

NMR structure calculation with ARIAweb and an initial 3D model

In this practical, we will calculate the structure of a **HRDC domain** ([Liu et al., 1999, Structure 15:7\(12\):1557-66](#)) with ARIA2. The data comprise two NOESY spectra, torsion angles from 3J coupling constants and hydrogen bond restraints.

Links:

- [ARIA website](#)
- [ARIAweb documentation](#)

IMPORTANT For this tutorial, you will use the ARIAweb server, for which you can either Register or use anonymously. Please go first to the login page and choose how to use the server. <https://ariaweb.pasteur.fr/login>

Once you have access to ARIAweb, here is how your homepage looks like.

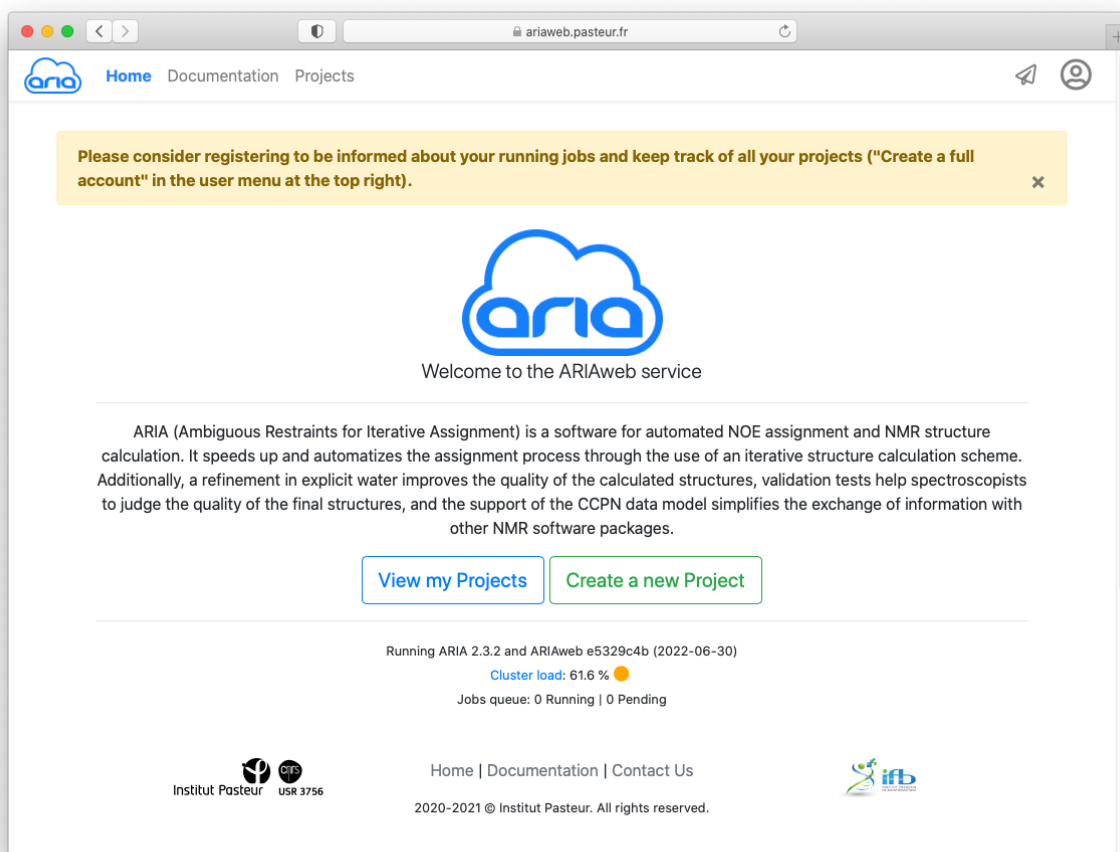


Figure 1: ARIAweb homepage

Intro:

In this tutorial, we will see how to use a starting model (e.g. from a previous calculation or from AlphaFold prediction) to help assigning NOEs. In turn the starting model(s) will mimic a previous iteration before iteration 0. Distances for possible NOE assignment will be measured used the coordinates the starting model(s) and used for calibration, elimination of spurious cross-peaks and unlikely assignment possibilities.

Outline

- [1. Generate an AlphaFold model of the protein](#)
- [2. Create an ARIAweb Project](#)
- [3. Convert the data](#)
- [4. Prepare your ARIA structure calculation project](#)
- [5. Analysing results](#)
- [6. Advantage of using an input model](#)

1. Generate an AlphaFold model of the protein

Note This step requires to be logged in with a Google account. If you prefer to skip this step, you can download a **pre-calculated model** [here](#)

To predict a 3D model of our protein of interest, we will use this [ColabFold notebook](#).

In the `query_sequence` field, input the one-letter sequence of the construct used for NMR experiment, *i.e.* including the expression tag:

```
MKHHHHHPMELNNLRMTYERLRELSLNLGNRMVPPVGNFMPDSILKKMAAILPMNDSAFATLGTVEDKYRRRFKYFK  
ATIADLSKKRSSE
```

In the `jobname` field, enter `hrdc` and set `num_relax` to `1`. Since we will use this model to evaluate inter-proton distances from NOE cross-peaks assignments, it is necessary to have a protonated model at the end and that is exactly what the relaxation with amber does.

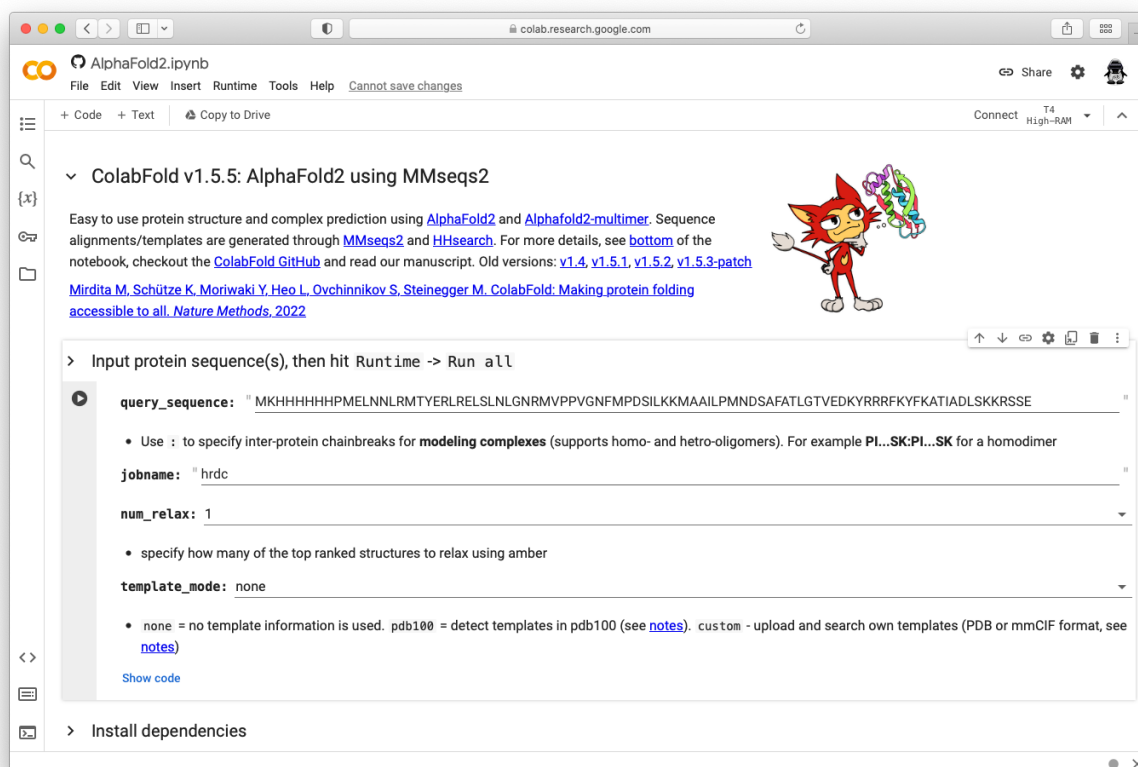


Figure 2: ColabFold notebook

Next, in the top menu of the ColabFold notebook, click `Runtime > Run all`. In the popup window, click `Run anyway`.

If you scroll down to the `> Run Prediction` section, you will see the predicted models as they are generated by AlphaFold.

After a few minutes, when all 5 models are generated, the notebook will generate a ZIP file (xxx.results.zip) to be saved. It will contains the predicted models along with some statistics. After unzipping the ZIP file, you will get the following files:

```
hrdc_5687f/  
|__cite.bibtex  
|__config.json  
|__hrdc_5687f.a3m  
|__hrdc_5687f.csv  
|__hrdc_5687f.done.txt  
|__hrdc_5687f_coverage.png  
|__hrdc_5687f_env  
|__hrdc_5687f_pae.png  
|__hrdc_5687f_plddt.png  
|__hrdc_5687f_predicted_aligned_error_v1.json  
|__hrdc_5687f_relaxed_rank_001_alphafold2_ptm_model_3_seed_000.pdb  
|__hrdc_5687f_scores_rank_001_alphafold2_ptm_model_3_seed_000.json  
|__hrdc_5687f_scores_rank_002_alphafold2_ptm_model_5_seed_000.json  
|__hrdc_5687f_scores_rank_003_alphafold2_ptm_model_4_seed_000.json  
|__hrdc_5687f_scores_rank_004_alphafold2_ptm_model_2_seed_000.json  
|__hrdc_5687f_scores_rank_005_alphafold2_ptm_model_1_seed_000.json  
|__hrdc_5687f_unrelaxed_rank_001_alphafold2_ptm_model_3_seed_000.pdb  
|__hrdc_5687f_unrelaxed_rank_002_alphafold2_ptm_model_5_seed_000.pdb  
|__hrdc_5687f_unrelaxed_rank_003_alphafold2_ptm_model_4_seed_000.pdb  
|__hrdc_5687f_unrelaxed_rank_004_alphafold2_ptm_model_2_seed_000.pdb  
|__hrdc_5687f_unrelaxed_rank_005_alphafold2_ptm_model_1_seed_000.pdb  
|__log.txt
```

The output files are:

- **log.txt**: log file that contains the scores (ptm, plddt etc)
- ***.pdb**: PDB predictions for the 5 models, ranked by pLDDT
- ***.json**: statistics of each model in json format
- **hrdc_5687f_coverage.png**: coverage of MSA
- **hrdc_5687f_plddt.png**: pLDDT score along the sequence
- **hrdc_5687f_pae.png**: PAE plot for the 5 models
- **hrdc_5687f.a3m**: MSA generate with MMseqs2 used for prediction

The model we will use for the rest of the tutorial is

hrdc_5687f_relaxed_rank_001_alphafold2_ptm_model_3_seed_000.pdb. It is the best ranked model and the one that has been relaxed and protonated with Amber.

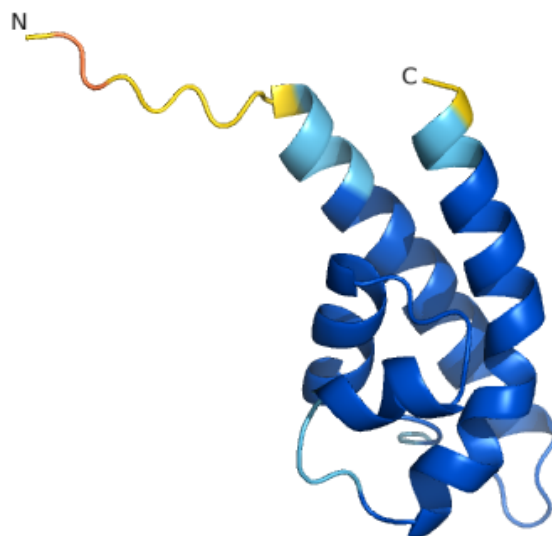


Figure 3: ColabFold model of the the HRDC domain colored by pLDDT

2. Create an ARIAweb Project

Before all, you will need to create a new project. First, go the ARIAweb homepage <https://ariaweb.pasteur.fr/> and log in.

1. Click on the Create a new Project button.
2. Specify the "Project name", e.g. *HRDC*
3. Click on the Save button.

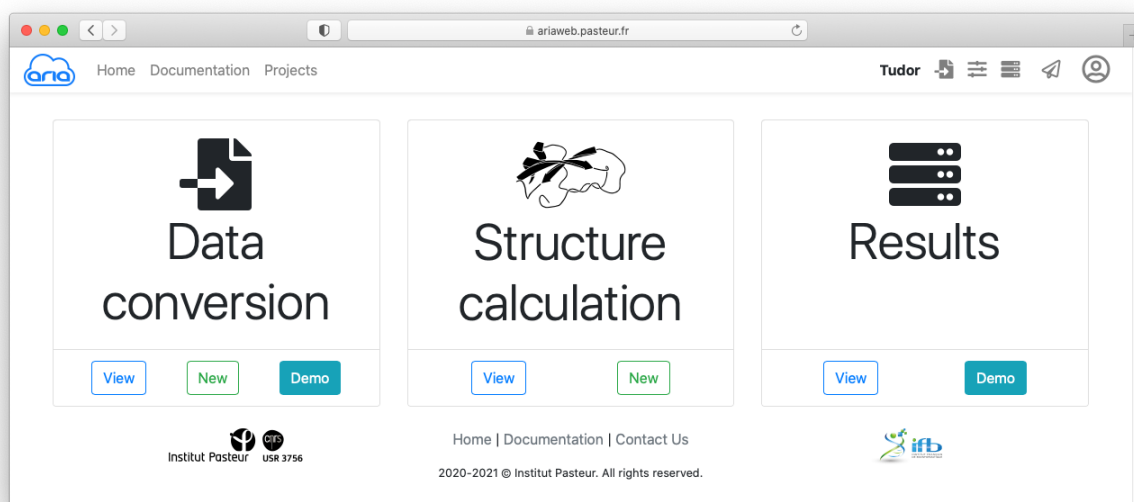


Figure 4: Project menu

3. Convert the data

To run ARIA, you will need:

- The sequence of your protein
- Chemical shifts assignments
- NOESY peaks-lists

ARIA can import NMR data from **various NMR formats (requires a conversion step)** or directly **from a CCPN project**. The first option is the one described in this part of the tutorial, by first converting the data from various format to the internal ARIA XML format.

Since the original data are not stored in ARIA XML format, the first thing that needs to be done is to convert the NOE spectrum files from the original XEASY format to the ARIA XML format. The conversion routines check the data for consistency, and convert the atom names to strict IUPAC convention. ARIA uses the XML format to describe and store most of the crucial input files.

Create a conversion project

1. Under **Data Conversion**, click the New button.
2. Specify the name for this conversion step, e.g. *xeasy*
3. Specify the name of the molecule, e.g. *hrdc*
4. Click on the Save button.

Load the tutorial data

For this tutorial, we will load the corresponding data simply by clicking on the Load demo data button.

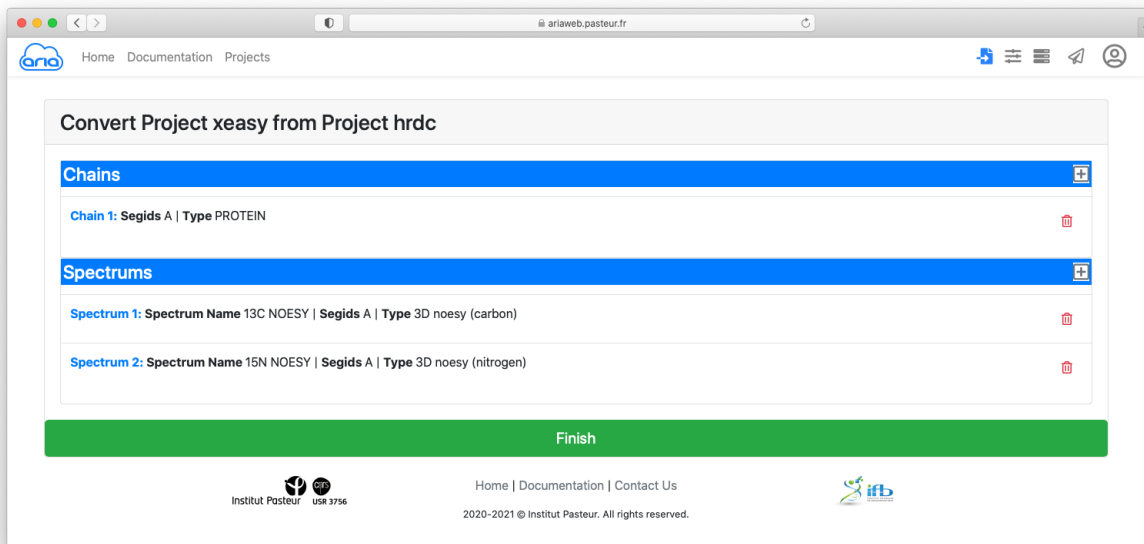


Figure 5: conversion project with sequence and spectra data

You can see that the data contain:

- a molecular chain definition
- two 3D NOE spectra (more precisely, NOE peak-lists) with the associated chemical shift list.
 - ¹³C-resolved NOESY
 - ¹⁵N-resolved NOESY

By clicking on each item (in blue), you can see details that must specified for conversion from Xeasy format to ARIA XML format.

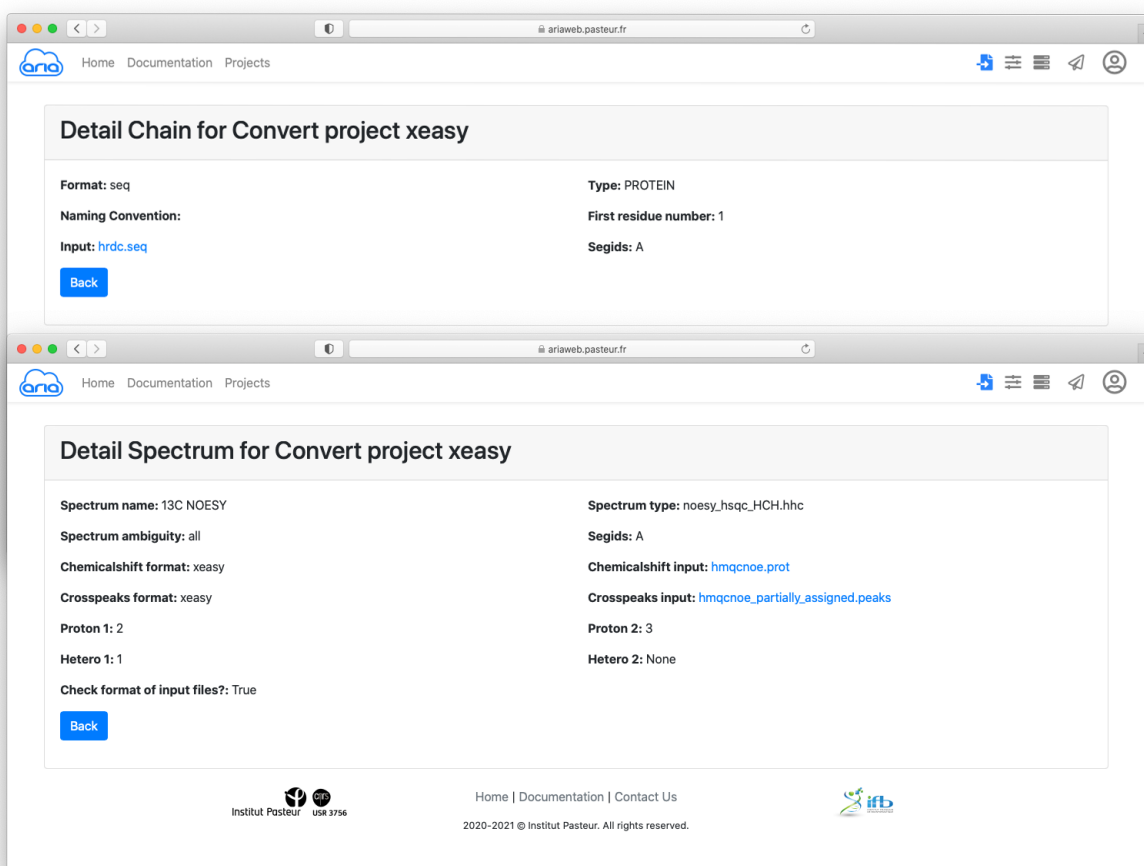


Figure 6: detailed view on Chain and Spectrum items for conversion

You can also have a look at the input files by clicking on the filenames (in blue) for the sequence, chemical shifts and cross-peak lists.

After clicking the **Back** button, you can save the conversion project by clicking on **Finish** (Figure 5).

Submit conversion

You will be then redirected to the page where you can submit the conversion data conversion on the server.

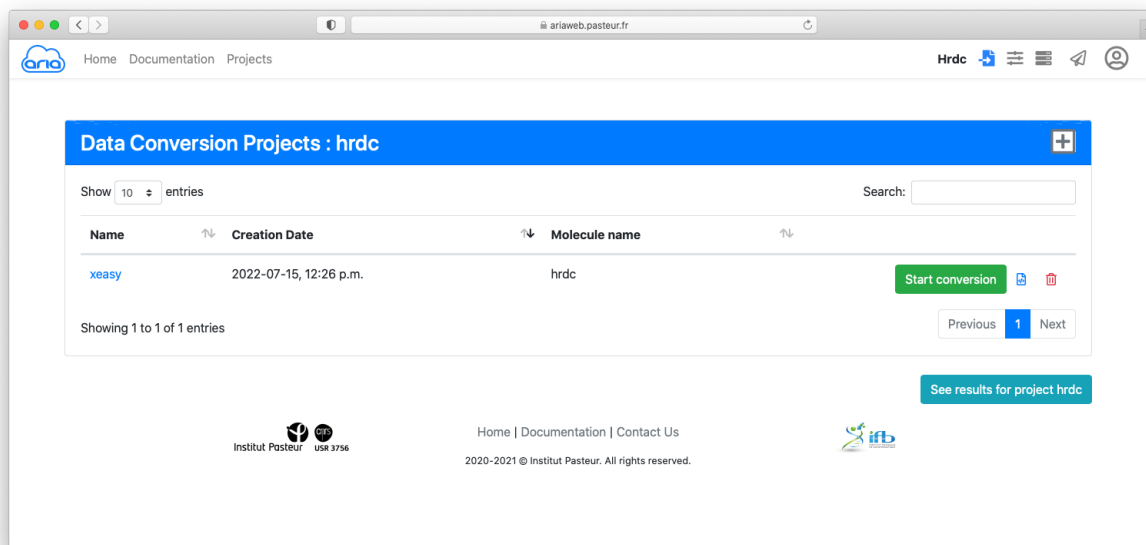


Figure 7: List of conversions ready to be submitted

Click the **Start conversion** button to submit your data conversion. You will be redirected to a *Results* page where you can see the status of your conversion job. See [here](#) for an explanation on the status icon (Figure 8).

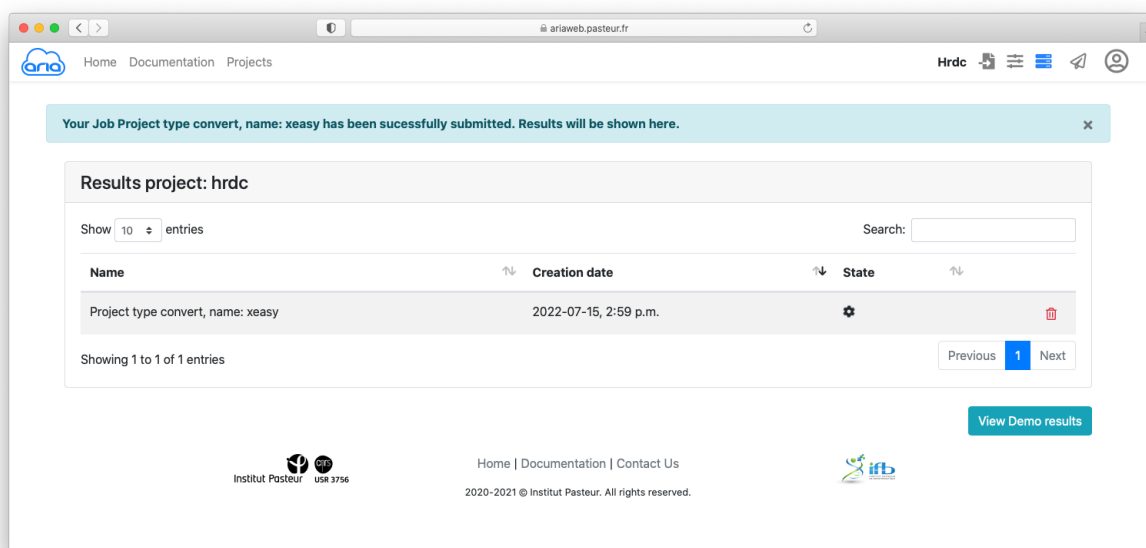





Figure 8: Results page where conversion is running

The status icon  indicates that the conversion job is running. Once the job is finished (status icon ) and if needed, you can download the convert files using the  icon (Figure 9).

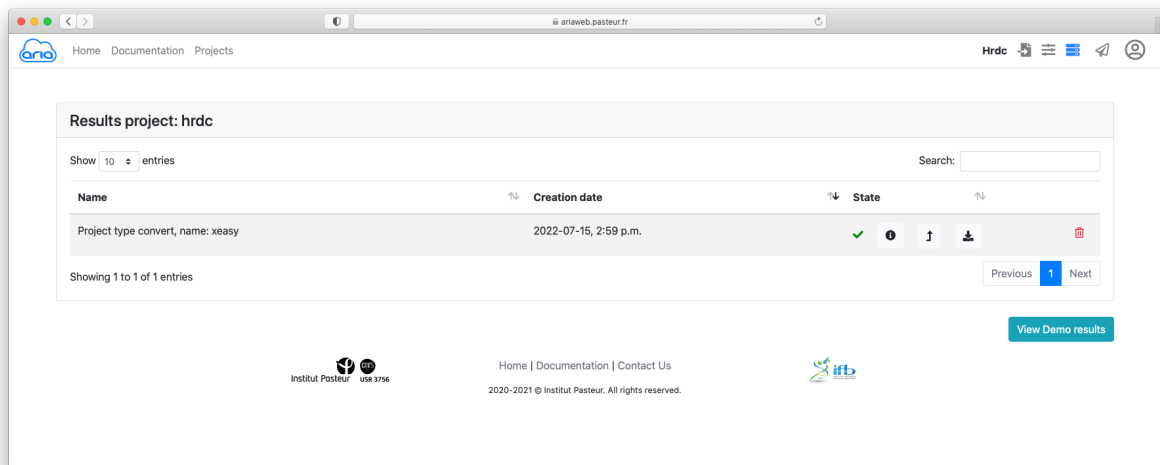



Figure 9: Conversion finished, ready to be downloaded or to start preparing a structure calculation

4. Prepare your ARIA structure calculation project

Now that you have converted your sequence and NOE data, you can prepare your ARIA structure calculation. ARIAweb allows to directly create a structure calculation project using the convert data by simply clicking on the  button (Figure 9).

You now have a new form where you will be able to validate or modify all the parameters for your ARIA run. The Structure Calculation form 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 4 main categories of parameters are:

- **SETUP:** name of the structure calculation job and other generic info
- **DATA:** all input data must be selected/uploaded here
- **PROTOCOL:** parameters related to the ARIA iterative protocol
- **STRUCTURE CALCULATION:** parameters related to generation of structures with [CNS](#)

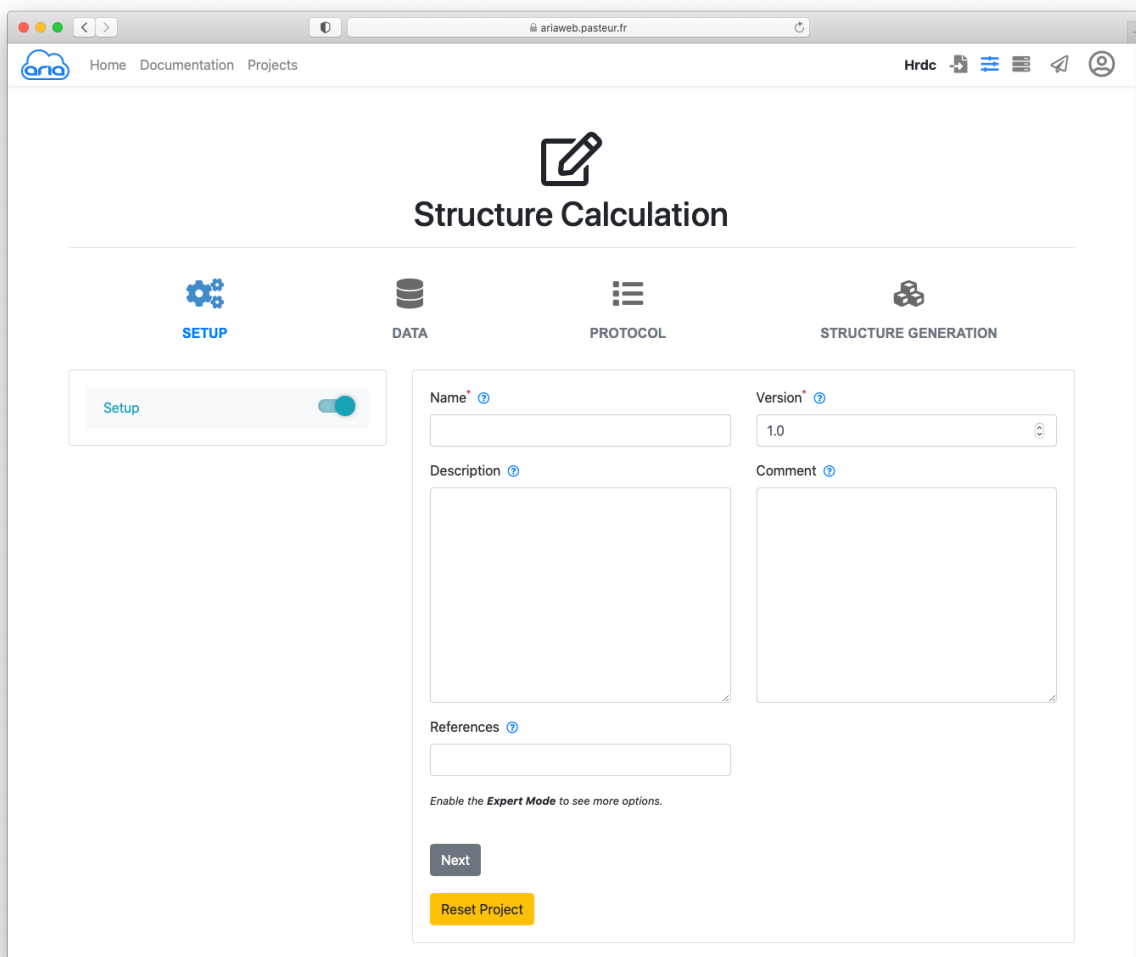


Figure 10: SETUP page of the structure calculation form

Start by giving a name to this ARIA job, e.g. *hrdc*.

Click the **Next** button to move to the next section of the form.

In the **DATA** section, you will see that the molecular system is already pre-filled with your converted file from the previous step. Click **Next** to move to the Spectra definitions.

NOE peak-list and chemical shifts

Again, your ^{13}C -NOESY and ^{15}N -NOESY peak-lists are already specified as 2 spectrum items with the files you converted before.

For the 2 spectra (scroll down to access Spectrum 2), uncheck the **Use assignments** box. Here we will not use any previous assignments (from manual analysis of the peaks).

Leave all parameters as default.

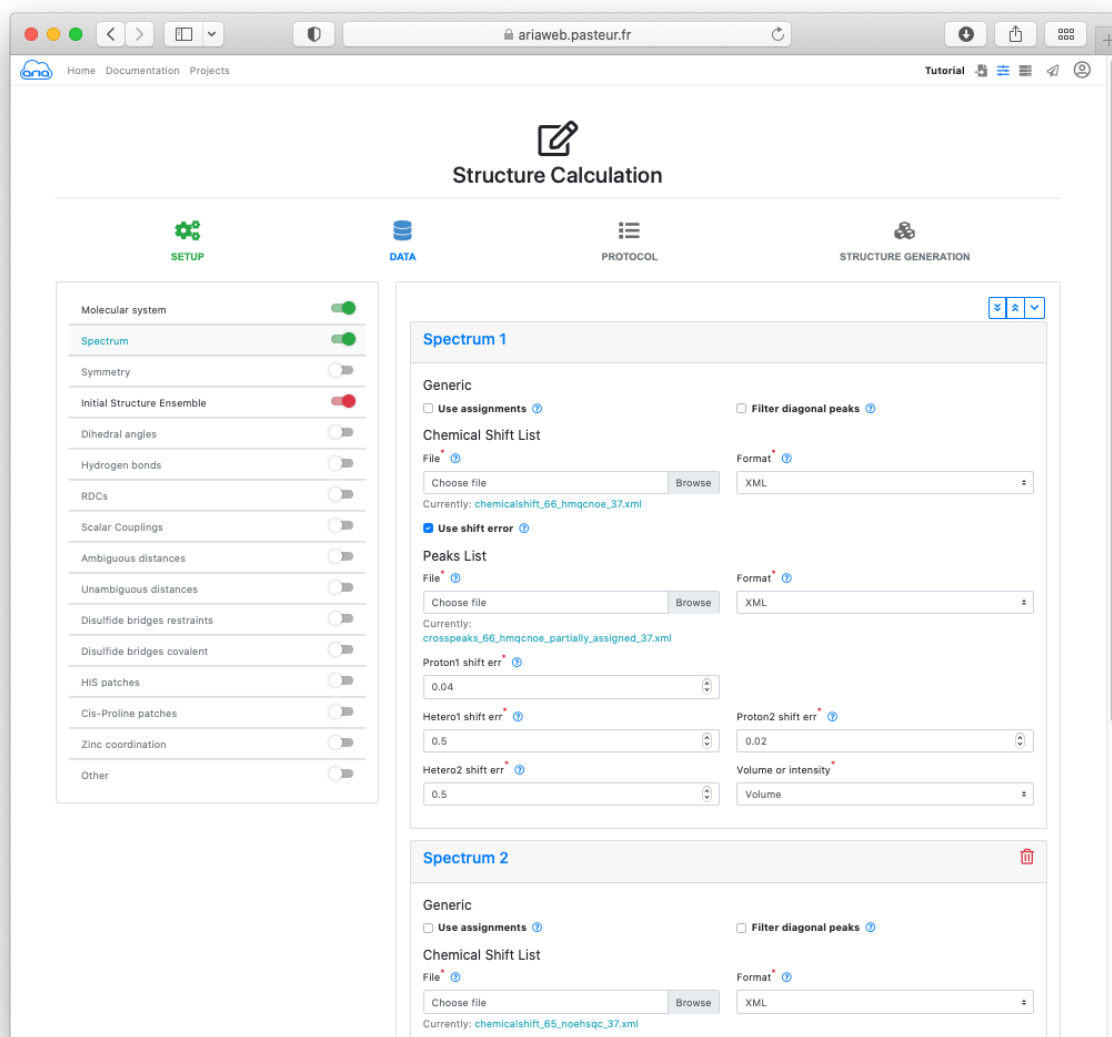


Figure 11: Spectra parameters

Adding the input model

We will use the AlphaFold model to help the initial NOE assignment. To do so, activate the slider on the left for *Initial Structure Ensemble* (Figure 12).

Click the **Next** button to move to *Initial Structure Ensemble*.

Now click the **Browse** button and select the AlphaFold model ([hrdc_5687f_relaxed_rank_001_alphafold2_ptm_model_3_seed_000.pdb](#)) so that it can be uploaded. Make sure to select the **relaxed** PDB file.

Select **IUPAC** for the Format.

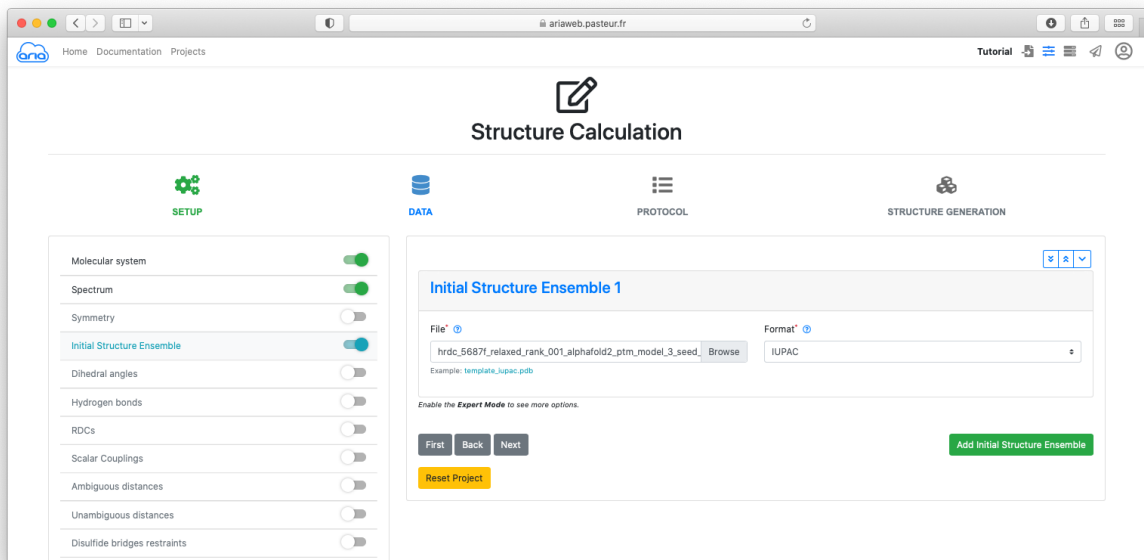


Figure 12: Initial ensemble form

Click the **Next** button to move to the *PROTOCOL* section.

Adapting the assignment parameters for early iterations

Since we will be using an initial input model, we need to adapt some parameters related to the stringency of the automated assignment method. To do so, you must enable the **Expert mode** from your User menu (👤 icon, top right).

Activate the slider on the left for *Iteration settings* (Figure 13).

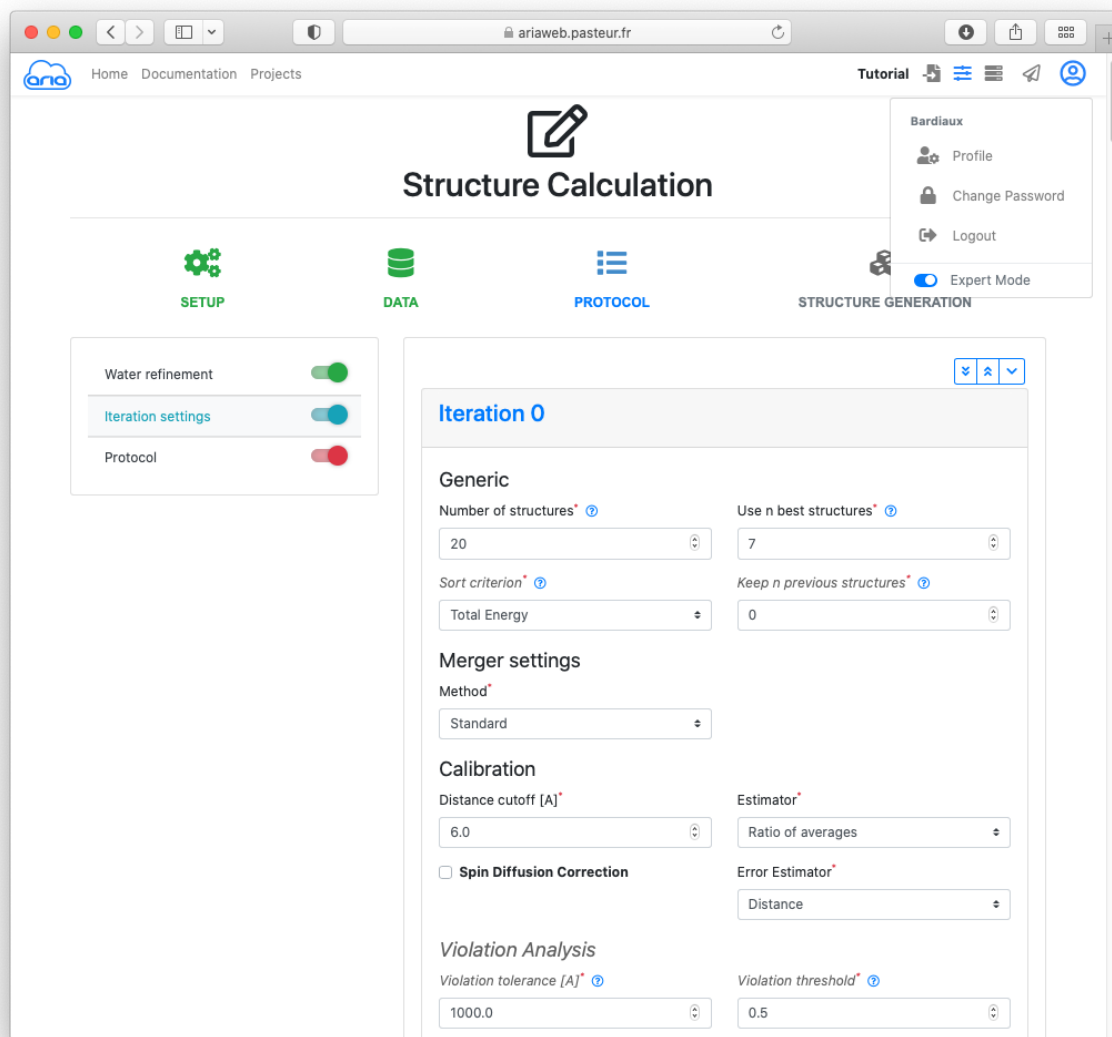


Figure 13: Expert mode and Iteration Settings

On the **Water refinement** page, click the **Next** button to move to the *Iteration settings*.

For the first 3 iterations, we will change the following parameters:

- **Violation tolerance** in the *Violation Analysis* section: when lowering this value (in Angstroms), you are reducing the tolerance for considering a peak as true
- **Ambiguity cutoff** in the *Partial Assignment* section: when decreasing this value, it will discard the less likely assignment possibilities

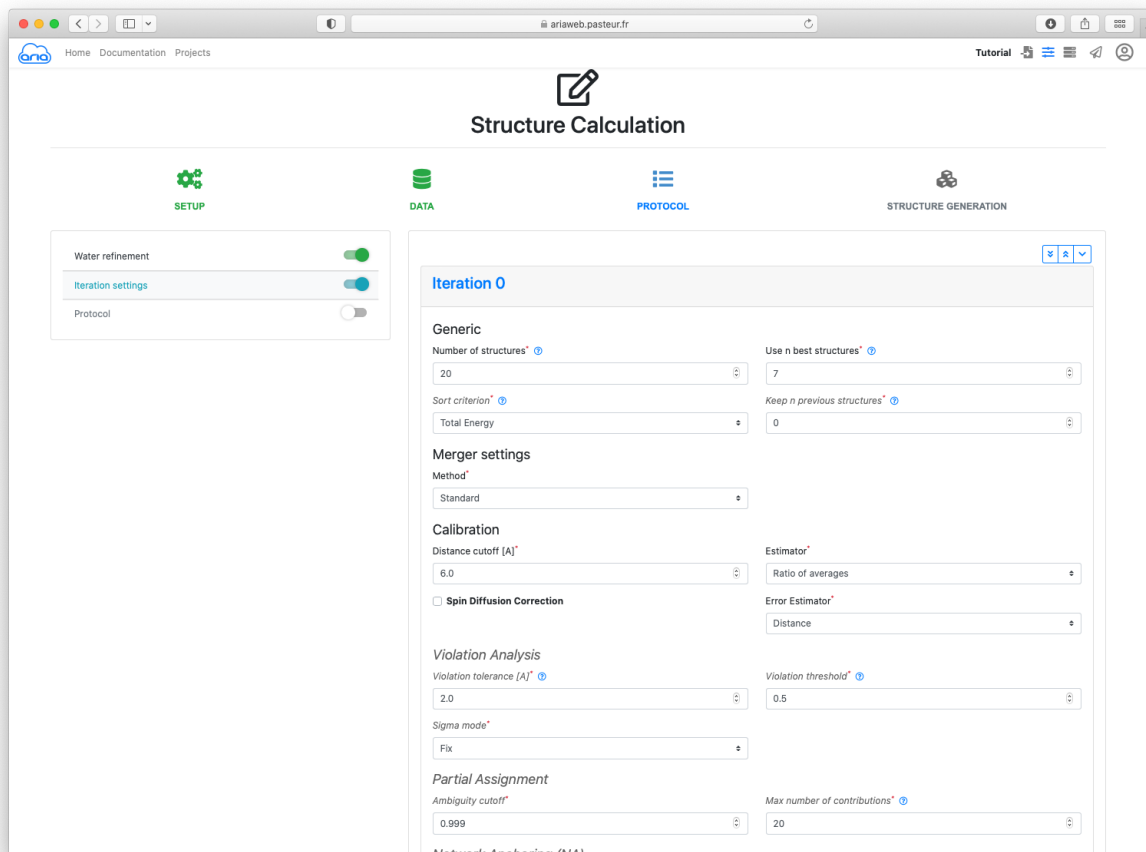


Figure 14: Adapted parameters in Iteration Settings

Set the following values for Iteration 0, 1 and 2 (scroll down to access other Iterations):

Iteration	Violation tolerance	Ambiguity cutoff
0	2.0	0.999
1	1.0	0.99
2	1.0	0.99

Scroll down to the bottom of the page and click the [Next](#) button again.

Submitting your ARIA calculation

For the practical, we use all parameters as default for **STRUCTURE GENERATION**. Continue clicking [Next](#) until you see the [Save](#) button.

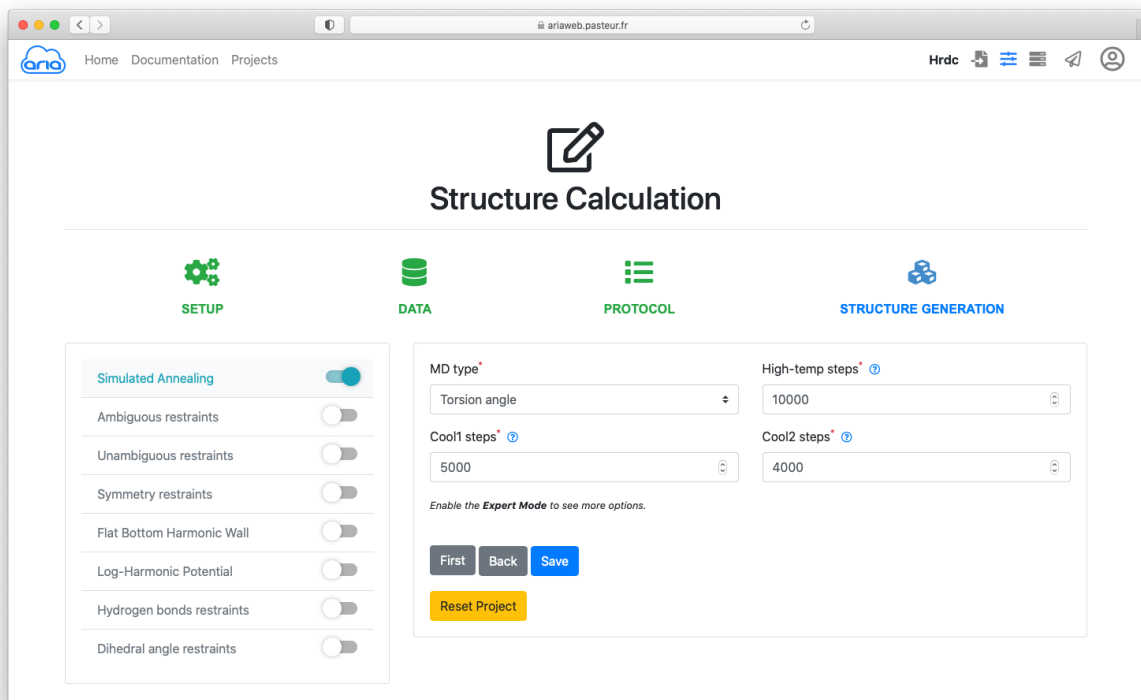


Figure 15: A Structure calculation form ready to be saved

Clicking **Save** will save (no surprise here) your parameters and you will be directed to the page where can submit the ARIA calculation job.

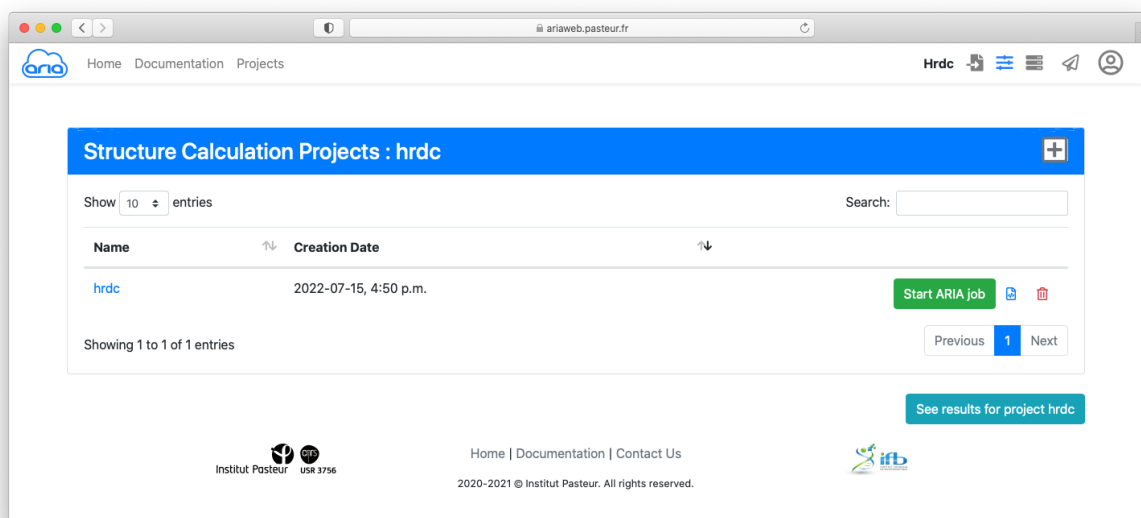



Figure 16: List of structure calculations ready to be submitted

Click the **Start ARIA job** button to submit you structure calculation. You will be redirected to a *Results* page where you can see the status of your ARIA job. See [here](#) for an explanation on the status icon.

Depending on the server load the calculation might take ~30 minutes to hours... For the rest of the practical you will not have to wait until the calculation is finished. We have pre-calculated results that will allow you to visualise and analyse the results of an ARIA job.

5. Analysing results

When your job is finished (status icon ) , click on the  button to start the job analysis and visualize the results.

Results archive can be downloaded using  .

**Tutorial results are already available so you don't have to wait.
Click [HERE](#) to go to the results page.**

The visualisation page allows to view an analysed the results of a structure calculation job. Final structure ensemble generated by ARIAweb are displayed interactively, with various representations. Final restraints statistics, structure quality checks and bundle RMSD are shown to help the user interprets the reliability of the results. On top of that, more graphs, restraints validation and PDB files can be downloaded directly (Figure 17).

Tutorial Report

Iterations tab →

IT 0 IT 1 IT 2 IT 3 IT 4 IT 5 IT 6 IT 7 IT 8 REFINE

Structures

- fitted_it8
7 models
- Sec. structure
- RMS NOE viol
- RMSF bundle
- Representation
Add a new representation

NGL Viewer

Viewer

Representations

Restrains statistics

USED RESTRAINTS

1698

↑ + 8 Since last iteration

VIOLATED RESTRAINTS

47

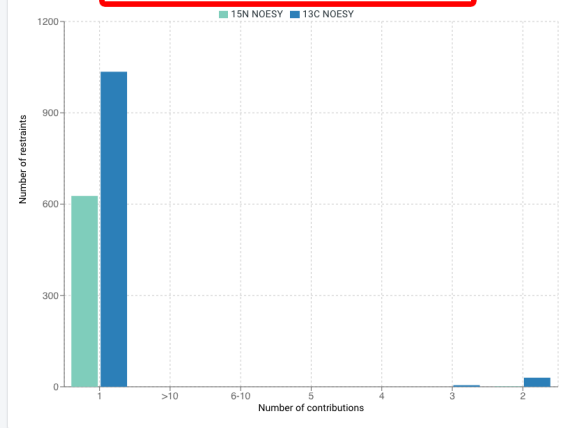
↓ - 12 Since last iteration

MERGED RESTRAINTS

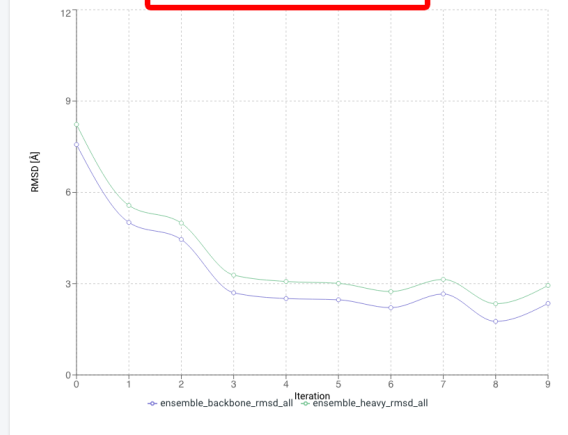
600

↓ - 0 Since last iteration

Contribution histogram



Ensemble RMSD



Quality checks

PROCHECK


- Most favoured regions : 74.8
- Allowed regions : 21
- Generously allowed regions : 3.53
- Disallowed regions : 0.686
- Labelled residues (all Ramachandrans) : 11.1
- Labelled residues (Chi1-chi2 plots) : 1.57
- Bad contacts : 4.86
- G-factor dihedrals : -0.253
- G-factor covalent : 0.633
- G-factor overall : 0.0829

More results

- PDB ensemble
- Per residue WHATCHECK graph
- RMS of NOE violations graph
- NOE restraints assignments
- NOE restraints violations
- Full ARIA run archive

Figure 17: Screenshot of a typical Visualisation page, where main components are marked in red.


Results for all ARIA iterations can be shown by selecting an iteration in the **Iterations tab**. The main component of the Visualisation page is the **NGL viewer**. By default, several representations of the structure ensemble generated by ARIA are shown:



- **Sec. structure** is a cartoon representation with secondary structure elements (**helices** and **strands**).
- **RMS NOE viol** shows side-chains colored by the average RMS (Root Mean Square) of NOE violations (per residue) from **low RMS (good)** to **high RMS (bad)** (not available for *Refine* iteration).
- **RMSF bundle** is a sausage representation with a radius proportional to the Root Mean Square Fluctuation (per residue) in the ensemble (hidden by default). More representations can be added with the  button.



Is the NOE violation statistics for residues better at iteration 2 (IT 2) or the last iteration (IT 8) ?

(Hint: Look at the color of side-chains)

The **Representation tab** (below the NGL viewer) allows to change the visibility, styling and coloring of a selected representation. By default, all structures in the bundle are shown; use the  button to select individual models for display).

The  button allows to upload another PDB file to superimpose on the ARIA structure ensemble (use the  button to superimpose on a selected structure).

Below, **Restrains statistics** recapitulate the number of restrains and the trend since the previous iteration for:

1. Restrains used for structure calculation (the more, the better)
2. Unsatisfied (i.e. consistently violated) restrains (less is better)
3. Merged (identical) restrains



How many NOE cross-peaks remain unassigned ?

How many restrains were used for structure calculation ?

The **Contribution histogram** gives count of restrains (from each input spectra) that have 1 (i.e. unambiguous) or more (i.e. ambiguous) assignment possibilities (or "contributions"). Ultimately, ARIA tries to reduce the ambiguity in NOE assignments, producing more unambiguous restrains. Truly ambiguous assignments can remain due to spectral overlap or chemical shifts degeneracy.

The **Ensemble RMSD** graph shows, for each iteration, the RMSD of the structure ensemble generated by ARIA ("bundle"), computed as the mean RMSD when superimposing of the ensemble average (after iterative superimposition). Low RMSD (< 1-2 Å) is good indicator of convergence of the NOE assignment and structure calculation process by ARIA.



How the ensemble precision evolves during the ARIA iterative process ?

Is there a convergence at the 1st iteration ? At the last iteration ?

What is the precision of the final structure ensemble ?

(Hint: Look at the "Summary table" under **More results**)

The **Quality checks** panel summarizes the results of 3 main structural quality validators: [WHAT-IF](#), [Procheck](#) and [Molprobability](#). Quality scores are shown on a slider from **bad** to **good** values. Bad quality score values indicate that the input data may contain errors/inconsistencies and that ARIA was not able to produce a high quality model. We provide here some indicators on how to judge the quality checks:

1. Procheck Ramachandran percentage: for typical NMR structures deposited in the PDB, 80% of the dihedral angles lie within the preferred region of the Ramachandran plot. For high-resolution NMR structures, a higher percentage is expected (90%).
2. WHAT-IF Z-scores: WHAT-IF results are presented in the form of overall Z-scores. In general, structures with Z-scores between -2 and +2 are considered to be within a normal range and are thus good structures, while structures with Z-scores lower than -2 should be inspected further. Useful indicators of good quality are *Backbone conformation* and *Packing quality*. The *bump-score* also reports the number of van der Waals violations per 100 residues.
- 3.

WHAT-IF profiles: recently, some studies have stressed that global structural indicators are not sufficient to detect errors in structures and suggested examining parameters on a per-residue basis. Such profiles for the WHAT-IF scores are produced by ARIA in the form of a PDF file. Thus, poor quality regions can be precisely identified.

4. Molprobability clashscore. this reports the number of overlaps $>0.4\text{\AA}$ per thousand atoms. For typical NMR structures deposited in the PDB, this score is generally high (>10). From our experience, the application of the log-harmonic potential along with automated weight estimation significantly improves this situation (see the [More results](#) panels).



Compare the overall quality scores (Quality checks) for the last iteration (IT8) and for the water refinement (REFINE).

The [More results](#) panel provides a link to a **standard summary table of structural and restraints statistics** (iteration 8 and refine)



Take a look at the restraint violations statistics for NOE restraints in the "Summary table"

Additional results are also accessible:

- PDB file of the structure ensemble
- graph of NOE restraints violations RMS (PDF)
- lists of NOE restraints assignments and violations (text)
- graph of WHATCHECK scores per residue (PDF, it8 and refine)
- a full ARIA run archive (ZIP)



Look at the "Per residue WHATCHECK graph". Do you spot regions of lower quality ?

6. Advantage of using an input model

If we would not have used the input model, the results would have been quite different. You can a look at the results of an ARIA run with the same data, but without the initial model [HERE](#) .

You can compare the **Quality Scores** and the **Ensemble RMSD** graph between the calculations with and without the initial model.



In the run performed without an initial model, how would you describe the ensemble in Iteration 0 ?

How does it compare with the Iteration 0 of the run from this tutorial (*i.e.* with an initial model) ?

We invite you to read the following book chapter to learn more about ARIA and on how to judge the quality and reliability of structures determined with ARIA from NMR data.

Bardiaux B, Malliavin T, Nilges M. *ARIA for solution and solid-state NMR*. **Methods Mol Biol**. 2012;831:453-83. https://doi.org/10.1007/978-1-61779-480-3_23

[Back to top](#) | [Home](#) | [PDF version](#)

