and performed on a benchmark of seven proteins or protein/RNA complexes, ranging from 56 to 160 amino-acids, for which the solution NMR structure has been solved by experts and deposited in the Protein Data Bank (33) (Table 1). Input data consisted in assigned chemical shift lists, lists of unassigned NOE cross-peaks (derived from 2D, 3D or 4D NOESY spectra) and dihedral angle restraints corresponding to the data originally used to determine the deposited PDB structure. The tested entries are: Tudor domain of the human Survival of Motor Neuron (SMN) protein (1G5V) (34), HRDC domain of the *S. cerevisiae* RecQ helicase (1D8B) (35), dimeric anti-sigma factor CsfB of *B. sub-*

*tilis* (5N7Y) (36), mouse RBM20 RRM domain in complex with RNA (6SO9) (37), MANEC domain of the human hepatocyte growth factor activator inhibitor-1 (HAI-1) (2MSX) (38), Ni metallochaperone HypA from *Helicobacter pylori* (6G81) (39) and the N-terminal domain of the human mitotic checkpoint serine/threonine protein kinase BUB1 (CASD-NMR target HR5460A, 2LAH) (40). The precision (mean pairwise-RMSD of backbone atoms of well-ordered residues) of ARIAweb generated ensembles is consistently <1 Å indicating well-converged calculations (Table 1). PROCHECK (16) analysis shows that >95% of the residues are in the allowed regions of the Ramachandran plots as is expected for high-resolution NMR structures. More importantly, the bundle accuracy (mean pairwise RMSD between ARIAweb and PDB ensembles) is <1.6 Å for all tested entries except HypA, revealing that automated NMR structure calculation with ARIAweb yields structures very similar to the reference PDB structures (Figure 4). The lower accuracy (1.96 Å) observed for the HypA protein can be explained by the slightly different orientation between the Zn- and Ni-binding domains with regards to the reference PDB structure, for which the authors already noticed a variability (39).

## CONCLUSIONS AND FUTURE WORK

In a field where the amount of experimental data and sizes of studied macromolecules are continuously growing, improving the usability and accessibility of appropriate software plays an important role in speeding up research. The development of the new ARIAweb server tends to solve most of the limitations of a standalone installation of the ARIA tool, not only for accessibility but also for gathering the successive conversion, structure calculation and analysis steps in one place. Adapting the tool as a web service facilitates the interaction and eases the learning curve with the user base and the deployment of new features on a cross-platform environment. ARIAweb will serve as a useful portal to narrow the differences in technical know-how of users by helping them to manage and perform NMR structure calculation projects through the different stages without the need to install ARIA and its dependencies, or to have access to a computing cluster. ARIAweb will continue to be enhanced by integrating more experimental or *in silico* data and by improving the robustness of its automated NOE assignments procedure, e.g. through the use of consensus bundles (41). Finally, we are committed to make ARIAweb compliant with the new NMR Exchange Format (NEF) standard (42) for better integration with next-generation NMR analysis software (43) and streamlined deposition of atomic coordinates and NMR restraints in public databases.

## ACKNOWLEDGEMENTS

This work used the computational and storage services (TARS cluster) provided by the IT department at Institut Pasteur, Paris. We thank Drs Tru Huynh, Bertrand Néron and Mathias Ferber for their help. We are grateful to Dr Cameron Mackereth, Dr Rivka Isaacson and the late Dr Chris Spronk for providing us data for the RBM20, CsfB and MANEC test cases, respectively.

## FUNDING

French Institute of Bioinformatics (IFB) [ANR-11-INBS-0013]; INCEPTION project [PIA/ANR-16-CONV-0005]; European Union [FP7-IDEAS-ERC 294809 to M.N.]. Funding for open access charge: Institut Pasteur, Paris. *Conflict of interest statement.* None declared.

## REFERENCES

1. Rout, M.P. and Sali, A. (2019) Principles for integrative structural biology studies. *Cell*, **177**, 1384–1403.
2. Markwick, P.R., Malliavin, T. and Nilges, M. (2008) Structural biology by NMR: structure, dynamics, and interactions. *PLoS Comput. Biol.*, **4**, e1000168.
3. Guerry, P. and Herrmann, T. (2011) Advances in automated NMR protein structure determination. *Q. Rev. Biophys.*, **44**, 257–309.
4. 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.
5. Guntert, P. and Buchner, L. (2015) Combined automated NOE assignment and structure calculation with CYANA. *J. Biomol. NMR*, **62**, 453–471.
6. Guerry, P. and Herrmann, T. (2012) Comprehensive automation for NMR structure determination of proteins. *Methods Mol. Biol.*, **831**, 429–451.
7. Huang, Y.J., Tejero, R., Powers, R. and Montelione, G.T. (2006) A topology-constrained distance network algorithm for protein structure determination from NOESY data. *Proteins*, **62**, 587–603.
8. Linge, J.P., Habeck, M., Rieping, W. and Nilges, M. (2003) ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics*, **19**, 315–316.
9. Linge, J.P., O'Donoghue, S.I. and Nilges, M. (2001) Automated assignment of ambiguous nuclear overhauser effects with ARIA. *Methods Enzymol.*, **339**, 71–90.
10. Rosato, A., Bagaria, A., Baker, D., Bardiaux, B., Cavalli, A., Doreleijers, J.F., Giachetti, A., Guerry, P., Guntert, P., Herrmann, T. *et al.* (2009) CASD-NMR: critical assessment of automated structure determination by NMR. *Nat. Methods*, **6**, 625–626.
11. Mareuil, F., Malliavin, T.E., Nilges, M. and Bardiaux, B. (2015) Improved reliability, accuracy and quality in automated NMR structure calculation with ARIA. *J. Biomol. NMR*, **62**, 425–438.
12. Bardiaux, B., Malliavin, T. and Nilges, M. (2012) ARIA for solution and solid-state NMR. *Methods Mol. Biol.*, **831**, 453–483.
13. 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.
14. Brunger, A.T. (2007) Version 1.2 of the crystallography and NMR system. *Nat. Protoc.*, **2**, 2728–2733.
15. Linge, J.P., Williams, M.A., Spronk, C.A., Bonvin, A.M. and Nilges, M. (2003) Refinement of protein structures in explicit solvent. *Proteins*, **50**, 496–506.
16. Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., Kaptein, R. and Thornton, J.M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**, 477–486.
17. Hoof, R.W., Vriend, G., Sander, C. and Abola, E.E. (1996) Errors in protein structures. *Nature*, **381**, 272.
18. Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B. *et al.* (2018) MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.*, **27**, 293–315.
19. Nilges, M., Bernard, A., Bardiaux, B., Malliavin, T., Habeck, M. and Rieping, W. (2008) Accurate NMR structures through minimization of an extended hybrid energy. *Structure*, **16**, 1305–1312.
20. Bernard, A., Vranken, W.F., Bardiaux, B., Nilges, M. and Malliavin, T.E. (2011) Bayesian estimation of NMR restraint potential and weight: a validation on a representative set of protein structures. *Proteins*, **79**, 1525–1537.



21. 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.
22. Bardiaux,B., Bernard,A., Rieping,W., Habeck,M., Malliavin,T.E. and Nilges,M. (2009) Influence of different assignment conditions on the determination of symmetric homodimeric structures with ARIA. *Proteins*, **75**, 569–585.
23. Nilges,M. (1993) A calculation strategy for the structure determination of symmetric dimers by <sup>1</sup>H NMR. *Proteins*, **17**, 297–309.
24. Vranken,W.F., Boucher,W., Stevens,T.J., Fogh,R.H., Pajon,A., Llinas,M., Ulrich,E.L., Markley,J.L., Ionides,J. and Laue,E.D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, **59**, 687–696.
25. Bardiaux,B., Bernard,A., Rieping,W., Habeck,M., Malliavin,T.E. and Nilges,M. (2008) Graphical analysis of NMR structural quality and interactive contact map of NOE assignments in ARIA. *BMC Struct. Biol.*, **8**, 30.
26. Skinner,S.P., Goult,B.T., Fogh,R.H., Boucher,W., Stevens,T.J., Laue,E.D. and Vuister,G.W. (2015) Structure calculation, refinement and validation using CcpNmr Analysis. *Acta Crystallogr. D. Biol. Crystallogr.*, **71**, 154–161.
27. Nabuurs,S.B., Spronk,C.A., Vuister,G.W. and Vriend,G. (2006) Traditional biomolecular structure determination by NMR spectroscopy allows for major errors. *PLoS Comput. Biol.*, **2**, e9.
28. Mareuil,F., Doppelt-Azeroual,O. and Ménager,H. (2017) A public Galaxy platform at Pasteur used as an execution engine for web services. *F1000Research*, **6**, 1030, poster: not peer-reviewed.
29. Sloggett,C., Goonasekera,N. and Afgan,E. (2013) BioBlend: automating pipeline analyses within Galaxy and CloudMan. *Bioinformatics*, **29**, 1685–1686.
30. Afgan,E., Baker,D., Batut,B., van den Beek,M., Bouvier,D., Cech,M., Chilton,J., Clements,D., Coraor,N., Gruning,B.A. *et al.* (2018) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.*, **46**, W537–W544.
31. Rose,A.S., Bradley,A.R., Valasatava,Y., Duarte,J.M., Prlic,A. and Rose,P.W. (2018) NGL viewer: web-based molecular graphics for large complexes. *Bioinformatics*, **34**, 3755–3758.
32. Rose,A.S. and Hildebrand,P.W. (2015) NGL Viewer: a web application for molecular visualization. *Nucleic Acids Res.*, **43**, W576–579.
33. Rose,P.W., Prlic,A., Altunkaya,A., Bi,C., Bradley,A.R., Christie,C.H., Costanzo,L.D., Duarte,J.M., Dutta,S., Feng,Z. *et al.* (2017) The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res.*, **45**, D271–D281.
34. Selenko,P., Sprangers,R., Stier,G., Buhler,D., Fischer,U. and Sattler,M. (2001) SMN tudor domain structure and its interaction with the Sm proteins. *Nat. Struct. Biol.*, **8**, 27–31.
35. Liu,Z., Macias,M.J., Bottomley,M.J., Stier,G., Linge,J.P., Nilges,M., Bork,P. and Sattler,M. (1999) The three-dimensional structure of the HRDC domain and implications for the Werner and Bloom syndrome proteins. *Structure*, **7**, 1557–1566.
36. Martinez-Lumbreras,S., Alfano,C., Evans,N.J., Collins,K.M., Flanagan,K.A., Atkinson,R.A., Krysztowska,E.M., Vydyanath,A., Jackter,J., Fixon-Owuo,S. *et al.* (2018) Structural and Functional Insights into Bacillus subtilis Sigma Factor Inhibitor, CsfB. *Structure*, **26**, 640–648.
37. Upadhyay,S.K. and Mackereth,C.D. (2020) Structural basis of UCUU RNA motif recognition by splicing factor RBM20. *Nucleic Acids Res.*, **48**, 4538–4550.
38. Hong,Z., Nowakowski,M., Spronk,C., Petersen,S.V., Andreassen,P.A., Kozminski,W., Mulder,F.A. and Jensen,J.K. (2015) The solution structure of the MANEC-type domain from hepatocyte growth factor activator inhibitor-1 reveals an unexpected PAN/apple domain-type fold. *Biochem. J.*, **466**, 299–309.
39. Spronk,C., Zerko,S., Gorka,M., Kozminski,W., Bardiaux,B., Zambelli,B., Musiani,F., Piccioli,M., Basak,P., Blum,F.C. *et al.* (2018) Structure and dynamics of Helicobacter pylori nickel-chaperone HypA: an integrated approach using NMR spectroscopy, functional assays and computational tools. *J. Biol. Inorg. Chem.*, **23**, 1309–1330.
40. Rosato,A., Vranken,W., Fogh,R.H., Ragan,T.J., Tejero,R., Pederson,K., Lee,H.W., Prestegard,J.H., Yee,A., Wu,B. *et al.* (2015) The second round of critical assessment of automated structure determination of proteins by NMR: CASD-NMR-2013. *J. Biomol. NMR*, **62**, 413–424.
41. Buchner,L. and Guntert,P. (2015) Increased reliability of nuclear magnetic resonance protein structures by consensus structure bundles. *Structure*, **23**, 425–434.
42. Gutmanas,A., Adams,P.D., Bardiaux,B., Berman,H.M., Case,D.A., Fogh,R.H., Guntert,P., Hendrickx,P.M., Herrmann,T., Kleywegt,G.J. *et al.* (2015) NMR Exchange Format: a unified and open standard for representation of NMR restraint data. *Nat. Struct. Mol. Biol.*, **22**, 433–434.
43. Skinner,S.P., Fogh,R.H., Boucher,W., Ragan,T.J., Mureddu,L.G. and Vuister,G.W. (2016) CcpNmr AnalysisAssign: a flexible platform for integrated NMR analysis. *J. Biomol. NMR*, **66**, 111–124.