# 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 <sup>3</sup>J 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. <u>https://ariaweb.pasteur.fr/login</u>

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

Please consider re	egistering to be info	rmed about your runnir	ng jobs and keep track of al	ll your projects ("Create a f	full
account" in the us	ser menu at the top r	ight).			×
			$\sim$		
			Ϋ́,		
		Welcome to th	e ARIAweb service		
ARIA (Ambi	iguous Restraints for	Iterative Assignment) is a	a software for automated NO	E assignment and NMR structure	cture
ARIA (Ambi calculation. It sp Additionally, a re	iguous Restraints for beeds up and automa finement in explicit w	terative Assignment) is a tizes the assignment pro ater improves the quality	a software for automated NO ocess through the use of an it y of the calculated structures	E assignment and NMR stru- terative structure calculation s, validation tests help specti	cture 1 scheme. roscopists
ARIA (Ambi calculation. It sp Additionally, a re to judge the qua	iguous Restraints for beeds up and automa finement in explicit w ality of the final struct	terative Assignment) is a tizes the assignment pro ater improves the quality ures, and the support of other NMR sc	a software for automated NO poess through the use of an i y of the calculated structures i the CCPN data model simpl oftware packages	E assignment and NMR stru terative structure calculation s, validation tests help spect ifies the exchange of informa	cture a scheme. roscopists ation with
ARIA (Ambi calculation. It sp Additionally, a re to judge the qua	iguous Restraints for seeds up and automa ifinement in explicit w ality of the final struct	terative Assignment) is a tizes the assignment pro ater improves the quality ures, and the support of other NMR so	a software for automated NO ocess through the use of an in y of the calculated structures f the CCPN data model simpl oftware packages.	E assignment and NMR stru- terative structure calculation s, validation tests help spectr ifies the exchange of informa	cture a scheme. roscopists ation with
ARIA (Ambi calculation. It sp Additionally, a re to judge the qua	iguous Restraints for beeds up and automa finement in explicit w ality of the final struct	terative Assignment) is a tizes the assignment pro ater improves the quality ures, and the support of other NMR so View my Projects	a software for automated NO cess through the use of an ir y of the calculated structures f the CCPN data model simpl oftware packages. Create a new Project	E assignment and NMR stru terative structure calculation s, validation tests help spectr ifies the exchange of informa	cture a scheme. roscopists ation with
ARIA (Ambi calculation. It sp Additionally, a re to judge the qua	iguous Restraints for beeds up and automa finement in explicit w ality of the final struct	terative Assignment) is a tizes the assignment pro ater improves the quality ures, and the support of other NMR so View my Projects	a software for automated NO occess through the use of an ir y of the calculated structures f the CCPN data model simpl oftware packages.	E assignment and NMR stru terative structure calculation s, validation tests help spect ifies the exchange of inform	cture a scheme. roscopists ation with
ARIA (Ambi calculation. It sp Additionally, a re to judge the qua	iguous Restraints for beeds up and automa finement in explicit w ality of the final struct	terative Assignment) is a tizes the assignment pro ater improves the quality ures, and the support of other NMR so View my Projects Running ARIA 2.3.2 and AF Cluster la	a software for automated NO ocess through the use of an in y of the calculated structures i the CCPN data model simpl oftware packages. Create a new Project	E assignment and NMR stru terative structure calculation , validation tests help spect ifies the exchange of inform	cture a scheme. roscopists ation with
ARIA (Ambi calculation. It sp Additionally, a re to judge the qua	iguous Restraints for beeds up and automa finement in explicit w ality of the final struct	terative Assignment) is a tizes the assignment pro ater improves the quality ures, and the support of other NMR so View my Projects Running ARIA 2.3.2 and AF Cluster lo Jobs queue: 0	a software for automated NO occess through the use of an ir y of the calculated structures t the CCPN data model simpl oftware packages. Create a new Project RAweb e5329c4b (2022-06-30) ocd: 61.6 % Running   0 Pending	E assignment and NMR stru terative structure calculation s, validation tests help spect ifies the exchange of inform	cture a scheme. roscopists ation with
ARIA (Ambi calculation. It sp Additionally, a re to judge the qua	iguous Restraints for beeds up and automa finement in explicit w ality of the final struct	terative Assignment) is a tizes the assignment pro ater improves the quality ures, and the support of other NMR so View my Projects Running ARIA 2.3.2 and AF Cluster ic Jobs queue: 0	a software for automated NO bocess through the use of an in y of the calculated structures if the CCPN data model simple oftware packages. Create a new Project RAweb e5329c4b (2022-06-30) bad: 61.6 % • Running   0 Pending	E assignment and NMR stru terative structure calculation s, validation tests help spectr ifies the exchange of informa	cture a scheme. roscopists ation with
ARIA (Ambi calculation. It sp Additionally, a re to judge the qua	iguous Restraints for beeds up and automa finement in explicit w ality of the final struct	terative Assignment) is a tizes the assignment pro ater improves the quality ures, and the support of other NMR so View my Projects Running ARIA 2.3.2 and AF Cluster & Jobs queue: 0	a software for automated NO boccess through the use of an if y of the calculated structures if the CCPN data model simpl offware packages. Create a new Project RAweb e5329c4b (2022-06-30) bad: 61.6 % Running   0 Pending entation   Contact Us	E assignment and NMR stru terative structure calculation s, validation tests help spectr ifies the exchange of informa	cture 1 scheme. roscopists ation with

## 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
- <u>3. Convert the data</u>
- <u>4. Prepare your ARIA structure calculation project</u>
- <u>5. Analysing results</u>
- 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 <u>here</u>

To predict a 3D model of our protein of interest, we will use this ColabFold notebook.

In the query\_sequence filed, input the one-letter sequence of the construct used for NMR experiment, *i.e.* including the expression tag:

MKHHHHHHPMELNNLRMTYERLRELSLNLGNRMVPPVGNFMPDSILKKMAAILPMNDSAFATLGTVEDKYRRFKYFK 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.

• •	• <		0	a colab.research.google.com	Ċ	± = +
co	R A File	lphaFold2.ipynb Edit View Insert Runtime	Tools Help <u>Cannot save c</u>	hanges		🖘 Share 🎄 👼
≔	+ Code	e + Text & Copy to Drive				Connect T4
ৎ {x} ে	✓ C Easy align notel <u>Mirdi</u> acce	ColabFold v1.5.5: Alp to use protein structure and ments/templates are genera sook, checkout the <u>ColabFolt</u> ta M. Schütze K. Moriwaki Y. ssible to all. <i>Nature Methods</i> .	bhaFold2 using MI complex prediction using ted through <u>MMseqs2</u> an <u>1 GitHub</u> and read our ma <u>Heo L. Ovchinnikov S. Ste</u> .2022	Mseqs2 <u>AlphaFold2</u> and <u>Alphafold2-multimer</u> . Sequence d <u>HIsearch</u> . For more details, see <u>bottom</u> of the nuscript. Old versions: <u>v1.4</u> , <u>v1.5.1</u> , <u>v1.5.2</u> , <u>v1.5.3</u> -p <u>inegger M. ColabFold: Making protein folding</u>	atch	
	<b>N</b> 1		a) than hit Dunting	> Dun all		↑↓⇔‡ [] 🔋 :
	· "	iput protein sequence(	s), then nit Runtime	-> Run att		
	0	query_sequence: "M	KHHHHHPMELNNLRMT	YERLRELSLNLGNRMVPPVGNFMPDSILKKMAAILF	PMNDSAFATLGTVEDKYRRR	RFKYFKATIADLSKKRSSE
		<ul> <li>Use : to specify inter</li> </ul>	r-protein chainbreaks for I	modeling complexes (supports homo- and hetro-o	ligomers). For example <b>PI</b>	.SK:PISK for a homodimer
		jobname: "hrdc				
		num_relax: 1				•
		<ul> <li>specify how many of</li> </ul>	the top ranked structures	to relax using amber		
		template_mode: none				<b>.</b>
<>		<ul> <li>none = no template ir notes)</li> </ul>	nformation is used. pdb10	00 = detect templates in pdb100 (see <u>notes</u> ). cust	om - upload and search ow	n templates (PDB or mmCIF format, see
=		Show code				
D	> lı	nstall dependencies				
						• ×



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:

nrdc_5687f/
<pre>cite.bibtex</pre>
config.json
hrdc_5687f.a3m
hrdc_5687f.csv
hrdc_5687f.done.txt
<pre> hrdc_5687f_coverage.png</pre>
hrdc_5687f_env
hrdc_5687f_pae.png
hrdc_5687f_plddt.png
<pre> hrdc_5687f_predicted_aligned_error_v1.json</pre>
<pre> hrdc_5687f_relaxed_rank_001_alphafold2_ptm_model_3_seed_000.pdb</pre>
<pre> hrdc_5687f_scores_rank_001_alphafold2_ptm_model_3_seed_000.json</pre>
hrdc_5687f_scores_rank_002_alphafold2_ptm_model_5_seed_000.json
<pre> hrdc_5687f_scores_rank_003_alphafold2_ptm_model_4_seed_000.json</pre>
<pre> hrdc_5687f_scores_rank_004_alphafold2_ptm_model_2_seed_000.json</pre>
hrdc_5687f_scores_rank_005_alphafold2_ptm_model_1_seed_000.json
<pre> hrdc_5687f_unrelaxed_rank_001_alphafold2_ptm_model_3_seed_000.pdb</pre>
<pre> hrdc_5687f_unrelaxed_rank_002_alphafold2_ptm_model_5_seed_000.pdb</pre>
<pre> hrdc_5687f_unrelaxed_rank_003_alphafold2_ptm_model_4_seed_000.pdb</pre>
<pre> hrdc_5687f_unrelaxed_rank_004_alphafold2_ptm_model_2_seed_000.pdb</pre>
<pre> hrdc_5687f_unrelaxed_rank_005_alphafold2_ptm_model_1_seed_000.pdb</pre>
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 pIDDT
- \*. j son: 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 pIDDT

# 2. Create an ARIAweb Project

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



#### 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.

	iii ariaweb.pasteur.fr		
Projects			- <b>5</b> = = 🦪 (
races from Broject h	rdo		
leasy non Project n			
			Đ
PROTEIN			Ŵ
			Đ
ame 13C NOESY   Segids A   Typ	e 3D noesy (carbon)		Û
ame 15N NOESY   Segids A   Typ	e 3D noesy (nitrogen)		Ŵ
	Finish		
Institut Pasteur USR 3756	Home   Documentation   Contact Us 2020-2021 © Institut Pasteur. All rights reserved.	<b><sup>2</sup> i</b> B	
	Projects  Reasy from Project h  ROTEIN  Ime 13C NOESY   Segids A   Typ  ame 15N NOESY   Segids A   Typ  Institut Poster State State  Institut Poster State State	Projects  Reasy from Project hrdc  ROTEIN  Imme 13C NOESY   Segids A   Type 3D noesy (carbon)  ame 15N NOESY   Segids A   Type 3D noesy (nitrogen)  Finish  Finish  Mome   Documentation   Contact Us 2020-2021 @ Institut Pasteur. All rights reserved.	Projects  Reasy from Project hrdc  ROTEIN  ROTEIN  Imme 13C NOESY   Segids A   Type 3D noesy (carbon)  ame 15N NOESY   Segids A   Type 3D noesy (nitrogen)  Finish  Finish  Lower   Documentation   Contact Us 2020-2021 @ Institut Pasteur. All rights reserved.

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.
  - $\,\circ\,$   $^{13}\text{C}\text{-resolved}$  NOESY
  - $\circ~^{15}\mathrm{N}\text{-resolved}$  NOESY

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

Home Documentation Projects				
-				
Detail Chain for Conver	t project xeasy			
Format: seq		Type: PROTEIN		
Naming Convention:		First residue number	r: 1	
Input: hrdc.seq		Segids: A		
Back				
	0	ariaweb.pasteur.fr	0	
Home Documentation Projects				🛃 🗄 🛃 🛷
Detail Spectrum for Co	nvert project xeas	у		
Detail Spectrum for Con Spectrum name: 13C NOESY	nvert project xeas	y Spectrum type: noes	y_hsqc_HCH.hhc	
Detail Spectrum for Con Spectrum name: 13C NOESY Spectrum ambiguity: all	nvert project xeas	<b>y</b> Spectrum type: noes Segids: A	y_hsqc_HCH.hhc	
Detail Spectrum for Con Spectrum name: 13C NOESY Spectrum ambiguity: all Chemicalshift format: xeasy	nvert project xeas	Y Spectrum type: noes Segids: A Chemicalshift input:	y_hsqc_HCH.hhc hmqcnoe.prot	
Detail Spectrum for Con Spectrum name: 13C NOESY Spectrum ambiguity: all Chemicalshift format: xeasy Crosspeaks format: xeasy	nvert project xeas	Y Spectrum type: noes Segids: A Chemicalshift input: Crosspeaks input: hr	y_hsqc_HCH.hhc hmqcnoe.prot mqcnoe_partially_assigned.peaks	
Detail Spectrum for Cod Spectrum name: 13C NOESY Spectrum ambiguity: all Chemicalshift format: xeasy Crosspeaks format: xeasy Proton 1: 2	nvert project xeas	y Spectrum type: noes Segids: A Chemicalshift input: Crosspeaks input: hu Proton 2: 3	y_hsqc_HCH.hhc hmqcnoe.prot mqcnoe_partially_assigned.peaks	
Detail Spectrum for Con spectrum name: 13C NOESY Spectrum ambiguity: all Chemicalshift format: xeasy Crosspeaks format: xeasy Proton 1: 2 Hetero 1: 1	nvert project xeas	y Spectrum type: noes Segids: A Chemicalshift input: Crosspeaks input: hr Proton 2: 3 Hetero 2: None	y_hsqc_HCH.hhc hmqcnoe.prot nqcnoe_partially_assigned.peaks	
Detail Spectrum for Con Spectrum name: 13C NOESY Spectrum ambiguity: all Chemicalshift format: xeasy Crosspeaks format: xeasy Proton 1: 2 Hetero 1: 1 Check format of input files?: True	nvert project xeas	y Spectrum type: noes Segids: A Chemicalshift input: Crosspeaks input: hr Proton 2: 3 Hetero 2: None	y_hsqc_HCH.hhc hmqcnoe.prot nqcnoe_partially_assigned.peaks	
Detail Spectrum for Cod Spectrum name: 13C NOESY Spectrum ambiguity: all Chemicalshift format: xeasy Crosspeaks format: xeasy Proton 1: 2 Hetero 1: 1 Check format of input files?: True	nvert project xeas	y Spectrum type: noes Segids: A Chemicalshift input: Crosspeaks input: hr Proton 2: 3 Hetero 2: None	y_hsqc_HCH.hhc hmqcnoe.prot nqcnoe_partially_assigned.peaks	
Detail Spectrum for Cod Spectrum name: 13C NOESY Spectrum ambiguity: all Chemicalshift format: xeasy Crosspeaks format: xeasy Proton 1: 2 Hetero 1: 1 Check format of input files?: True Back	nvert project xeas	y Spectrum type: noes Segids: A Chemicalshift input: Crosspeaks input: hr Proton 2: 3 Hetero 2: None	y_hsqc_HCH.hhc hmqcnoe.prot nqcnoe_partially.assigned.peaks	
Detail Spectrum for Cod Spectrum name: 13C NOESY Spectrum ambiguity: all Chemicalshift format: xeasy Crosspeaks format: xeasy Proton 1: 2 Hetero 1: 1 Check format of input files?: True Back	nvert project xeas	y Spectrum type: noes Segids: A Chemicalshift input: Crosspeaks input: hr Proton 2: 3 Hetero 2: None	y_hsqc_HCH.hhc hmqcnoe.prot mqcnoe_partially.assigned.peaks	

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.

## Submit conversion

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

Data Conv	version Projects : hrdc		±.
Show 10 + en	ntries		Search:
Name	↑↓ Creation Date	∿ Molecule name	∿
xeasy	2022-07-15, 12:26 p.m.	hrdc	Start conversion
Showing 1 to 1 of	1 entries		Previous 1 Next
			See results for project hrdc
		Home   Documentation   Contact Us	2 100
	(Y) @	Home   Documentation   Contact Us	See results for project hrdc

Figure 7: List of conversions ready to be submitted

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

Your Job Project type conv	vert, name: xeasy has been suc	cessfully submitted. Results will be shown here.			×
Results project: h	nrdc				
Show 10 ¢ entries			Search	:	
Name		↑↓ Creation date	∿↓ State	î.↓	
Project type convert, na	ame: xeasy	2022-07-15, 2:59 p.m.	۵	Û	
Showing 1 to 1 of 1 entrie	s			Previous 1 Next	
				View Demo result	s
	Institut Pasteur USR 3756	Home   Documentation   Contact Us	<b>2</b>		

Figure 8: Results page where conversion is running

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

	0	ariaweb.pasteur.fr	Ċ			
Home Documentation Projects				Hrdc ·	8 = =	∅ @
Results project: hrdc						
Show 10 + entries			S	Search:		
Name		∿ Creation date	1∿ State	ŤΨ		
Project type convert, name: xeasy		2022-07-15, 2:59 p.m.	<ul> <li>✓ 0</li> </ul>	± t	Ŵ	
Showing 1 to 1 of 1 entries				Previo	ous 1 Next	
					view Demo resu	ults
	Institut Pasteur USR 3756	Home   Documentation   Contact Us 2020-2021 © Institut Pasteur. All rights reserved.	<b>B</b>			

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 **1** 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

Home Docu	mentation Projects		i ariaweb.pasteur.tr	C	c - <u>1) ≓</u> ≡ ∅ (
		Stru	icture Calculati	ion	
	100		:=	æ	
	SETUP	DATA	PROTOCOL	STRUCTURE GE	NERATION
Setup	(	Name	0	Version <sup>*</sup> (?)	
				1.0	٢
		Descri	ption 🔞	Comment (?)	
				li li	
		Refere	nces 🕐		
		Enable t	he <b>Expert Mode</b> to see more options.		
		Next			
		Rese	t Project		

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 form the previous step. Click Next to move to the Spectra definitions.

## NOE peak-list and chemical shifts

Again, your <sup>13</sup>C-NOESY and <sup>15</sup>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.

Home Documentation Projects		Ľ			Tutorial 🦓 🗮 📰 🤇
		Structure Calo	ulation		
SETUP		DATA	PROTOCOL	STRUCTURE GEN	IERATION
Molecular system	-				* * ~
Spectrum		Spectrum 1			
Symmetry	)	0			
Initial Structure Ensemble	-	Generic Use assignments ⑦		🗌 Filter diagonal peaks 🕐	
Dihedral angles		Chemical Shift List			
Hydrogen bonds		File 😗		Format 💿	
RDCs		Choose file	Browse	XML	*
Scalar Couplings		Use shift error ③	ioe_37.xmi		
Ambiguous distances		Peaks List			
Unambiguous distances		File <sup>®</sup> 😗		Format <sup>*</sup> 😗	
Disulfide bridges restraints		Choose file	Browse	XML	+
Disulfide bridges covalent	0	crosspeaks_66_hmqcnoe_partially_	assigned_37.xml		
UIC antabas		Proton1 shift err* 💿			
nis patches		0.04	٢		
Cis-Proline patches		Hetero1 shift err 🕐	A	Proton2 shift err 0	
Zinc coordination		Hatara2 shift arr	۲	Volume or intensity	
Other		0.5	٢	Volume	+
		Spectrum 2			Û
		Generic Use assignments ® Chemical Shift List		🗌 Filter diagonal peaks 🕑	
		File <sup>®</sup> (9)		Format 🔊	
		Choose file	Browse	XML	+

Figure 11: Spectra parameters

# Adding the input model

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.

	0	🗎 ariaweb.pasteur.fr	Ċ	0 1
Home Documentation Projects				Tutorial 📲 🧮 🛒
		Structure Calculation		
<b>10</b> 2		:=		æ
SETUP	DATA	PROTOCOL		STRUCTURE GENERATION
Molecular system	-			* * *
Spectrum	-	Initial Structure Ensemble 1		
Symmetry		File* 🔿	Format* (2)	
Initial Structure Ensemble		hrdc_5687f_relaxed_rank_001_alphafold2_ptm_model_3_seed_ Brows	IUPAC	•
Dihedral angles		Example: template_jupac.pdb		
Hydrogen bonds	Ena	able the Expert Mode to see more options.		
RDCs		Tint Deals Mont		Add Initial Oterration Franklin
Scalar Couplings		ITSL DOCK INEXT		Add Initial Structure Ensemble
Ambiguous distances	<u> </u>	Reset Project		
Unambiguous distances				
Disulfide bridges restraints				
	~			

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 (<sup>(2)</sup>) icon, top right).

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

•	$\langle \rangle$	0		ariaweb.pasteur.fr		C	0 1
9	Home Documentation P	rojects				Tutorial	5 ≓ ≣ ∅
				<b>Г</b> /9		В	ardiaux
			-				🔓 Profile
			S	tructure Calculat	ion		Change Password
							🕒 Logout
	<b>\$</b> \$					ß	<ul> <li>Expert Mode</li> </ul>
	SETUP		DATA	PROTOCOL		STRUCTURE GENE	RATION
	Water refinement	-					× * ~
	Iteration settings			Iteration 0			
	Protocol						
				Generic			
				Number of structures* (2)		Use n best structures* ⑦	
				20	٢	7	٢
				Sort criterion <sup>*</sup> ⑦		Keep n previous structures*	0
				Total Energy	٥	0	٢
				Merger settings			
				Method*			
				Standard	٥		
				Calibration			
				Distance cutoff [A]*		Estimator	
				6.0	٢	Ratio of averages	\$
				Spin Diffusion Correction		Error Estimator*	
						Distance	\$
				Violation Analysis			
				Violation tolerance [A]* ③		Violation threshold <sup>*</sup> (?)	
				1000.0	٢	0.5	٢

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

	0	ariaweb.pasteur.fr	Ċ	• • • • +
Home Documentation Projects		-		Tutorial 📲 🗮 🛷 🔘
		Ľ		
		Structure Calculation		
		-	8	
SETIP		PROTOCOL	STRUCTURE GENI	RATION
52.51	2010	- A COULT	STRUGTURE DEM	
Water refinement	-			* * ~
Iteration settings	= It	eration 0		
Protocol				
	Nu	eneric imber of structures* ()	Use n best structures" 💿	
		80 8	7	٥
	Sa	rt criterion" ()	Keep n previous structures* 💿	
		Total Energy +	0	٢
	M	erger settings		
	M	sthod"		
	1	Standard +		
	с	alibration		
	Di	stance cutoff [A]	Estimator	
		.0 ()	Ratio of averages	•
		Spin Diffusion Correction	Error Estimator	
			Distance	•
	V	iolation Analysis		
	Vi	plation tolerance [A]* ()	Violation threshold"	
			0.5	
	Si	ima mode		
		₩ ¥		
	P	artial Assignment	Max number of contributions*	
		1999 (a)	20	8
	- M	etwork Apchoring (NA)		

Figure 14: Adapted parameters in Iteration Settings

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

Iteration	<b>Violation tolerance</b>	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.

Home Documentation	Projects	≡ anaveu,pastou.n	Ŭ	Hrdc 🎝 🗮 📰 🦪	
		[CP			
		Structure Calcul	ation		
\$	3	∎ ≔		&	
SETUP	DATA	PROTO	COL ST	RUCTURE GENERATION	
Simulated Annealing	-	MD type <sup>*</sup>	High-temp steps	s* ⑦	
Ambiguous restrainte		Torsion angle	÷ 10000	٢	
Anbiguous restraints		Cool1 steps 🖁 🔞	Cool2 steps* 📀		
Unambiguous restraint	s	5000	3 4000	٢	
Symmetry restraints		Enable the Expert Mode to see more options.			
Flat Bottom Harmonic	Wall				
Log-Harmonic Potentia	al 🔎	First Back Save			
Hydrogen bonds restra	aints	Reset Project			
Dihedral angle restrain	ts 🕥				

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.

me Documentation Proie				
	ects		Hrdc 🎝 🗮 📰 🦂	a (
ructure Calculat	ion Projects : hrdc		+	
w 10 ¢ entries			Search:	
ame	↑↓ Creation Date	1↓		
dc	2022-07-15, 4:50 p.m.		Start ARIA job 🔮 🛍	
wing 1 to 1 of 1 entries			Previous 1 Next	
			See results for project hrd	с
Institut Pa	Steur USR 3756	Home   Documentation   Contact Us 2020-2021 © Institut Pasteur. All rights reserved.	Ž 🏤	
) )	ructure Calculat         w 10 + entries         ame         dc         wing 1 to 1 of 1 entries         Institut Po	ructure Calculation Projects : hrdc w 10 € entries ame 1 Creation Date dc 2022-07-15, 4:50 p.m. wing 1 to 1 of 1 entries Winstitut Pasteur Sease	ructure Calculation Projects : hrdc w 10 € entries ame N Creation Date V dc 2022-07-15, 4:50 p.m. wing 1 to 1 of 1 entries Wing 1 to 1 of 1 entries Home   Documentation   Contact Us 2020-021 @ Institut Pasteur. All rights reserved.	Image: Calculation Projects : hrdc       Search:         w 10 • entries       Search:         ame       1 • creation Date       1 •         dc       2022-07-15, 4:50 p.m.       Start ARIA job @ 0         wing 1 to 1 of 1 entries       Previous 1 • Next         Image: Calculation Project hrd         Image: Calculation Date         dc       2022-07-15, 4:50 p.m.         wing 1 to 1 of 1 entries       Previous 1 • Next         Image: Calculation Project hrd         Mome   Documentation   Contact Us         Search:

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  $\checkmark$ ), click on the  $\clubsuit$  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 <u>HERE</u> 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).



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  $\bigoplus$  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 D button to select individual models for display).

The  $\textcircled{\bullet}$  button allows to upload another PDB file to superimpose on the ARIA structure ensemble (use the  $\diamondsuit$  button to superimpose on a selected structure).

Below, Restraints statistics recapitulate the number of restraints and the trend since the previous iteration for:

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

## ?

How many NOE cross-peaks remain unassigned ?

How many restraints were used for structure calculation ?

The Contribution histogram gives count of restraints (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 restraints. 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 Molprobity. 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.

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. Molprobity clashscore. this reports the number of overlaps >0.4Å 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

© 2024 Benjamin Bardiaux (Institut Pasteur/CNRS)