

NMR structure calculation with ARIA

Practical Course, 2016 Munich

B. Using CCPNmr

Benjamin Bardiaux and Michael Nilges

Unité de Bio-Informatique Structurale, Institut Pasteur, Paris, France

bardiaux@pasteur.fr; nilges@pasteur.fr

Abstract

In this practical, we will calculate the structure of the Tudor domain (Selenko P, Sprangers R, Stier G, Buhler D, Fischer U, Sattler M., 2001, Nat Struct Biol 8:27-31) with ARIA2. The data comprise two NOESY spectra, torsion angles from coupling constants, hydrogen bond restraints, and residual dipolar couplings.



1 Start ARIA

ARIA is written in the python language and need a valid python version with some particular packages as Numpy. For this practical, this installation is already done, but you can get more details about it on the ARIA web site <http://www.aria.pasteur.fr>.

First, setup the environment for ARIA:

1. Open a terminal window
2. Type the following command in the terminal

```
source /ms/data4/embo2015/ARIA/setup-aria.csh
```

3. Then, copy the ARIA example directory:

```
cp -r /ms/data4/embo2015/ARIA/example ./
```

4. Go to the `example` directory

```
cd example
```

5. Start ARIA

```
aria2 --help
```

This gives you an overview of the options.

2 Create your ARIA project

To create an ARIA project, you need:

- The sequence of your protein
- Chemical shifts assignments
- NOESY peaks-lists
- Additional NMR data if available

ARIA can import NMR data from various NMR formats (requires a conversion step) or directly from a CCPN project. The latter option is the one described in this tutorial.

ARIA provides a Graphical User Interface (GUI) to fill in the specified information (Input data and location of the output).

The GUI indicates where information are missing by a red exclamation mark . You also need to set parameters for ARIA and, if you want, to modify parameters for CNS (Figure 1).

1. Start the ARIA user interface (GUI) with the command

```
aria2 -g
```

2. Select `Project`→`New` in the menu to create a new project
3. Fill in the missing information in the `Project` panel
4. Enter `tudor` in the `File Root` field

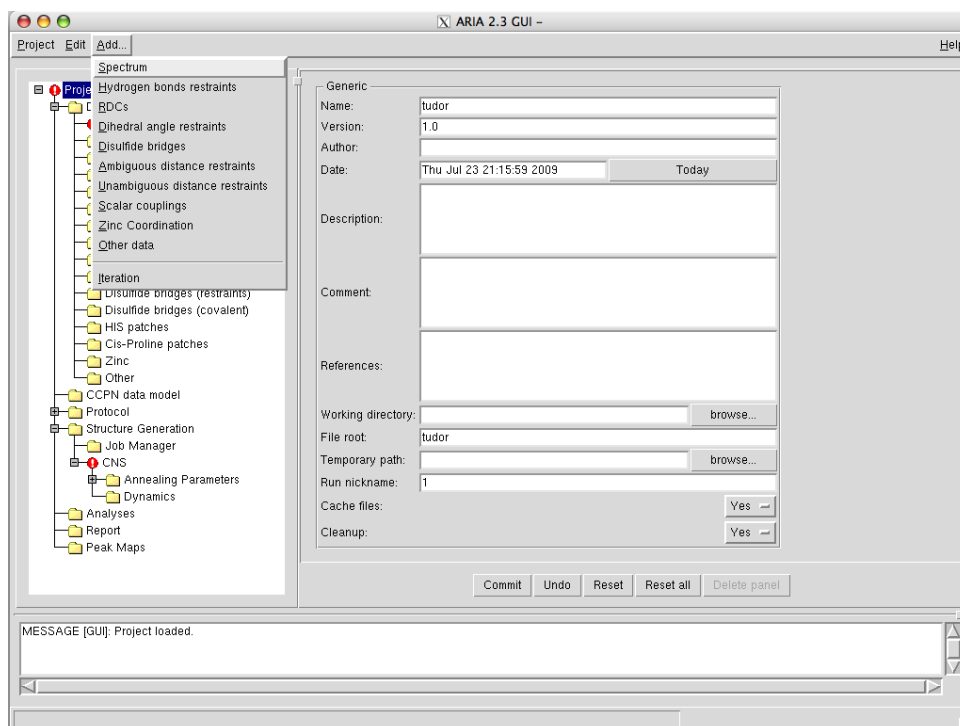


Figure 1: ARIA Graphical User Interface (GUI).

5. Set Working directory and Temporary path to the example/ directory in your home directory.
6. Select the CCPN project file. To do this, you need to fill the Project field in the CCPN data model node. Use the Browse button and select the ccpn/tudor directory that contains the CCPN project. *Note:* Do not worry for the warning pop-up window and just hit Ok.

Since data are taken from a CCPN project, you can also export ARIA results (calculated structures, restraints lists and NOE assignments) back to CCPN.

7. In the Options frame, set
 - Export restraints list** : Last iteration
 - Export assignments** : Yes
 - Export structures** : Yes
8. Select the molecular system (Figure 3). In the node Data → Molecular System, choose the format CCPN and use the Select... button to open a CCPN window, in which you can choose the Tudor Molecular System from the CCPN project.

Adding NOESY data

1. Add a NOESY spectrum panel by selecting Spectrum in the "Add..." menu.
2. In the new node Data → Spectra → #x, select the CCPN format for both the Chemical Shift List and the Spectrum.
3. With the Select... button, choose the ¹³C Shift List for the **Chemical Shift List**

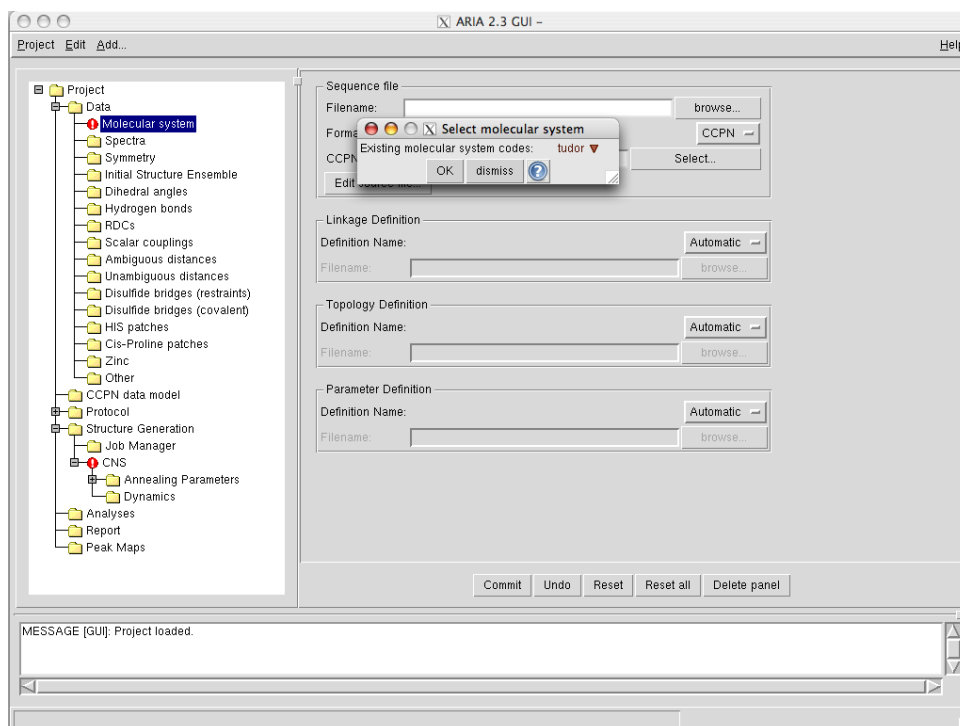


Figure 2: Use a CCPN project in an ARIA project.

4. Again with the `Select...` button, choose `13C_NOESY` as **Spectrum**
5. Repeat point 1 to 4 for the **15N 3D NOESY**.
6. Specify the chemical shift tolerances for the different NOE spectra (`freq. window`)
7. Leave the the default value (0.5 ppm) for the heteronuclear dimensions.
8. For the proton dimensions, use 0.04 ppm for the indirect dimension (`proton1`) and 0.02 ppm for the direct dimension (`proton2`).
9. For each spectra, turn off the use of the manual assignments by setting `Use manual assignment` to `No`. Otherwise, leave `Use manual assignment` to `Yes` if you want ARIA to use the NOE assignment already present in the peak list. You may also want to remove the diagonal peaks by enabling the `Filter diagonal peaks` option.

Adding other NMR data

The CCPN project contains also **hydrogen bonds** and **dihedral angle restraints**.

1. For each of them, add the corresponding restraints list with the “`Add...`” menu.
2. Select `CCPN` format and use the `Select...` button to locate them in the CCPN project (Note: as there is only one constraint list of each, the default one is automatically selected).
3. In our example, **RDC restraints** were not stored in the CCPN project. You can still add them in the ARIA project with the “`Add...`” menu and using the original data file `./data/rdcs/rdc2.tbl` (Figure 1).

4. Specify the alignment tensor for the RDC data. Go to Structure Generation → CNS → Annealing Parameters → RDCs → class 1. Rhombicity is 0.22 and magnitude 7.5.
5. Go to node Structure Generation → CNS and fill in with the path to the CNS program which is used to perform the actual structure calculation.

```
/ms/data4/embo2015/ARIA/cns.exe
```

6. Save this project into the file `aria_project.xml` in the `example/` directory (Menu Project→Save as...)
7. Quit the GUI (Menu Project→Exit)

Analysis parameters

- Structures quality can be analyzed with Whatif, Procheck, Prosa and Molprobit. One could enter the corresponding paths in the node Analyses. To save time, these analysis have been already performed for this example.
- To see a 2-dimensional “Peak Map” in the GUI after the structures calculation, you need to enable the `python pickle` output in the node Report.

Do not forget to **save** the project as `aria_project.xml` in the `example/` directory.

3 Running ARIA

1. Setup the ARIA project directory by typing

```
aria2 --setup aria_project.xml
```

This command will create the working directory (`run1/`) and copy all files needed to perform structures calculation.

2. To **speed up the calculation for the practical**, copy the pre-calculated structures to your directory tree:

```
cp -r /ms/data4/embo2015/ARIA/structures ./run1/
```

3. Then start ARIA with

```
aria2 aria_project.xml
```

First, the data are filtered for errors and inconsistencies. ARIA then proceeds with creation of the molecular topology. Afterwards a initial assignment of NOE cross-peaks is derived from the chemical shift lists. Finally, ARIA determines which files are missing and calculates them.

Time for a break ?

4 Analysing ARIA output

4.1 PDB files

For each iteration, you will find the PDB files of the generated structures in their respective directories `run1/structures/itxxx`.



Visualise the ensemble of solvent-refined structures with the program Pymol:

```
pymol run1/structures/refine/fitted_tudor_water.pdb
```

To show secondary structures, type the command **dss** in Pymol and then choose **S** → as → cartoon in the right menu. You can visualise all superimposed structures with **Movie** → Show all states.

4.2 Text output

ARIA generates various output files to report analysis results. For every iteration, the program creates the following report files:

1. `noe_restraints.unambig` and `noe_restraints.ambig`
These files tabulate unambiguous and ambiguous restraints, respectively. Restraints discarded by the merging procedure are excluded. For every restraint, information is given on its reference cross-peak, restraint bounds, the average distance found in the ensemble and the result of violation analysis.
2. `noe_restraints.violations`
Lists all violated restraints sorted with respect to their upper bound violations and contains the same information as (1).
3. `noe_restraints.assignments`
Lists primarily restraint-wise assignments and gives information whether the assignment(s) stem from fully, partially or unassigned cross-peaks.
4. `noe_restraints.merged`
Reports all restraints that have been discarded by the merging procedure.
5. `report`
Summarises analyses of the restraint lists and the structure ensemble.



Look at the report for the last iteration `run1/structures/it8/report`.

- *How many peaks remain ambiguous ?*
- *What is the precision of the final structure ensemble ?*

Quality checks. ARIA uses the programs WHAT IF, PROCHECK, PROSA II and MOL-PROBITY to evaluate the quality of both the final set of structures and the solvent-refined ensemble. For every program separate report files, `quality_checks.*`, are stored in the directories of the respective ensembles.



Compare the overall quality scores (`quality_checks`) for the last iteration (`run1/structures/it8/`) and for the water refinement (`run1/structures/refine/`).

Miscellaneous analyses. Several CNS scripts calculate restraint energies, ensemble RMSDs, and an average structure. Results are stored in `/analysis/`.



Take a look at the restraint violations statistics for distance restraints (`noes_disp`), dihedral angles (`dihedrals_disp`) and RDCs (`sani_disp`) in the `refine/analysis/` directory.

4.3 Graphics output

In the `it8/graphics` directory, you can find two PDF files (use `evince` to open them).

- The first one, shows the RMS violation by residue in 1- and 2-dimensional plots:
`run1/structures/it8/graphics/rms_analysis.pdf`
- The second one shows different WHAT-IF Z-scores along the protein sequence:
`run1/structures/it8/graphics/whatif_profiles.pdf`



Regions of the structures with lower quality can be easily detected from this plots

4.4 GUI output

This option is available only if you enable the `python pickle` output of the Report node. Open your `aria_project.xml` project file with the ARIA Gui (`aria2 -g aria_project.xml`) and open the node Peak Maps → `it8`.



You can see a contact map based on the NOE assignments and get details on inter-residue constraints by clicking on the crosses of the map.

4.5 CCPN output

ARIA can export results back to the CCPN project. You can visualise the content of the CCPN project with the program **CcpNmr Analysis**.

1. Start **CcpNmr Analysis** by typing

```
analysis
```

2. With Project → Open Project, select the ccpn/tudor project.
3. New NOE peak-lists containing assignments made by ARIA are visible in the Peak → Peak Lists window (Figure 3)
4. Distance restraints created by ARIA are listed in the Structure → Restraints and Violations window (Restraint Set 2)
5. Structures of the last iteration and of the water refinement are also present in the Structure → Structures window.

Figure 3: ARIA output in CCPNmr Analysis

5 Advanced exercises

1. Copy the advanced files to your home directory

```
cp -r /ms/data4/embo2015/ARIA/advanced ./
```

2. Go to the advanced/ directory.

```
cd advanced/
```

5.1 Spot the difference

The directory `run2/` contains the result of an ARIA calculation for the tudor domain, with the same data as in the tutorial. Here, a parameter has been changed in the definition of the ARIA protocol.



- Compare this result with the one obtained in `run1` (violations, ensemble precision, structure quality...)
- What parameter could explain the differences ?

5.2 Obvious mistake

The directory `run3/` contains the result of another ARIA calculation for the tudor domain. Yet, the result obtained in this case may not be very satisfactory :



- Why is this result not satisfactory ?
- If you think something went wrong, could you point it out ?
- What would be the possible origin(s) for this (deliberate) error ?

Links

ARIA <http://aria.pasteur.fr>

CCPN <http://www.ccpn.ac.uk>

CNS <http://cns-online.org>

WHAT IF <http://swift.cmbi.ru.nl/whatif/>

PROCHECK <http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/>

MolProbity <http://molprobity.biochem.duke.edu/>

Further reading

- [1] B Bardiaux, A Bernard, W Rieping, M Habeck, TE Malliavin, and M Nilges. Graphical analysis of NMR structural quality and interactive contact map of NOE assignments in ARIA. *BMC Struct Biol*, 8(1):30, Jun 2008.
- [2] Benjamin Bardiaux, Thérèse E Malliavin, and Michael Nilges. ARIA for Solution and Solid-State NMR. *Methods Mol Biol*, 831:453–483, 2012.
- [3] Aymeric Bernard, Wim F Vranken, Benjamin Bardiaux, Michael Nilges, and Thérèse E Malliavin. Bayesian estimation of NMR restraint potential and weight: A validation on a representative set of protein structures. *Proteins: Structure, Function, and Bioinformatics*, January 2011.
- [4] M. Habeck, W. Rieping, J. P. Linge, and M. Nilges. *NOE assignment with ARIA 2.0: the nuts and bolts.*, volume 208 of *Methods in Molecular Biology*, pages 379–402. Humana Press, Totowa, NJ 07512, USA, 08 2004.
- [5] J P. Linge, M A Williams, C A Spronk, A M Bonvin, and M Nilges. Refinement of protein structures in explicit solvent. *Proteins Struct. Funct. Genet.*, 20(3):496–506, 2003.

- [6] Fabien Mareuil, Thérèse E Malliavin, Michael Nilges, and Benjamin Bardiaux. Improved reliability, accuracy and quality in automated NMR structure calculation with ARIA. *J Biomol NMR*, April 2015.
- [7] Michael Nilges, Aymeric Bernard, Benjamin Bardiaux, Thérèse Malliavin, Michael Habeck, and Wolfgang Rieping. Accurate NMR structures through minimization of an extended hybrid energy. *Structure*, 16(9):1305–1312, September 2008.
- [8] W Rieping, M Habeck, B Bardiaux, A Bernard, TE Malliavin, and M Nilges. ARIA2: automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics*, 23(3):381–382, 2007.
- [9] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, and E. D. Laue. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, 59(4):687–696, Jun 2005.

6 Answers for advanced exercises

6.1 Spot the difference

Here, we used a **log-harmonic potential** instead of the classical *flat bottom harmonic wall* potential for the distance restraints. In addition, the optimal weight applied on the distance restraints was determined automatically. The final average weight for the NOE restraints is $10.5 \text{ kcal.mol}^{-1}$ (where it was $50 \text{ kcal.mol}^{-1}.\text{\AA}^{-2}$ with the flat potential). The general quality of the structures is better with fewer clashes. See Nilges *et al.* Structure, 2008 and Mareuil *et al.* J Biomol NMR, 2015 for details.

6.2 Obvious mistake

Here, we deliberately exchanged assignments of Cys20 and Cys45. Moreover, we asked ARIA to use the manual NOE assignments and to keep them until the end of the calculation. The resulting structures failed to converge due to the incoherent restraints with the two cysteines. One can detect such errors by looking at the file `rms_analysis.pdf` for instance (large peaks for residues 20 and 45).