SHREDDER Tutorial for Arcimboldo

Aims of the tutorial

This tutorial shows how to create a library of models with ARCIMBOLDO-SHREDDER. SHREDDER evaluates a model template against the rotation function, generates ad-hoc models and uses them within ARCIMBOLDO to solve the structure.

Data tutorial: tutorial files

Experimental details

Test data for our tutorial have been obtained from the protein MltE (PDB ID: 2Y8P). It is a bacterial outer membrane-anchored endolytic peptidoglycan lytic transglycosylase. This structure was solved with data extending to 2.0 Å.

Data are summarized in the following table:

PDBID	2Y8P
Space group	C2221
Unit cell (a, b, c) (Å)	123.324 183.932 35.292
Resolution Range (Å)	102.5 - 2.0 Å
Residues in asymmetric unit	2 x 194

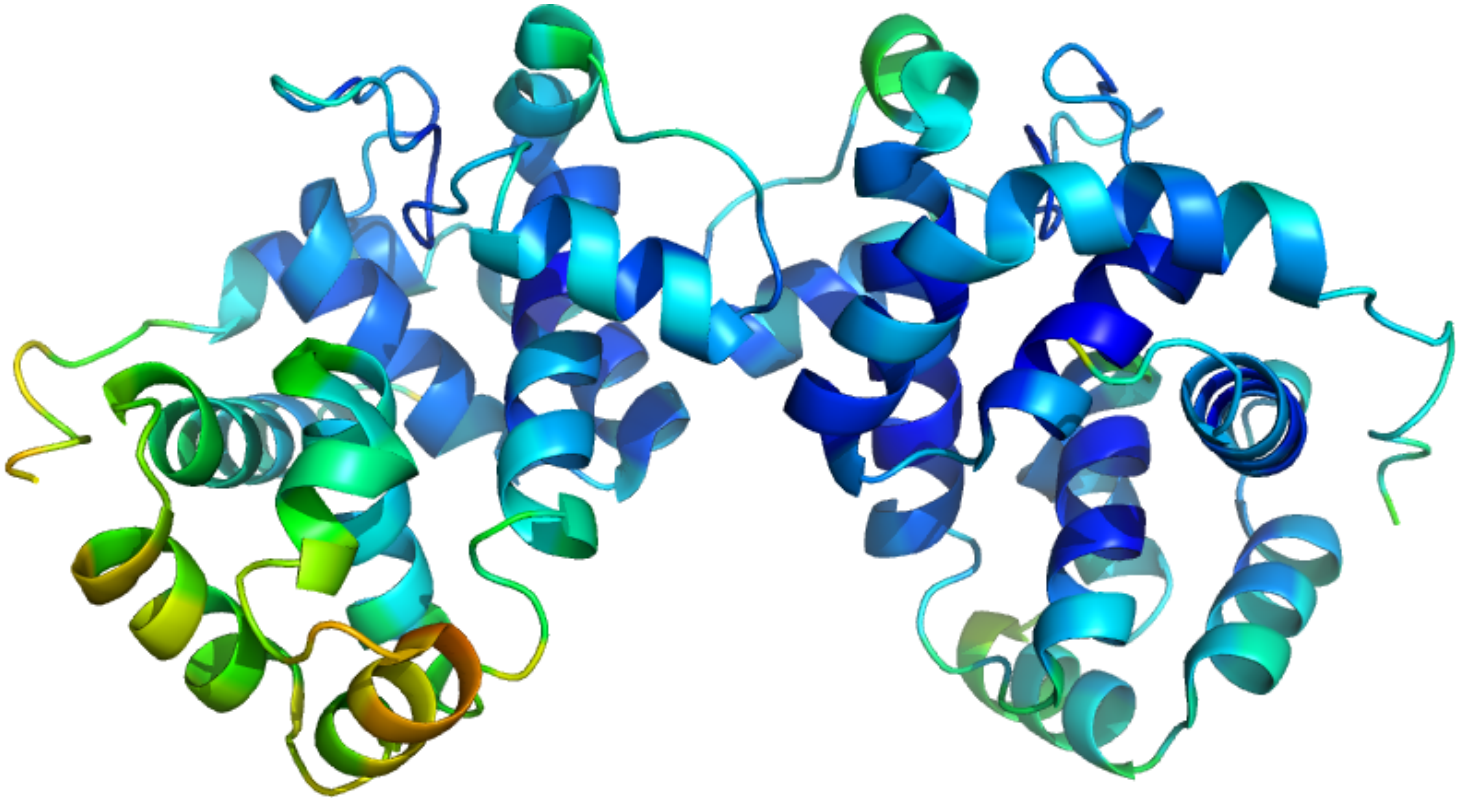


Figure 1. MltE Structure colored after B-value.

Step by Step tutorial

1. Creation of the library from a selected homology model

BLAST search of the MltE sequence against PDB can be used to find a model, 1QTE, with similarity of predicted fold and 35% of identity over a range of 66 amino acids.

```
mlte          KHDYTNPPWNAKVPVQRA-MQWMPISQKAGAAWG----------VDPQLITAIIAIESG 48
1QTE|A        RYAFNNQWWDLSVQATIAGKLWDHLEERFPLAYNDLFKRYTSGKEIPQSYAMAIARQESA 480
              :: :.*  *: .* .  *   *  :.::   *:.          :  .   **   **.

mlte          GNPNAVSKSNAIGLMQLKASTSGRDVYRRMGWSGEPTTSELKNPERNISMGAAYLNILET 108
1QTE|A        WNPKVKSPVGASGLMQIMPGT-ATHTVKMFSIPGYSSPGQLLDPETNINIGTSYLQYVYQ 539
               **:. *  .* ****: ..* . .. : :. .* .:..:* :** **.:*::**: :

mlte          GPLAGIEDPKVLQYALVVSYANGAGALLRTFSSDRKKAISKINDLDADEFLEHVARNHPA 168
1QTE|A        ----QFGNNRIFSSAAYNAGPGRVRTWLG-NSAGRIDAVAFVESIPFSETRGYVKNVLAY 594
                   : : :::. *   : .. . : *   *:.* .*:: ::.:  .*   :* .  .

mlte          PQAPRYIYKLEQALDAMLEHHH-- 190
1QTE|A        DAYYRYFMGDKPTLMSATEWGRRY 618
                  **:   : :* :  *  :
```

Figure 2. MltE alignment against 1QTE

A classical molecular replacement approach with the 1QTE model trimmed in standard ways fails due to several local structural differences.

Input

We will need:

- An .mtz file containing the reflection data
- The configuration .bor file
- The model .pdb file

The description of the configuration file follows:

```
[CONNECTION]:
use_remote_grid: False
remote_frontend_passkey:False
[LOCAL]
# Third party software paths
path_local_phaser: /path/to/local_phaser
path_local_shelxe: /path/to/local_shelxe
[GENERAL]:
working_directory: /path/to/working_directory

mtz_path: %(working_directory)s/cecilia.mtz
hkl_path: %(working_directory)s/cecilia.hkl

[BORGES-SHREDDER]
name_job: 1qte
molecular_weight: 20000
f_label: F
sigf_label: SIGF
number_of_component: 2
identity_shredder: 1.2
identity_arcimboldo: 0.6
model_file: %(working_directory)s/1qte.pdb

shelxe_line_fast: -m20 -v0 -a5 -t10 -q -s0.5 -u2000
pack_clashes: 6
fragment_to_search: 2

[MYSQL]
borges_userdb_hostname: localhost
borges_userdb_port: 3306
```

In the .bor file you need to specify:

- the [CONNECTION], [LOCAL] and [GENERAL], sections, which are common to all BORGES-ARCIMBOLDO programs

- a [BORGES-SHREDDER] section

The mySql parameters depend on your particular setup. use_remote_grid and remote_frontend_passkey have to be set up for connection to a remote grid. If true, the passkey should be given, otherwise it will ask for the password. In the general section you need to give the path to your input files.

In the [BORGES-SHREDDER] you will need to define the contents of the asymmetric unit, the model pdb to use, the number of copies to search and the labels for the mtz file. In this case, we also have identity values (as rmsd) to define to PHASER: one to be used in the Shred-LLG evaluation and optimization and one for the ARCIMBOLDO runs launched with the models generated in the first step. Considering that models have been improved and reduced in size with respect to the original template, it is advisable to use a smaller value for the rms deviation for ARCIMBOLDOs than for SHREDDER. We can also set a number of clashes to be allowed in the packing analysis for the ARCIMBOLDO's runs.

The key part of this section is the one defining the shredding, which in this case will be a sequential shred from 4 to 19 residues with step size is 1. As we input the keyword omit, models will be generated by omitting shreds of 4,5,6....19 residues at every possible starting position.

Execution

You can run the program interactively, getting the output displayed on the screen and inputting the password manually. Alternatively, run in background, redirecting an input file with the passwords you need and passing the output to a log file.

1. Interactively

```
BORGES-SHREDDER.py 1qte.bor
```

2. In background

```
BORGES-SHREDDER.py 1qte.bor < password >& log &
```

Output and Results

In the directory where you launched BORGES-SHREDDER, you have a ./library/ folder containing all models and a set of folders called ARCI_*/, where * refers to the number of the rotation cluster. Inside each folder, there is another set of sub-folders with ARCIMBOLDO runs that are called as the models that have been used. You will find inside of each of them an html file summarising the ARCIMBOLDO run. For our test case, we find that the model percentile75 has been able to solve the structure at the second fragment search. You can check the html output here.

The first section echoes all parameters used for the run, so that defaults are listed along with those set in the .bor file. This allows to reproduce the run even if defaults may change in future versions.The next section displays a sortable table summarizing the results for each step.

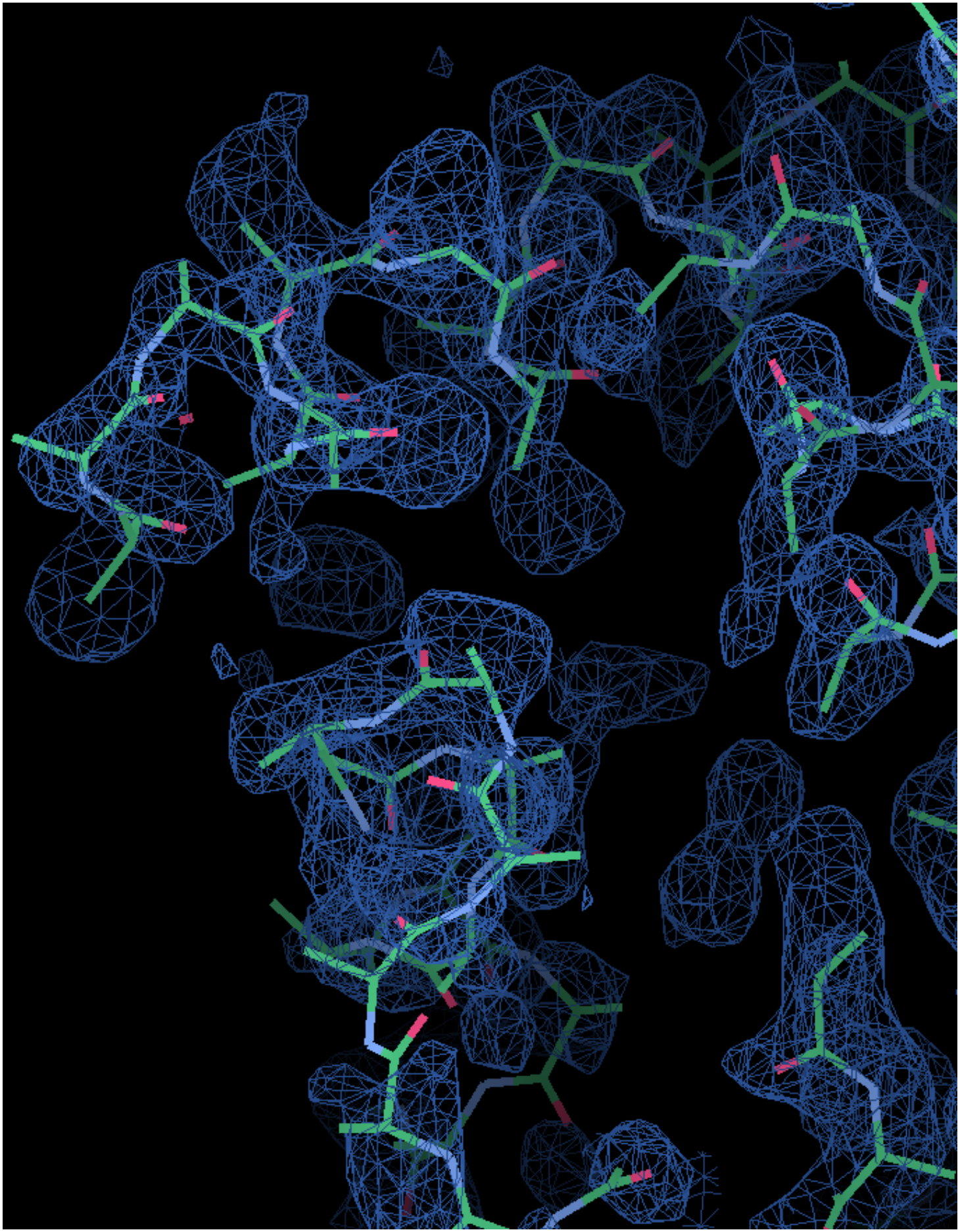
Fragment 1

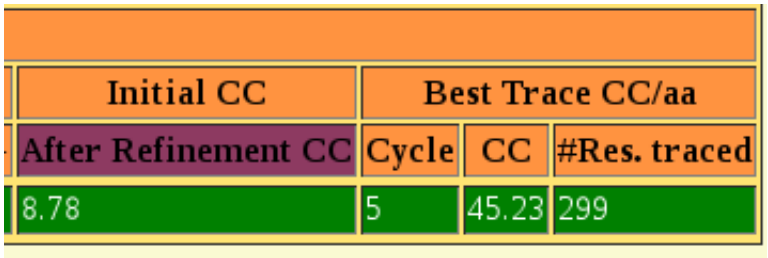
Cluster	Rotation Function					Translation Function					Packing					Rigid Body Refinement			Initial CC	Best Trace CC/aa		
	#Rots.	Top LLG	Mean LLG	Top Zscore	Mean Zscore	#Trans.	Top LLG	Mean LLG	Top Zscore	Mean Zscore	#Sol.	Top LLG	Mean LLG	Top Zscore	Mean Zscore	#Sol.	Top LLG	Mean LLG	After Refinement CC	Cycle	CC	#Res. traced
0	4	16.78	14.02	4.60	4.25	50	22.00	14.25	5.17	4.45	13	19.73	14.26	5.16	4.41	13	21.10	16.92	4.74			
1	4	15.44	12.46	4.43	4.04	49	26.49	13.21	5.67	4.64	8	26.49	15.31	5.67	4.50	8	37.50	19.14	5.42			
2	2	12.49	12.46	4.05	4.04	33	20.18	10.96	5.30	4.33	9	20.18	11.46	5.30	4.44	9	22.40	13.77	5.08			
3	1	11.78	11.78	3.96	3.96	17	15.93	10.93	5.16	4.56	6	15.93	10.80	5.12	4.56	6	18.60	14.12	4.62			
4	3	11.27	9.80	3.90	3.71	14	18.76	10.89	5.47	4.66	5	13.68	11.27	5.27	4.65	5	17.10	13.46	4.73			
5	2	10.41	9.68	3.78	3.69	11	11.05	9.74	4.81	4.56	6	10.85	9.69	4.81	4.58	5	17.60	14.04	4.88			
6	1	9.39	9.39	3.65	3.65	4	9.91	8.81	4.79	4.57	2	9.91	9.14	4.57	4.48	2	10.30	9.80	4.40			
8	1	8.14	8.14	3.49	3.49	5	19.66	12.88	5.10	4.63	2	14.73	12.12	4.95	4.54	2	19.20	15.90	4.38	5	19.02	184

Fragment 2

Cluster	Rotation Function					Translation Function					Packing					Rigid Body Refinement			Initial CC	Best Trace CC/aa		
	#Rots.	Top LLG	Mean LLG	Top Zscore	Mean Zscore	#Trans.	Top LLG	Mean LLG	Top Zscore	Mean Zscore	#Sol.	Top LLG	Mean LLG	Top Zscore	Mean Zscore	#Sol.	Top LLG	Mean LLG	After Refinement CC	Cycle	CC	#Res. traced
(0, 1)	65	34.71	25.57	4.62	3.93	142	-39.97	-58.55	6.31	4.92	1	1.28	1.28	6.48	6.48	1	89.10	89.10	8.78	5	45.23	299

Show All | Hide Not Relevant





Initial CC		Best Trace CC/aa		
After Refinement CC		Cycle	CC	#Res. traced
8.78		5	45.23	299

As you can observe in the html file, in this case the structure is solved after location of the second fragment, with a CC of 45.23 for 299 residues traced. The last section of the html contains links to access the best current solution: the pdb of the traced structure and the map in .phs format. The map is interpretable, and the model can be improved to the full trace. The html file also lists the backtracing for the best solution.

The work folder also contains all the files, that allow the program to be rerun from the break point in case of interruption.