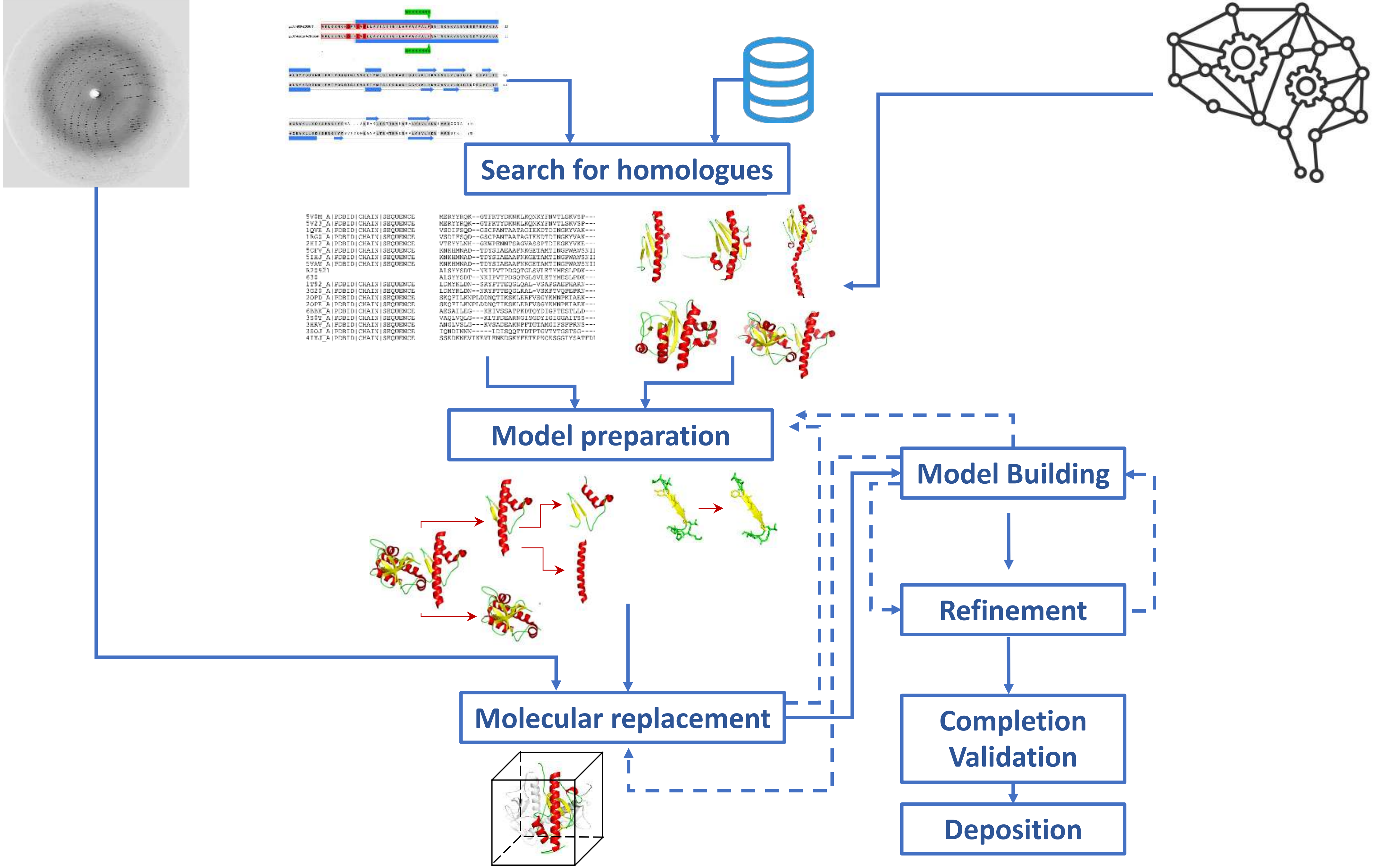
MR case studies

Paula S. Salgado
DLS/CCP4 Workshop 2023

SALGADO LAB

@paulasalgado@mastodon.social @pssalgado salgadolab.org



Search for homologues

Model preparation

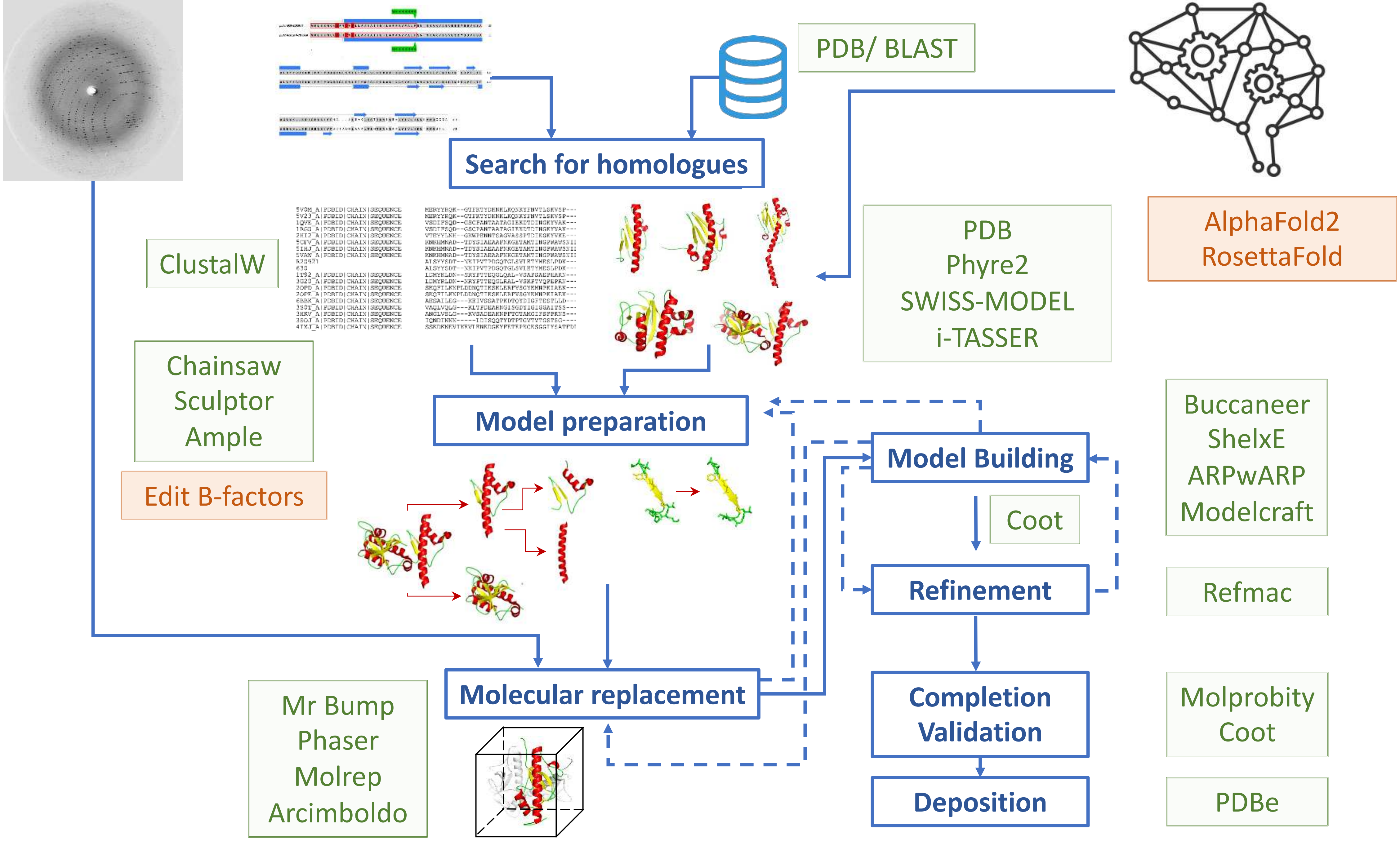
Model Building

Refinement

Molecular replacement

Completion
Validation

Deposition



PDB/ BLAST

Search for homologues

AlphaFold2
RosettaFold

ClustalW

PDB
Phyre2
SWISS-MODEL
i-TASSER

Chainsaw
Sculptor
Ample

Edit B-factors

Model preparation

Model Building

Buccaneer
ShelxE
ARPwARP
Modelcraft

Coot

Refinement

Refmac

Molecular replacement

Mr Bump
Phaser
Molrep
Arcimboldo

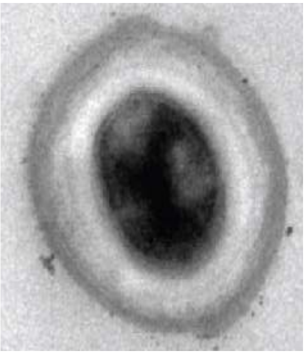
Completion
Validation

Molprobity
Coot

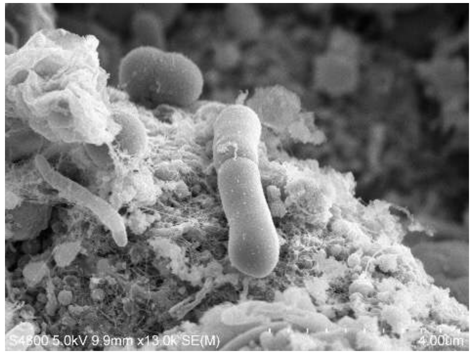
Deposition

PDBe

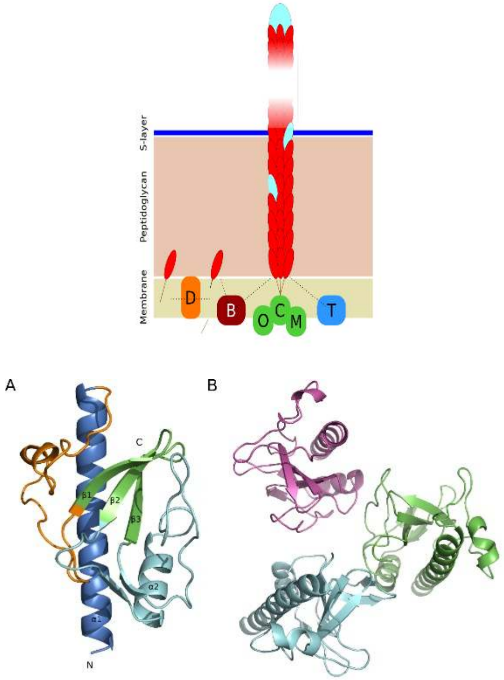
Case studies from *C. difficile* proteins



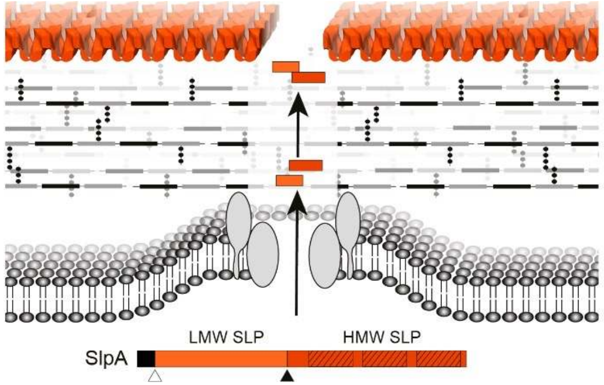
+
Antibiotics



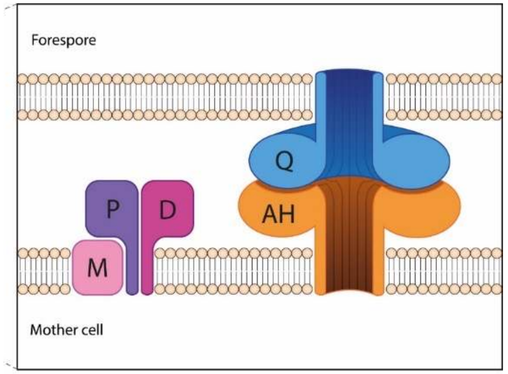
Inflammatory
Complications



Type IV pili

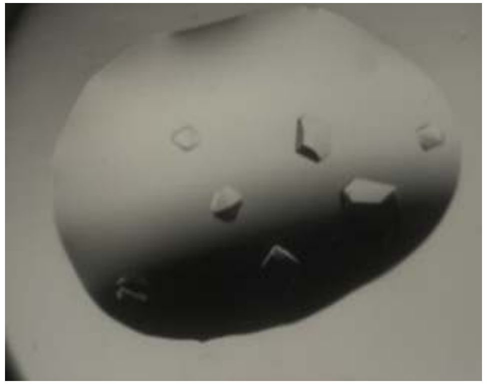


S-layer

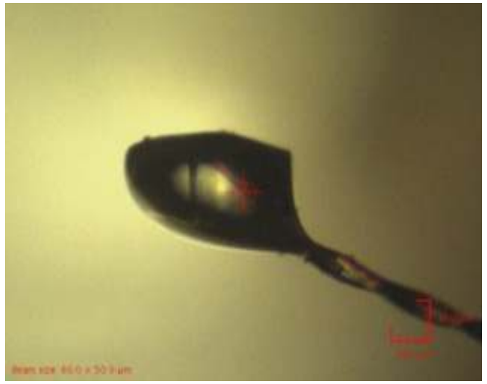


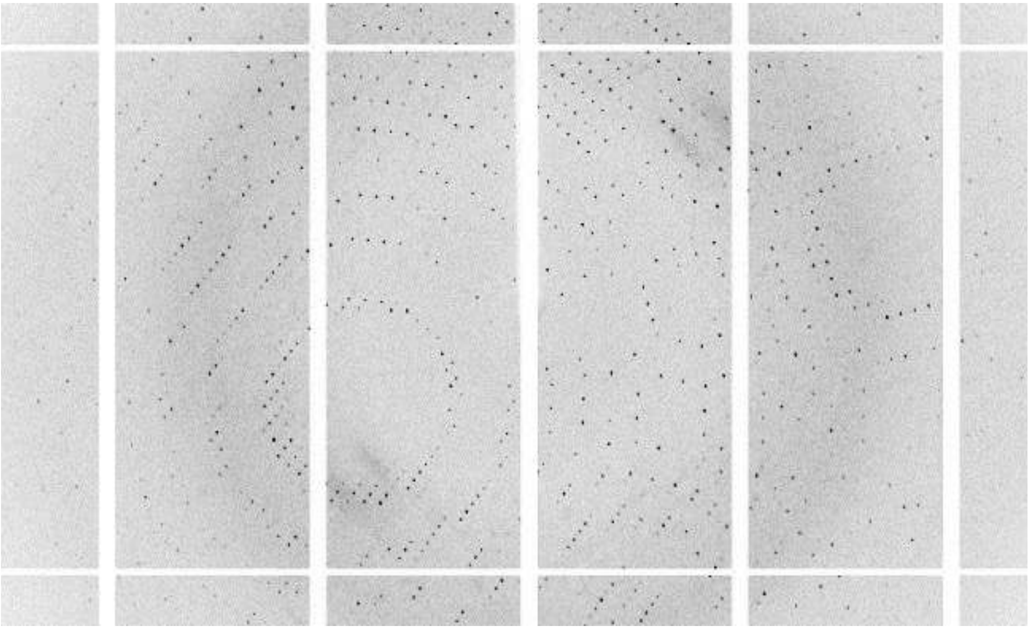
**Sporulation/
engulfment**

Example 1: CdPilA1

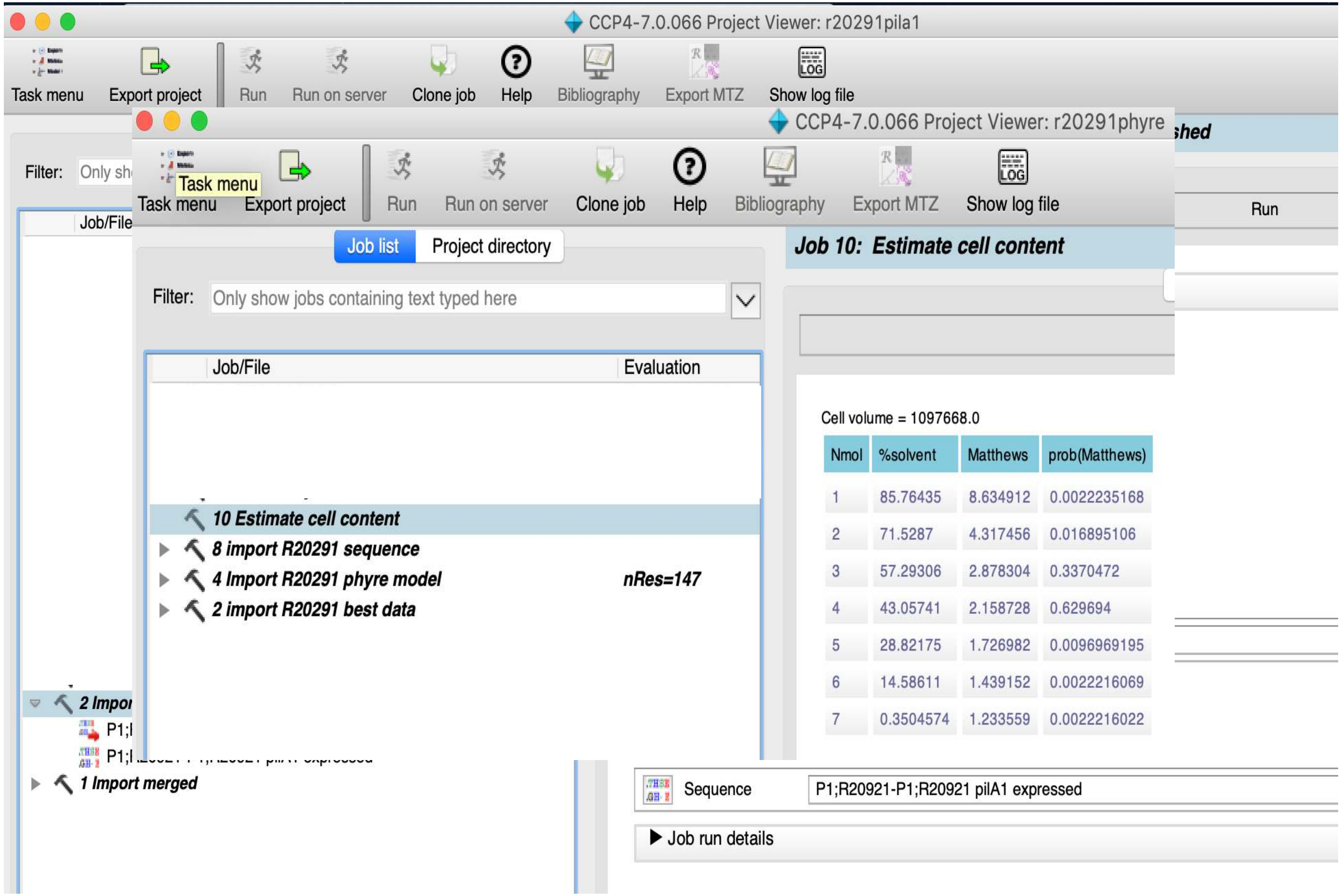


A. Crawshaw, C. Davies





	R20291 Native
Resolution (Å)	51.21 - 1.65
Wavelength (Å)	
Unit cell	
a, b, c (Å)	102.42, 102.42, 104.21
α=β=γ (°)	90
Spacegroup	$P4_12_12$
R_{merge}	0.136 (1.887)
I/sig I	10.7 (1.5)
Mean intensity CC1/2	0.998 (0.323)



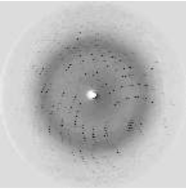
Job/File	Evaluation
10 Estimate cell content	
8 import R20291 sequence	
4 Import R20291 phyre model	nRes=147
2 import R20291 best data	

Job 10: Estimate cell content

Cell volume = 1097668.0

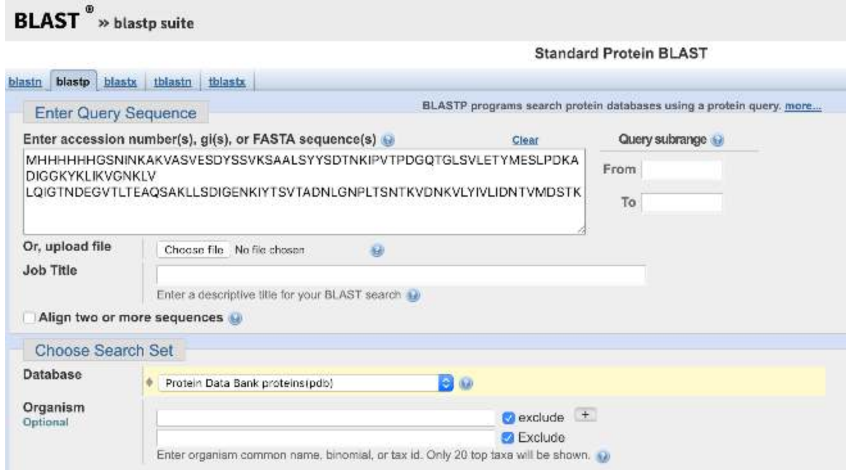
Nmol	%solvent	Matthews	prob(Matthews)
1	85.76435	8.634912	0.0022235168
2	71.5287	4.317456	0.016895106
3	57.29306	2.878304	0.3370472
4	43.05741	2.158728	0.629694
5	28.82175	1.726982	0.0096969195
6	14.58611	1.439152	0.0022216069
7	0.3504574	1.233559	0.0022216022

Sequence: P1;R20921-P1;R20921 pilA1 expressed
]



Search for homologues

PDB/ BLAST



BLAST® » blastp suite

Standard Protein BLAST

Enter Query Sequence

BLASTP programs search protein databases using a protein query. more...

Enter accession number(s), gi(s), or FASTA sequence(s)

```
MHHHHHHGSNINKAKVASVESDYSSVKSAALSYYSDTNKIPVTPDGQTGLSVLETYMESLPDKA
DIGGKYKLIKVGNKLV
LQIGTNDEGVTLTEAQSAKLLSDIGENKIYTSVTADNLGNPLTSNTKVDNKVLYIVLIDNTVMDSTK
```

Query subrange

From

To

Or, upload file: Choose file No file chosen

Job Title

Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

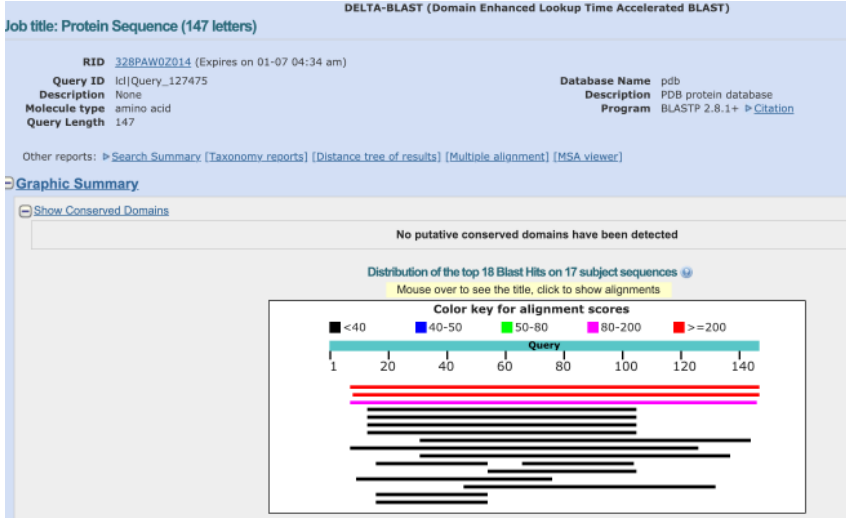
Database: Protein Data Bank proteins(pdb)

Organism
Optional

☑ exclude

☑ Exclude

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.



DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Job title: Protein Sequence (147 letters)

RID: 328PAW0Z014 (Expires on 01-07 04:34 am)

Query ID: lcl|Query_127475
Description: None
Molecule type: amino acid
Query Length: 147

Database Name: pdb
Description: PDB protein database
Program: BLASTP 2.8.1+ ▷ Citation

Other reports: ▷ Search Summary [Taxonomy reports] [Distance tree of results] [Multiple alignment] [MSA viewer]

Graphic Summary

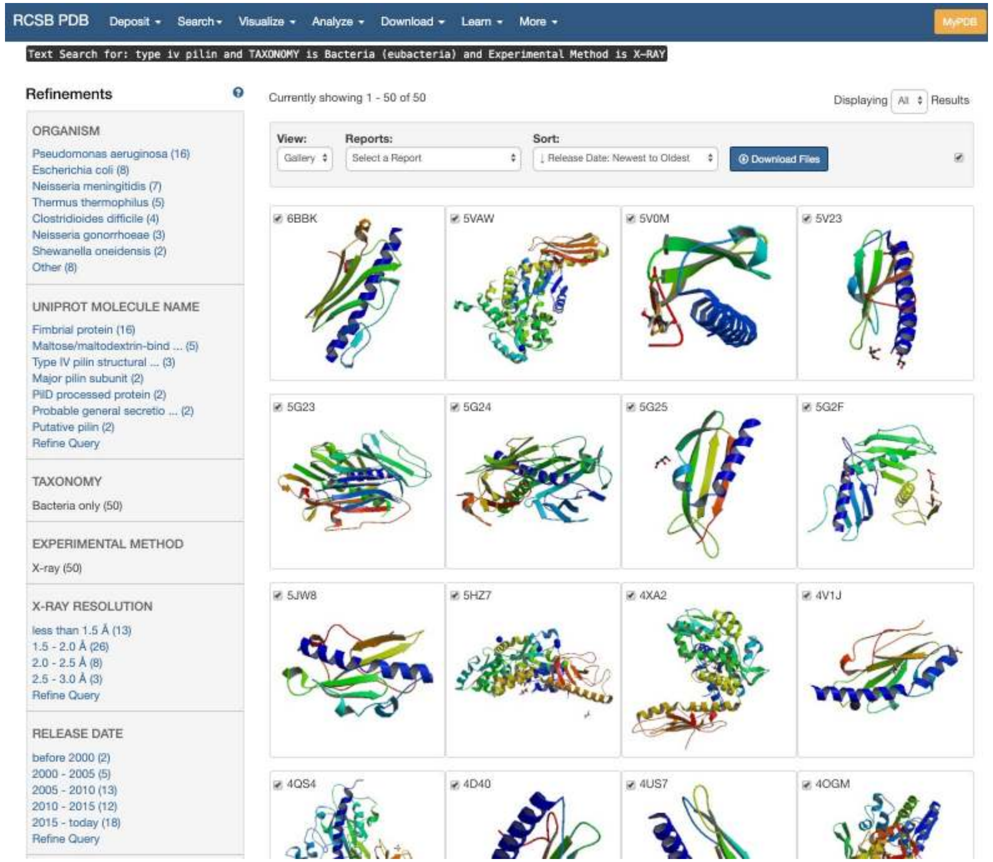
Show Conserved Domains

No putative conserved domains have been detected

Distribution of the top 18 Blast Hits on 17 subject sequences

Mouse over to see the title, click to show alignments

Color key for alignment scores



`Text Search for: type iv pilin and TAXONOMY is Bacteria (eubacteria) and Experimental Method is X-RAY`

Refinements

ORGANISM
Pseudomonas aeruginosa (16)
Escherichia coli (8)
Neisseria meningitidis (7)
Thermus thermophilus (5)
Clostridioides difficile (4)
Neisseria gonorrhoeae (3)
Shewanella oneidensis (2)
Other (8)

UNIPROT MOLECULE NAME
Fimbrial protein (16)
Maltose/maltodextrin-bind ... (5)
Type IV pilin structural ... (3)
Major pilin subunit (2)
PilD processed protein (2)
Probable general secretio ... (2)
Putative pilin (2)
Refine Query

TAXONOMY
Bacteria only (50)

EXPERIMENTAL METHOD
X-ray (50)

X-RAY RESOLUTION
less than 1.5 Å (13)
1.5 - 2.0 Å (26)
2.0 - 2.5 Å (8)
2.5 - 3.0 Å (3)
Refine Query

RELEASE DATE
before 2000 (2)
2000 - 2005 (5)
2005 - 2010 (13)
2010 - 2015 (12)
2015 - today (18)
Refine Query

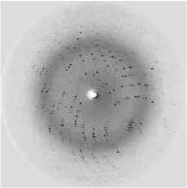
Currently showing 1 - 50 of 50

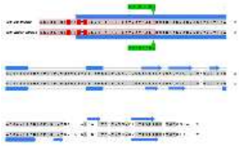
Displaying All Results

View: Gallery

Reports: Select a Report

Sort: ↓ Release Date: Newest to Oldest

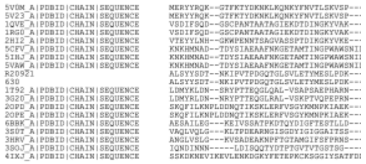


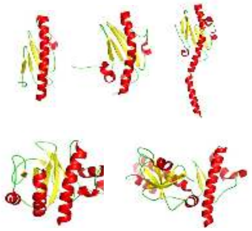


Search for homologues

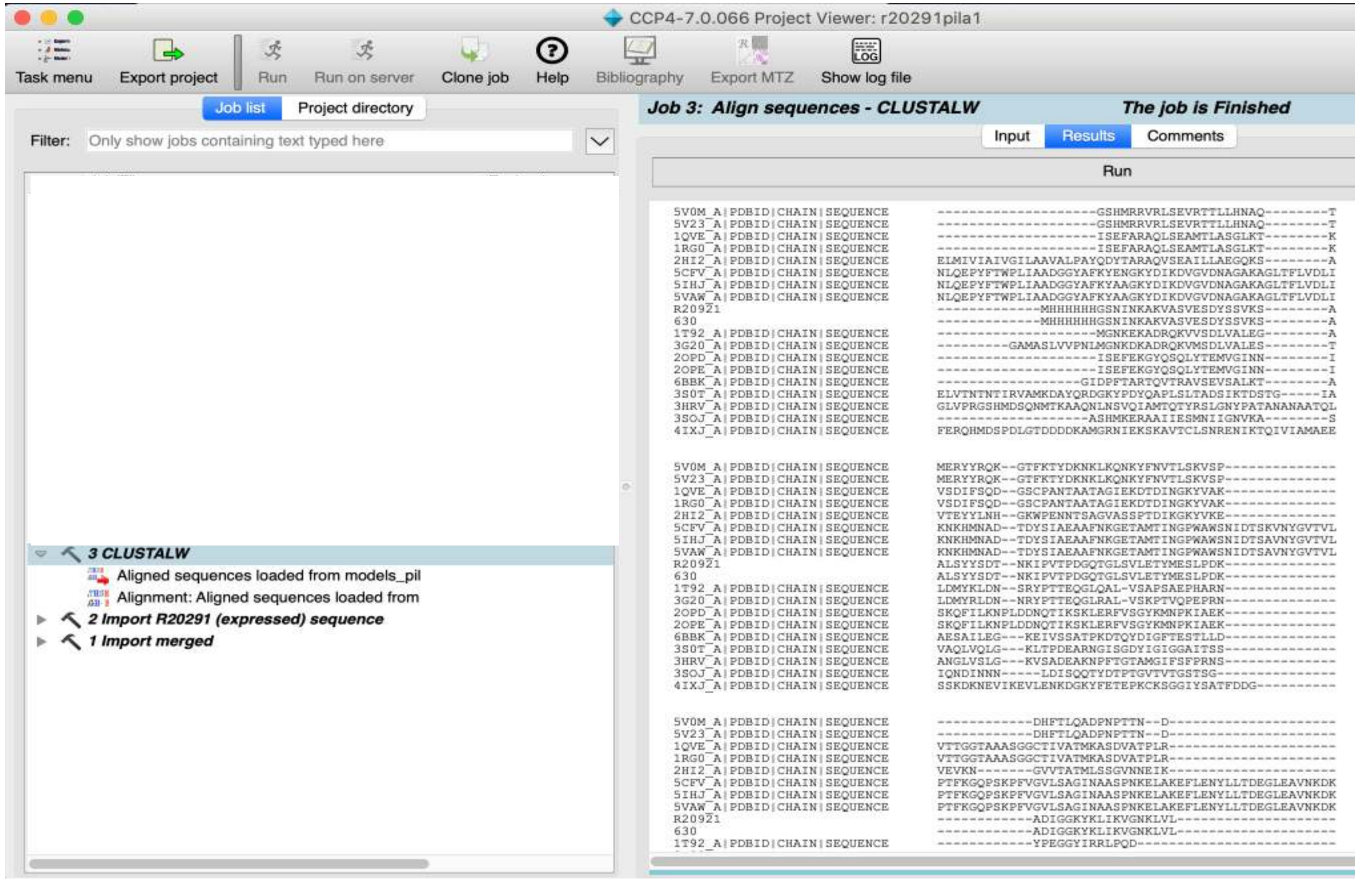
PDB/ BLAST

ClustalW





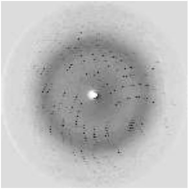
PDB

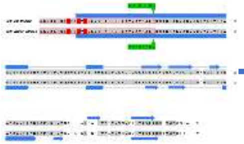


Job 3: Align sequences - CLUSTALW *The job is Finished*

- *3 CLUSTALW*
  - Aligned sequences loaded from models_pil
  - Alignment: Aligned sequences loaded from
- *2 Import R20291 (expressed) sequence*
- *1 Import merged*

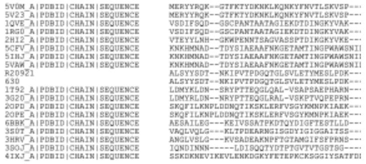
```
5V0M_A|PDBID|CHAIN|SEQUENCE      --------------------GSHMRRVRLSEVRTTLLHNAQ--------T
5V23_A|PDBID|CHAIN|SEQUENCE      --------------------GSHMRRVRLSEVRTTLLHNAQ--------T
1QVE_A|PDBID|CHAIN|SEQUENCE      --------------------ISEFARAQLSEAMTLASGLKT--------K
1RG0_A|PDBID|CHAIN|SEQUENCE      --------------------ISEFARAQLSEAMTLASGLKT--------K
2HI2_A|PDBID|CHAIN|SEQUENCE      EIMIVIAIVGILAAVALPAYQDYTARAQVSEAILLAEGQKS--------A
5CFV_A|PDBID|CHAIN|SEQUENCE      NLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLI
5IHJ_A|PDBID|CHAIN|SEQUENCE      NLQEPYFTWPLIAADGGYAFKYAAGKYDIKDVGVDNAGAKAGLTFLVDLI
5VAW_A|PDBID|CHAIN|SEQUENCE      NLQEPYFTWPLIAADGGYAFKYAAGKYDIKDVGVDNAGAKAGLTFLVDLI
R20921                           -------------MHHHHHHGSNINKAKVASVESDYSSVKS--------A
630                              -------------MHHHHHHGSNINKAKVASVESDYSSVKS--------A
1T92_A|PDBID|CHAIN|SEQUENCE      --------------------MGNKEKADRQKVVSDLVALEG--------A
3G20_A|PDBID|CHAIN|SEQUENCE      ---------GAMASLVVPNLMGNKDKADRQKVMSDLVALES--------T
2OPD_A|PDBID|CHAIN|SEQUENCE      --------------------ISEFEKGYQSQLYTEMVGINN--------I
2OPE_A|PDBID|CHAIN|SEQUENCE      --------------------ISEFEKGYQSQLYTEMVGINN--------I
6BBK_A|PDBID|CHAIN|SEQUENCE      ------------------GIDPFTARTQVTRAVSEVSALKT--------A
3S0T_A|PDBID|CHAIN|SEQUENCE      ELVTNTNTIRVAMKDAYQRDGKYPDYQAPLSLTADSIKTDSTG-----IA
3HRV_A|PDBID|CHAIN|SEQUENCE      GLVPRGSHMDSQNMTKAAQNLNSVQIAMTQTYRSLGNYPATANANAATQL
3SOJ_A|PDBID|CHAIN|SEQUENCE      -------------------ASHMKERAAIIESMNIIGNVKA--------S
4IXJ_A|PDBID|CHAIN|SEQUENCE      FERQHMDSPDLGTDDDDKAMGRNIEKSKAVTCLSNRENIKTQIVIAMAEE

5V0M_A|PDBID|CHAIN|SEQUENCE      MERYYRQK--GTFKTYDKNKLKQNKYFNVTLSKVSP--------------
5V23_A|PDBID|CHAIN|SEQUENCE      MERYYRQK--GTFKTYDKNKLKQNKYFNVTLSKVSP--------------
1QVE_A|PDBID|CHAIN|SEQUENCE      VSDIFSQD--GSCPANTAATAGIEKDTDINGKYVAK--------------
1RG0_A|PDBID|CHAIN|SEQUENCE      VSDIFSQD--GSCPANTAATAGIEKDTDINGKYVAK--------------
2HI2_A|PDBID|CHAIN|SEQUENCE      VTEYYLNH--GKWPENNTSAGVASSPTDIKGKYVKE--------------
5CFV_A|PDBID|CHAIN|SEQUENCE      KNKHMNAD--TDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVL
5IHJ_A|PDBID|CHAIN|SEQUENCE      KNKHMNAD--TDYSIAEAAFNKGETAMTINGPWAWSNIDTSAVNYGVTVL
5VAW_A|PDBID|CHAIN|SEQUENCE      KNKHMNAD--TDYSIAEAAFNKGETAMTINGPWAWSNIDTSAVNYGVTVL
R20921                           ALSYYSDT--NKIPVTPDGQTGLSVLETYMESLPDK--------------
630                              ALSYYSDT--NKIPVTPDGQTGLSVLETYMESLPDK--------------
1T92_A|PDBID|CHAIN|SEQUENCE      LDMYKLDN--SRYPTTEQGLQAL-VSAPSAEPHARN--------------
3G20_A|PDBID|CHAIN|SEQUENCE      LDMYRLDN--NRYPTTEQGLRAL-VSKPTVQPEPRN--------------
2OPD_A|PDBID|CHAIN|SEQUENCE      SKQFILKNPLDDNQTIKSKLERFVSGYKMNPKIAEK--------------
2OPE_A|PDBID|CHAIN|SEQUENCE      SKQFILKNPLDDNQTIKSKLERFVSGYKMNPKIAEK--------------
6BBK_A|PDBID|CHAIN|SEQUENCE      AESAILEG---KEIVSSATPKDTQYDIGFTESTLLD--------------
3S0T_A|PDBID|CHAIN|SEQUENCE      VAQLVQLG---KLTPDEARNGISGDYIGIGGAITSS--------------
3HRV_A|PDBID|CHAIN|SEQUENCE      ANGLVSLG---KVSADEAKNPFTGTAMGIFSFPRNS--------------
3SOJ_A|PDBID|CHAIN|SEQUENCE      IQNDINNN-----LDISQQTYDTPTGVTVTGSTSG---------------
4IXJ_A|PDBID|CHAIN|SEQUENCE      SSKDKNEVIKEVLENKDGKYFETEPKCKSGGIYSATFDDG----------

5V0M_A|PDBID|CHAIN|SEQUENCE      ------------DHFTLQADPNPTTN--D---------------------
5V23_A|PDBID|CHAIN|SEQUENCE      ------------DHFTLQADPNPTTN--D---------------------
1QVE_A|PDBID|CHAIN|SEQUENCE      VTTGGTAAASGGCTIVATMKASDVATPLR---------------------
1RG0_A|PDBID|CHAIN|SEQUENCE      VTTGGTAAASGGCTIVATMKASDVATPLR---------------------
2HI2_A|PDBID|CHAIN|SEQUENCE      VEVKN-------GVVTATMLSSGVNNEIK---------------------
5CFV_A|PDBID|CHAIN|SEQUENCE      PTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDK
5IHJ_A|PDBID|CHAIN|SEQUENCE      PTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDK
5VAW_A|PDBID|CHAIN|SEQUENCE      PTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDK
R20921                           ------------ADIGGKYKLIKVGNKLVL--------------------
630                              ------------ADIGGKYKLIKVGNKLVL--------------------
1T92_A|PDBID|CHAIN|SEQUENCE      ------------YPEGGYIRRLPQD-------------------------
```



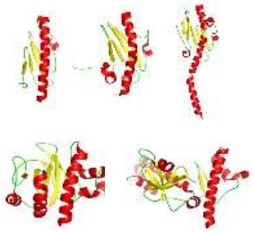


Search for homologues

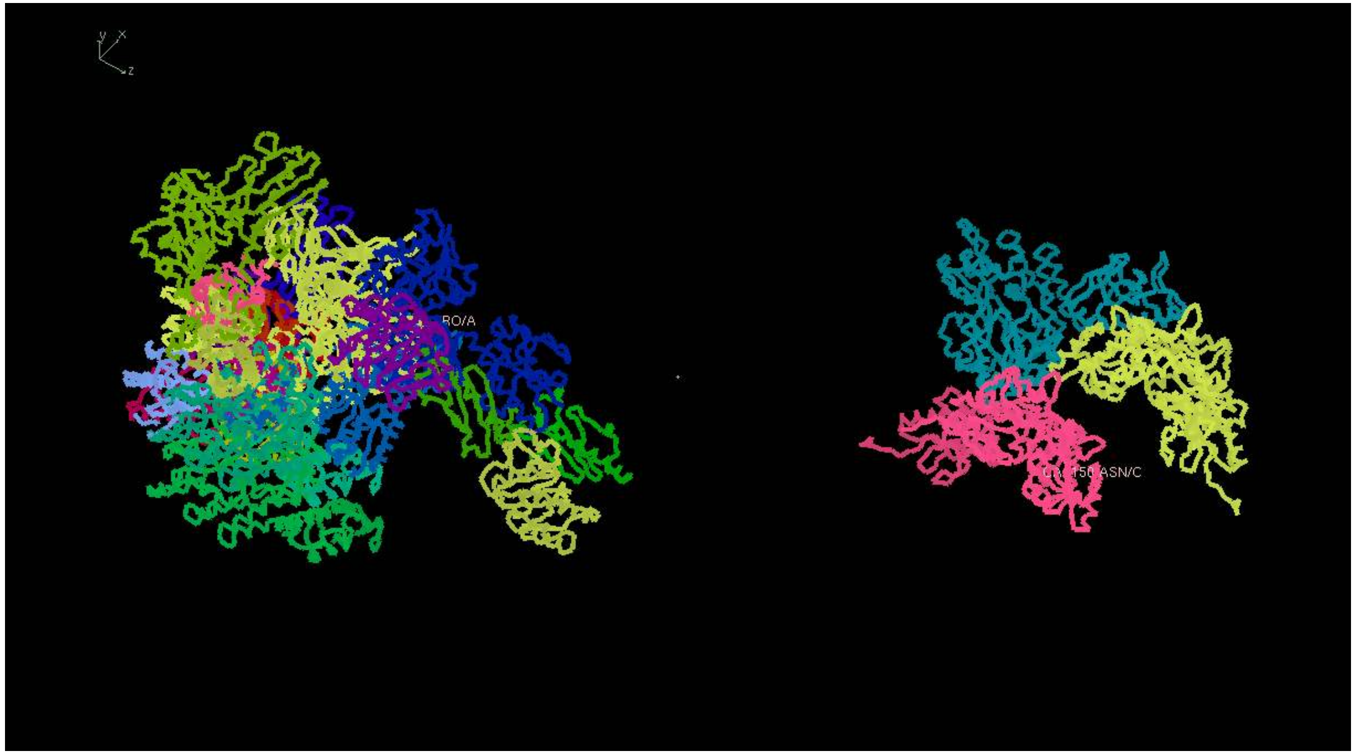
ClustalW

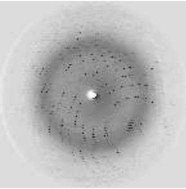


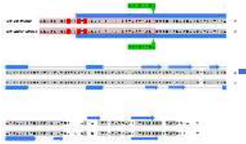
```
5VDM_A|PDBID|CHAIN|SEQUENCE      MERYYRQK--GTFKTYDKNKLKQNKYFNVTLSKVSP--
5V23_A|PDBID|CHAIN|SEQUENCE      MERYYRQK--GTFKTYDKNKLKQNKYFNVTLSKVSP--
1QVE_A|PDBID|CHAIN|SEQUENCE      VSDIFSQD--GSCPANTAATAGIEKDTDINGKYVAK--
1RGD_A|PDBID|CHAIN|SEQUENCE      VSDIFSQD--GSCPANTAATAGIEKDTDINGKYVAK--
2HI2_A|PDBID|CHAIN|SEQUENCE      VTEYYLNH--GKWPENNTSAGVASSPTDIKGKYVKE--
5CFV_A|PDBID|CHAIN|SEQUENCE      KNKHMNAD--TDYSIAEAAFNKGETAMTINGPWAWSNI
5IHJ_A|PDBID|CHAIN|SEQUENCE      KNKHMNAD--TDYSIAEAAFNKGETAMTINGPWAWSNI
5VAW_A|PDBID|CHAIN|SEQUENCE      KNKHMNAD--TDYSIAEAAFNKGETAMTINGPWAWSNI
R20921                           ALSYYSDT--NKIPVTPDGQTGLSVLETYMESLPDK--
630                              ALSYYSDT--NKIPVTPDGQTGLSVLETYMESLPDK--
1T92_A|PDBID|CHAIN|SEQUENCE      LDMYKLDN--SRYPTTEQGLQAL-VSAPSAEPHARN--
3G20_A|PDBID|CHAIN|SEQUENCE      LDMYRLDN--NRYPTTEQGLRAL-VSKPTVQPEPRN--
2OPD_A|PDBID|CHAIN|SEQUENCE      SKQFILKNPLDDNQTIKSKLERFVSGYKMNPKIAEK--
2OPE_A|PDBID|CHAIN|SEQUENCE      SKQFILKNPLDDNQTIKSKLERFVSGYKMNPKIAEK--
6BBK_A|PDBID|CHAIN|SEQUENCE      AESAILEG---KEIVSSATPKDTQYDIGFTESTLLD--
3SDT_A|PDBID|CHAIN|SEQUENCE      VAQLVQLG---KLTPDEARNGISGDYIGIGGAITSS--
3HRV_A|PDBID|CHAIN|SEQUENCE      ANGLVSLG---KVSADEAKNPFTGTAMGIFSFPRNS--
3SOJ_A|PDBID|CHAIN|SEQUENCE      IQNDINNN-----LDISQQTYDTPTGVTVTGSTSG---
4IXJ_A|PDBID|CHAIN|SEQUENCE      SSKDKNEVIKEVLENKDGKYFETEPKCKSGGIYSATFD
```



PDB



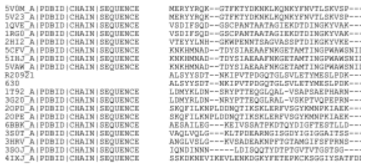




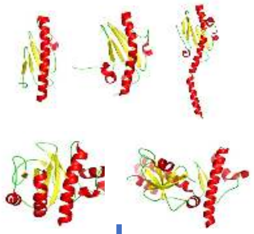
Search for homologues

PDB

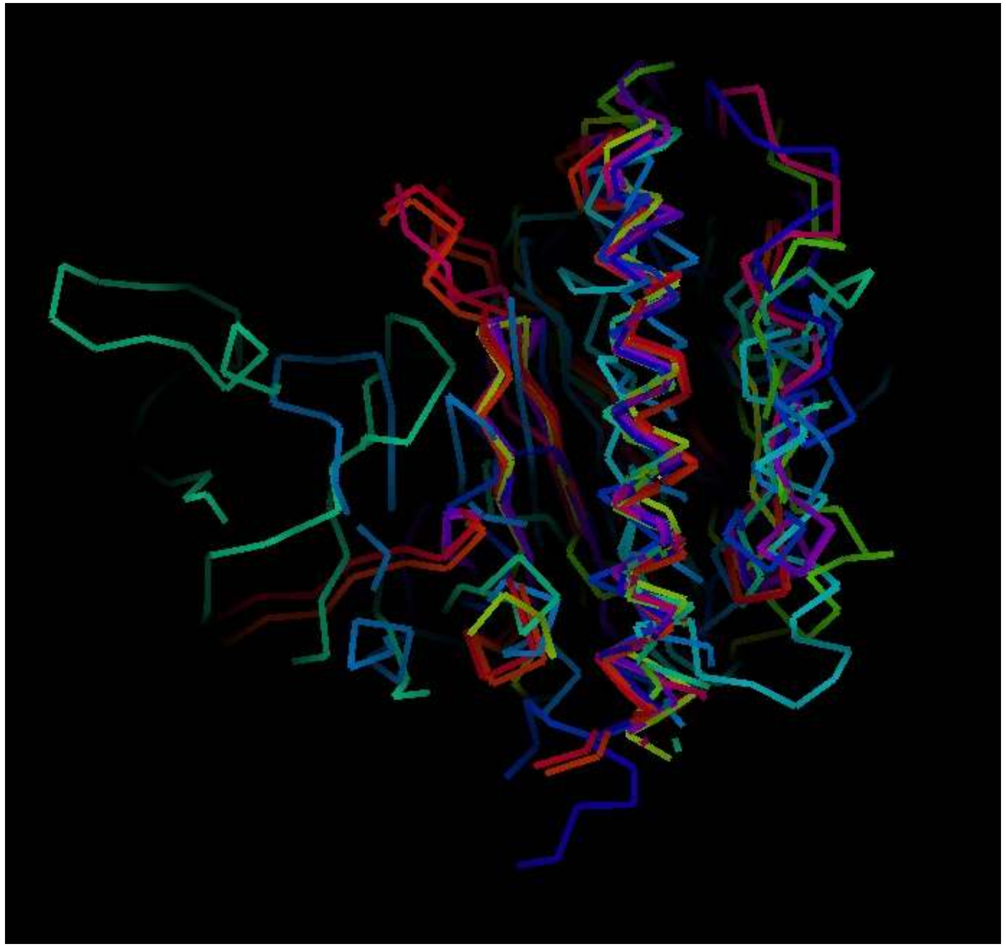
ClustalW



```
5V0M_A|PDBID|CHAIN|SEQUENCE      MERYYRQK--GTFKTYDKNKLKQNKYFNVTLSKVSP---
5V23_A|PDBID|CHAIN|SEQUENCE      MERYYRQK--GTFKTYDKNKLKQNKYFNVTLSKVSP---
1QVE_A|PDBID|CHAIN|SEQUENCE      VSDIFSQD--GSCPANTAATAGIEKDTDINGKYVAK---
1RG0_A|PDBID|CHAIN|SEQUENCE      VSDIFSQD--GSCPANTAATAGIEKDTDINGKYVAK---
2HI2_A|PDBID|CHAIN|SEQUENCE      VTEYYLNH--GKWPENNTSAGVASSPTDIKGKYVKE---
5CFV_A|PDBID|CHAIN|SEQUENCE      KNKHMNAD--TDYSIAEAAFNKGETAMTINGPWAWSNII
5IHJ_A|PDBID|CHAIN|SEQUENCE      KNKHMNAD--TDYSIAEAAFNKGETAMTINGPWAWSNII
5VAW_A|PDBID|CHAIN|SEQUENCE      KNKHMNAD--TDYSIAEAAFNKGETAMTINGPWAWSNII
R20921                           ALSYYSDT--NKIPVTPDGQTGLSVLETYMESLPDK---
630                              ALSYYSDT--NKIPVTPDGQTGLSVLETYMESLPDK---
1T92_A|PDBID|CHAIN|SEQUENCE      LDMYKLDN--SRYPTTEQGLQAL-VSAPSAEPHARN---
3G20_A|PDBID|CHAIN|SEQUENCE      LDMYRLDN--NRYPTTEQGLRAL-VSKPTVQPEPRN---
2OPD_A|PDBID|CHAIN|SEQUENCE      SKQFILKNPLDDNQTIKSKLERFVSGYKMNPKIAEK---
2OPE_A|PDBID|CHAIN|SEQUENCE      SKQFILKNPLDDNQTIKSKLERFVSGYKMNPKIAEK---
6BBK_A|PDBID|CHAIN|SEQUENCE      AESAILEG---KEIVSSATPKDTQYDIGFTESTLLD---
3S0T_A|PDBID|CHAIN|SEQUENCE      VAQLVQLG---KLTPDEARNGISGDYIGIGGAITSS---
3HRV_A|PDBID|CHAIN|SEQUENCE      ANGLVSLG---KVSADEAKNPFTGTAMGIFSFPRNS---
3SOJ_A|PDBID|CHAIN|SEQUENCE      IQNDINNN-----LDISQQTYDTPTGVTVTGSTSG---
4IXJ_A|PDBID|CHAIN|SEQUENCE      SSKDKNEVIKEVLENKDGKYFETEPKCKSGGIYSATFDI
```

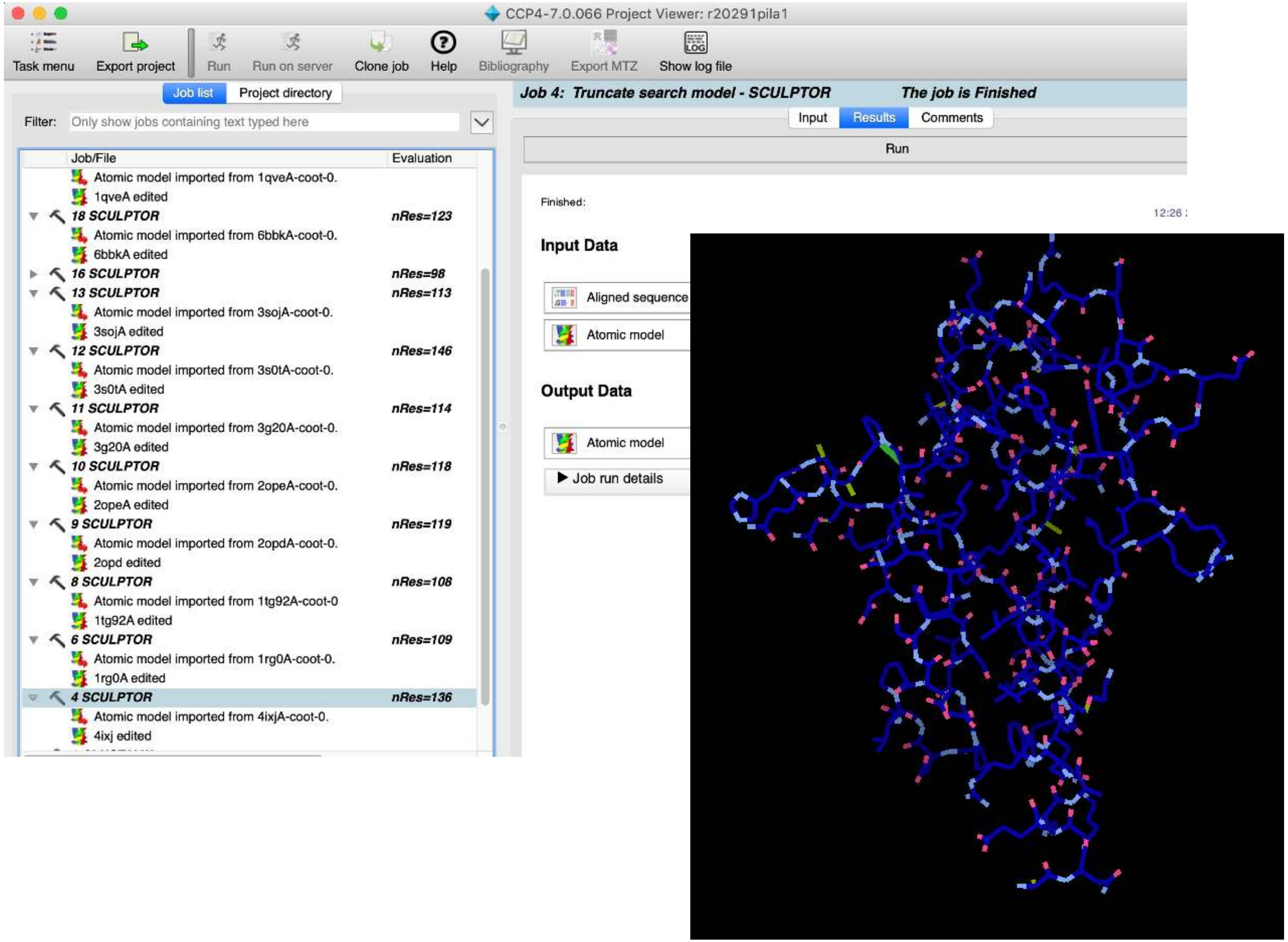


Model preparation



Sculptor

Model preparation



CCP4-7.0.066 Project Viewer: r20291pila1

Job/File	Evaluation
Atomic model imported from 1qveA-coot-0.	
1qveA edited	
18 SCULPTOR	*nRes=123*
Atomic model imported from 6bbkA-coot-0.	
6bbkA edited	
16 SCULPTOR	*nRes=98*
13 SCULPTOR	*nRes=113*
Atomic model imported from 3sojA-coot-0.	
3sojA edited	
12 SCULPTOR	*nRes=146*
Atomic model imported from 3s0tA-coot-0.	
3s0tA edited	
11 SCULPTOR	*nRes=114*
Atomic model imported from 3g20A-coot-0.	
3g20A edited	
10 SCULPTOR	*nRes=118*
Atomic model imported from 2opeA-coot-0.	
2opeA edited	
9 SCULPTOR	*nRes=119*
Atomic model imported from 2opdA-coot-0.	
2opd edited	
8 SCULPTOR	*nRes=108*
Atomic model imported from 1tg92A-coot-0	
1tg92A edited	
6 SCULPTOR	*nRes=109*
Atomic model imported from 1rg0A-coot-0.	
1rg0A edited	
4 SCULPTOR	*nRes=136*
Atomic model imported from 4ixjA-coot-0.	
4ixj edited	

Job 4: Truncate search model - SCULPTOR *The job is Finished*

Finished: 12:26

Input Data

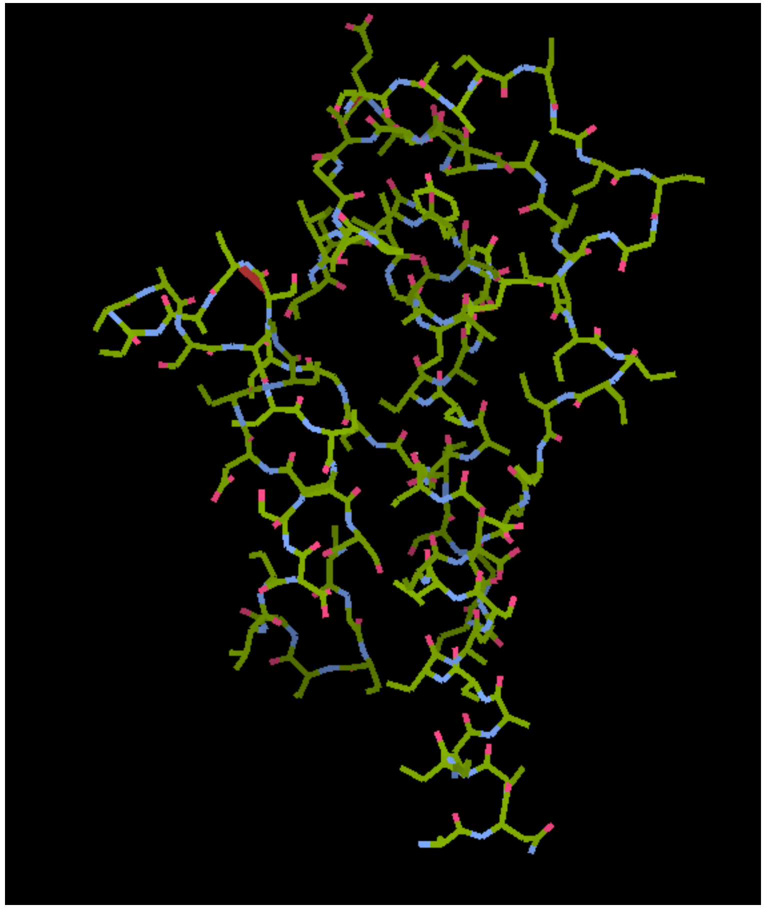
Aligned sequence

Atomic model

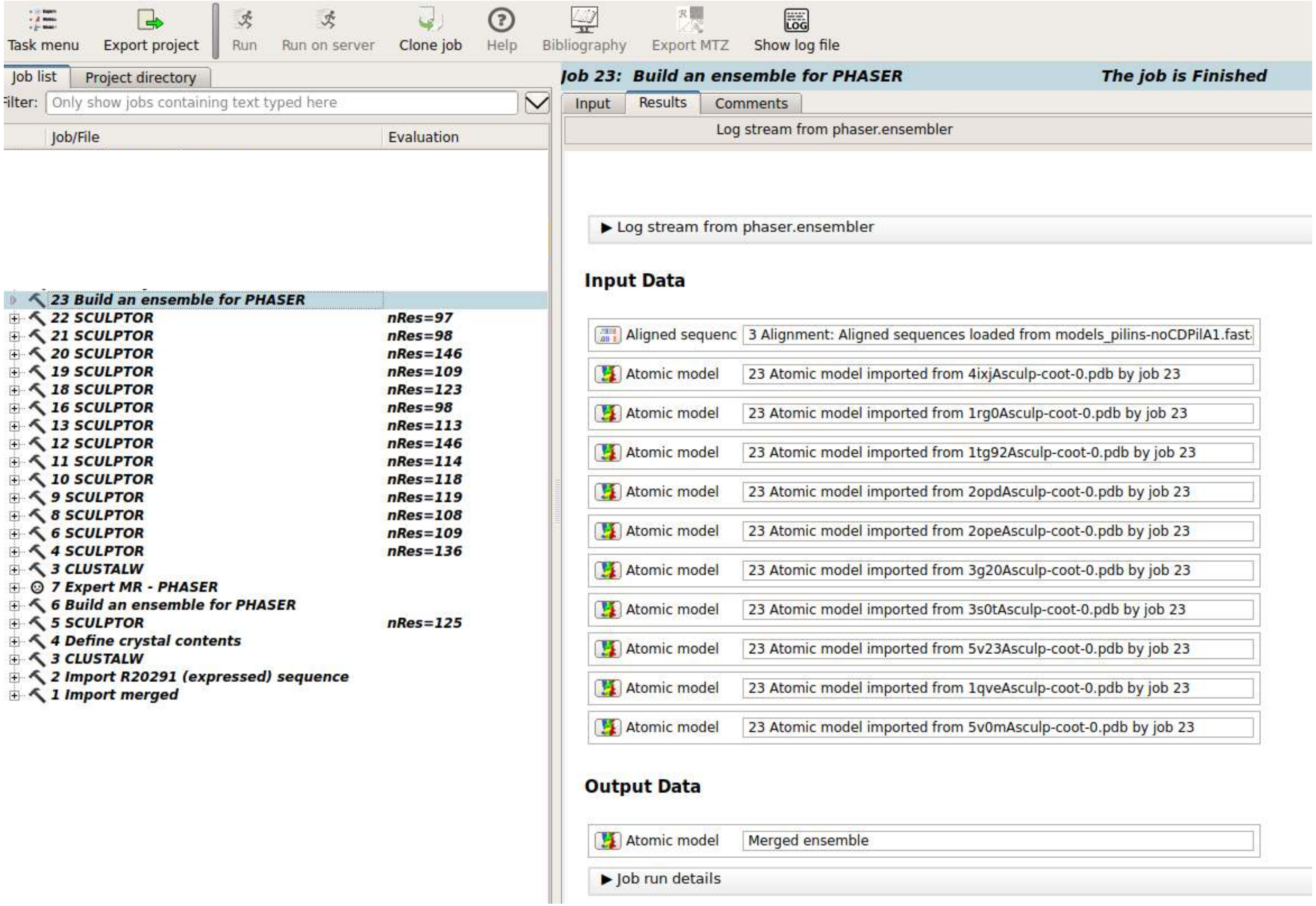
Output Data

Atomic model

▶ Job run details



Model preparation



Job 23: Build an ensemble for PHASER — **The job is Finished**

Job/File	Evaluation
23 Build an ensemble for PHASER	
22 SCULPTOR	**nRes=97**
21 SCULPTOR	**nRes=98**
20 SCULPTOR	**nRes=146**
19 SCULPTOR	**nRes=109**
18 SCULPTOR	**nRes=123**
16 SCULPTOR	**nRes=98**
13 SCULPTOR	**nRes=113**
12 SCULPTOR	**nRes=146**
11 SCULPTOR	**nRes=114**
10 SCULPTOR	**nRes=118**
9 SCULPTOR	**nRes=119**
8 SCULPTOR	**nRes=108**
6 SCULPTOR	**nRes=109**
4 SCULPTOR	**nRes=136**
3 CLUSTALW	
7 Expert MR - PHASER	
6 Build an ensemble for PHASER	
5 SCULPTOR	**nRes=125**
4 Define crystal contents	
3 CLUSTALW	
2 Import R20291 (expressed) sequence	
1 Import merged	

Log stream from phaser.ensembler

▶ Log stream from phaser.ensembler

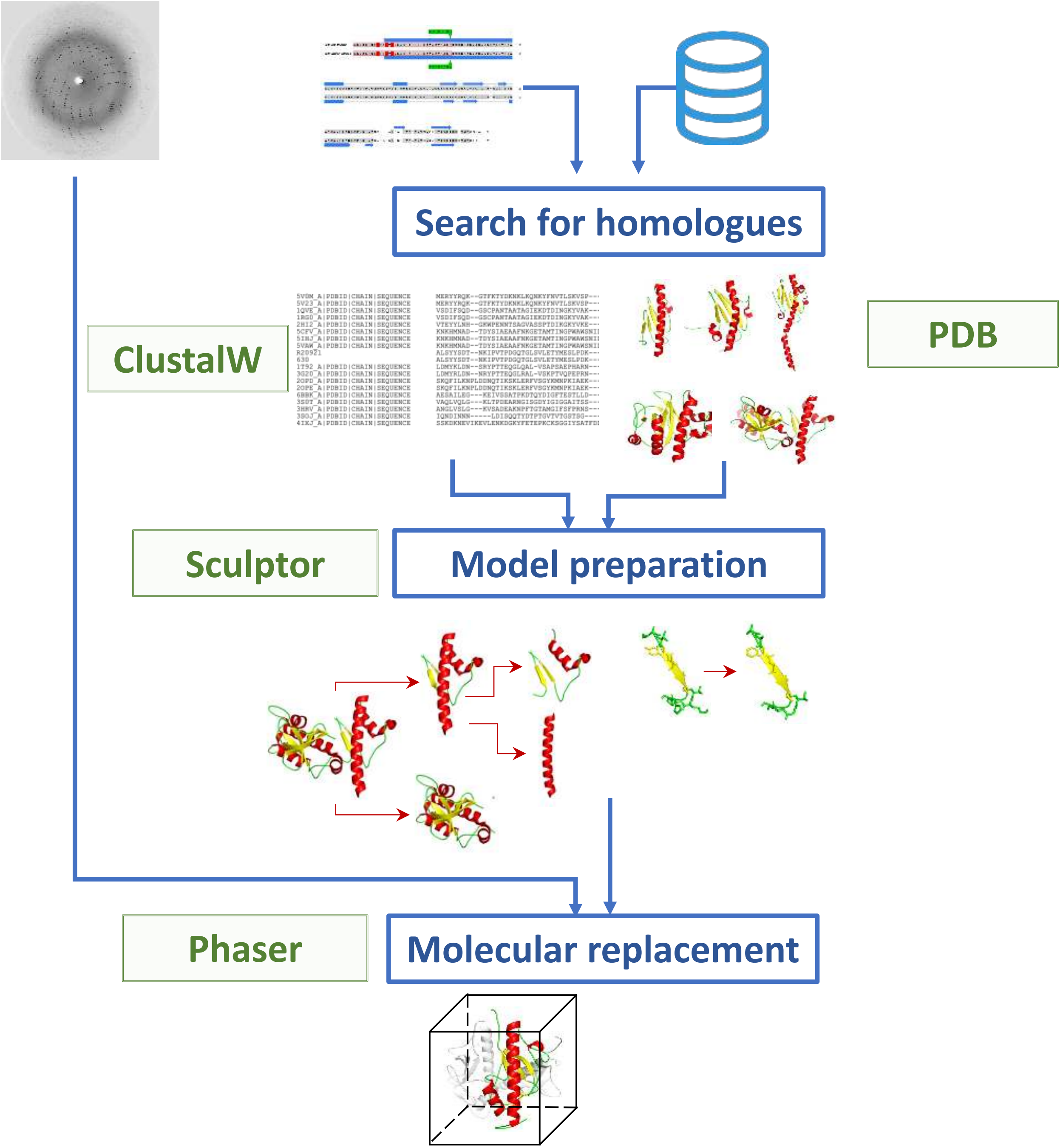
Input Data

Aligned sequenc	3 Alignment: Aligned sequences loaded from models_pilins-noCDPilA1.fast
Atomic model	23 Atomic model imported from 4ixjAsculp-coot-0.pdb by job 23
Atomic model	23 Atomic model imported from 1rg0Asculp-coot-0.pdb by job 23
Atomic model	23 Atomic model imported from 1tg92Asculp-coot-0.pdb by job 23
Atomic model	23 Atomic model imported from 2opdAsculp-coot-0.pdb by job 23
Atomic model	23 Atomic model imported from 2opeAsculp-coot-0.pdb by job 23
Atomic model	23 Atomic model imported from 3g20Asculp-coot-0.pdb by job 23
Atomic model	23 Atomic model imported from 3s0tAsculp-coot-0.pdb by job 23
Atomic model	23 Atomic model imported from 5v23Asculp-coot-0.pdb by job 23
Atomic model	23 Atomic model imported from 1qveAsculp-coot-0.pdb by job 23
Atomic model	23 Atomic model imported from 5v0mAsculp-coot-0.pdb by job 23

Output Data

Atomic model	Merged ensemble

▶ Job run details



Search for homologues

ClustalW

PDB

Sculptor

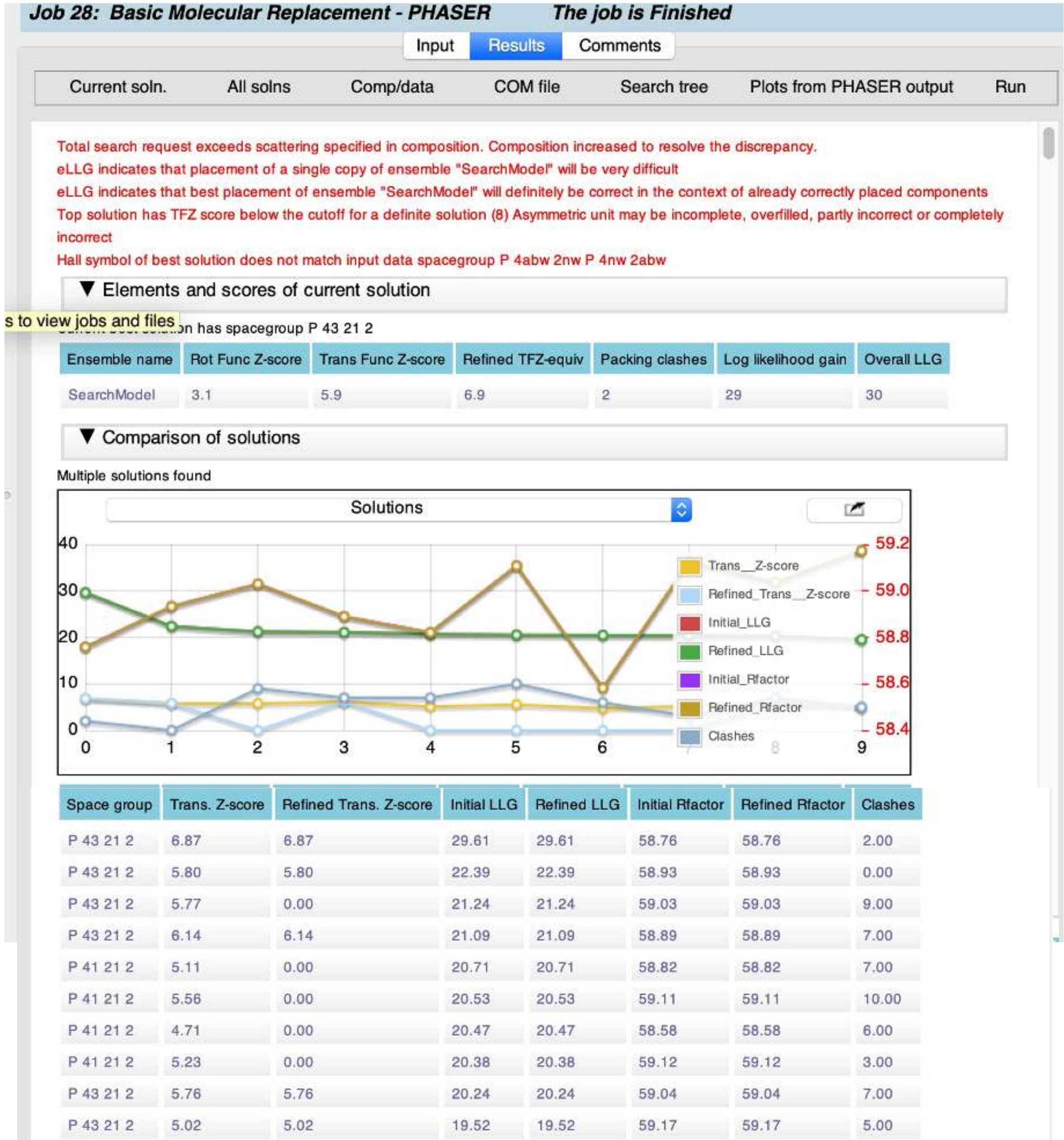
Model preparation

Phaser

Molecular replacement

Phaser

Molecular replacement



Job 28: Basic Molecular Replacement - PHASER **The job is Finished**

Total search request exceeds scattering specified in composition. Composition increased to resolve the discrepancy.
eLLG indicates that placement of a single copy of ensemble "SearchModel" will be very difficult
eLLG indicates that best placement of ensemble "SearchModel" will definitely be correct in the context of already correctly placed components
Top solution has TFZ score below the cutoff for a definite solution (8) Asymmetric unit may be incomplete, overfilled, partly incorrect or completely incorrect
Hall symbol of best solution does not match input data spacegroup P 4abw 2nw P 4nw 2abw

▼ Elements and scores of current solution

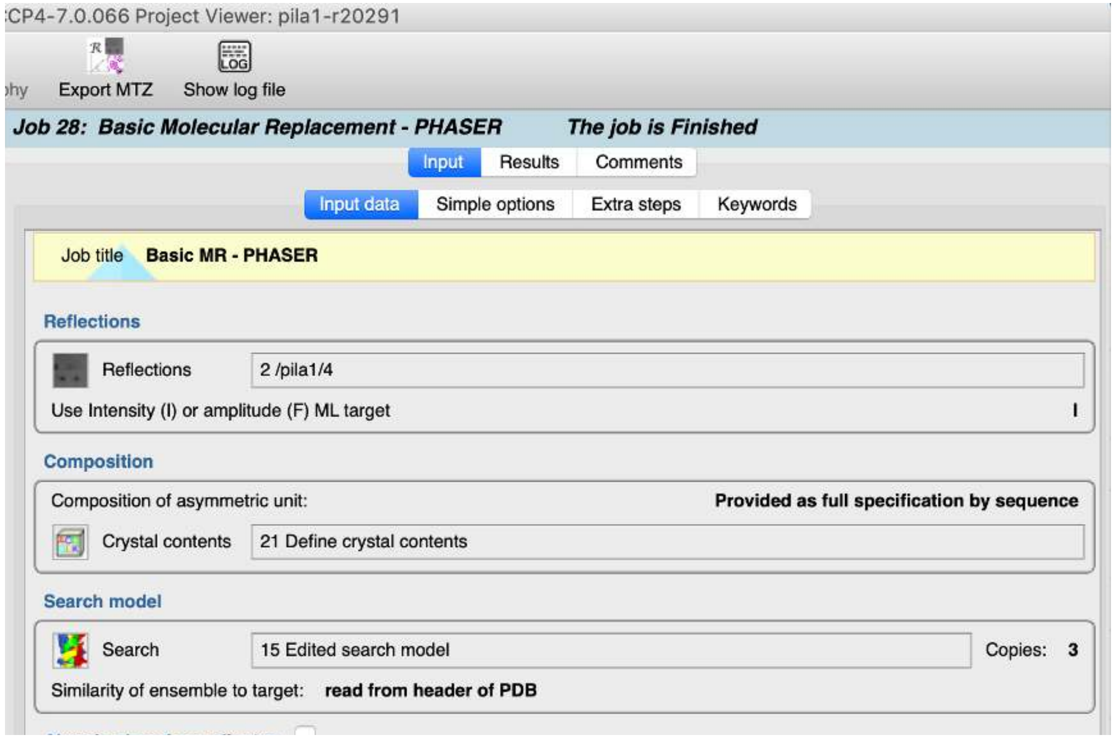
s to view jobs and files ...on has spacegroup P 43 21 2

Ensemble name	Rot Func Z-score	Trans Func Z-score	Refined TFZ-equiv	Packing clashes	Log likelihood gain	Overall LLG
SearchModel	3.1	5.9	6.9	2	29	30

▼ Comparison of solutions

Multiple solutions found

Space group	Trans. Z-score	Refined Trans. Z-score	Initial LLG	Refined LLG	Initial Rfactor	Refined Rfactor	Clashes
P 43 21 2	6.87	6.87	29.61	29.61	58.76	58.76	2.00
P 43 21 2	5.80	5.80	22.39	22.39	58.93	58.93	0.00
P 43 21 2	5.77	0.00	21.24	21.24	59.03	59.03	9.00
P 43 21 2	6.14	6.14	21.09	21.09	58.89	58.89	7.00
P 41 21 2	5.11	0.00	20.71	20.71	58.82	58.82	7.00
P 41 21 2	5.56	0.00	20.53	20.53	59.11	59.11	10.00
P 41 21 2	4.71	0.00	20.47	20.47	58.58	58.58	6.00
P 41 21 2	5.23	0.00	20.38	20.38	59.12	59.12	3.00
P 43 21 2	5.76	5.76	20.24	20.24	59.04	59.04	7.00
P 43 21 2	5.02	5.02	19.52	19.52	59.17	59.17	5.00



:CP4-7.0.066 Project Viewer: pila1-r20291

Job 28: Basic Molecular Replacement - PHASER **The job is Finished**

Job title **Basic MR - PHASER**

Reflections

Reflections: 2 /pila1/4

Use Intensity (I) or amplitude (F) ML target: I

Composition

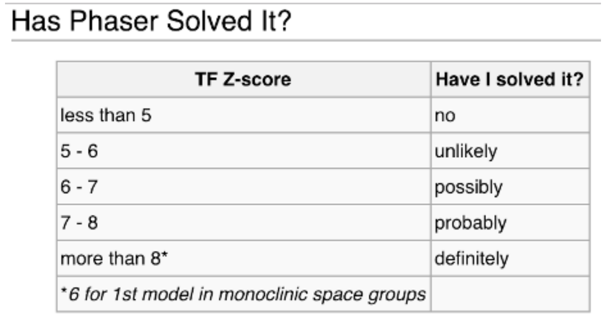
Composition of asymmetric unit: **Provided as full specification by sequence**

Crystal contents: 21 Define crystal contents

Search model

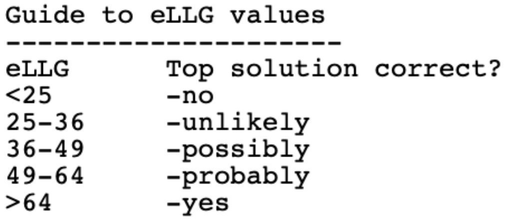
Search: 15 Edited search model Copies: 3

Similarity of ensemble to target: **read from header of PDB**

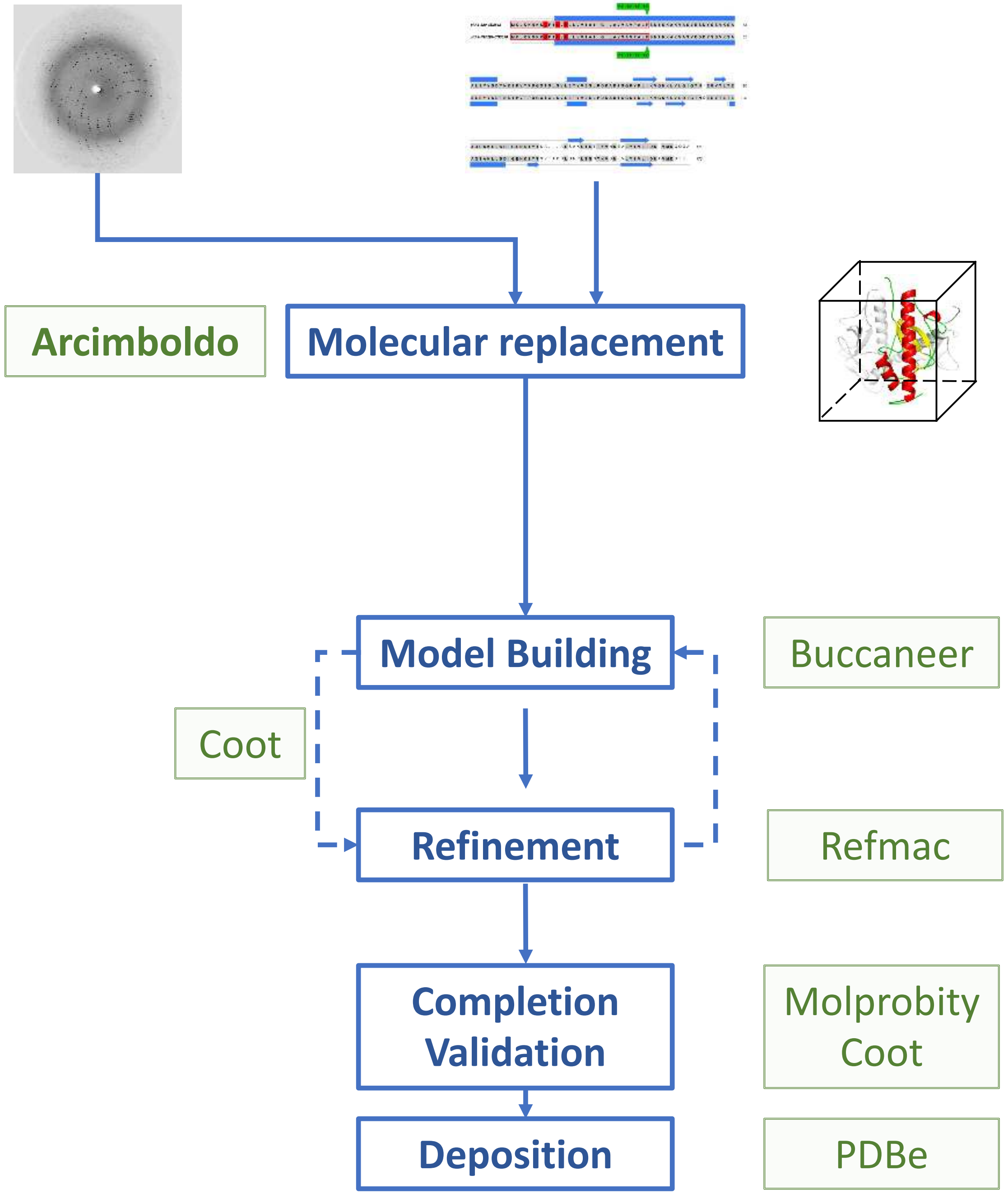


Has Phaser Solved It?

TF Z-score	Have I solved it?
less than 5	no
5 - 6	unlikely
6 - 7	possibly
7 - 8	probably
more than 8*	definitely
*6 for 1st model in monoclinic space groups	



```
Guide to eLLG values
--------------------
eLLG       Top solution correct?
<25        -no
25-36      -unlikely
36-49      -possibly
49-64      -probably
>64        -yes
```



Arcimboldo

Molecular replacement

Model Building

Buccaneer

Coot

Refinement

Refmac

Completion
Validation

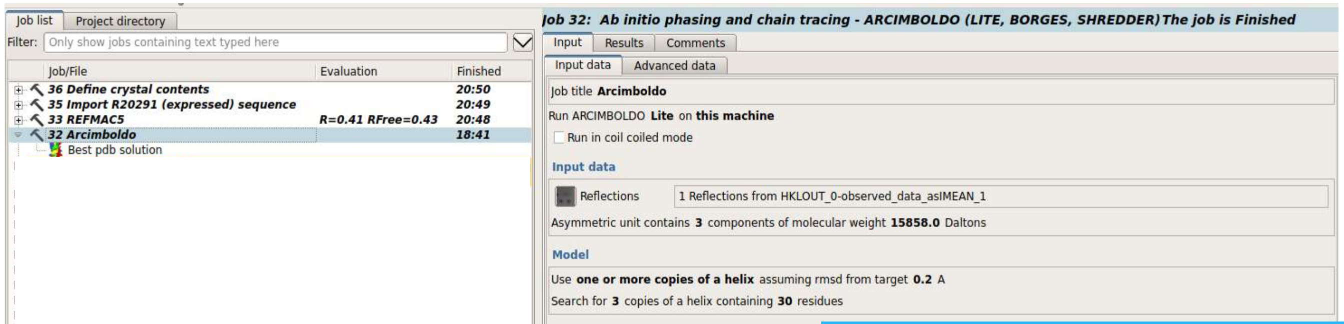
Molprobity
Coot

Deposition

PDBe

Arcimboldo

Molecular replacement



Job/File	Evaluation	Finished
36 Define crystal contents		***20:50***
35 Import R20291 (expressed) sequence		***20:49***
33 REFMAC5	***R=0.41 RFree=0.43***	***20:48***
32 Arcimboldo		***18:41***
Best pdb solution		

Job 32: Ab initio phasing and chain tracing - ARCIMBOLDO (LITE, BORGES, SHREDDER) The job is Finished

Job title **Arcimboldo**

Run ARCIMBOLDO **Lite** on **this machine**

☐ Run in coil coiled mode

Input data

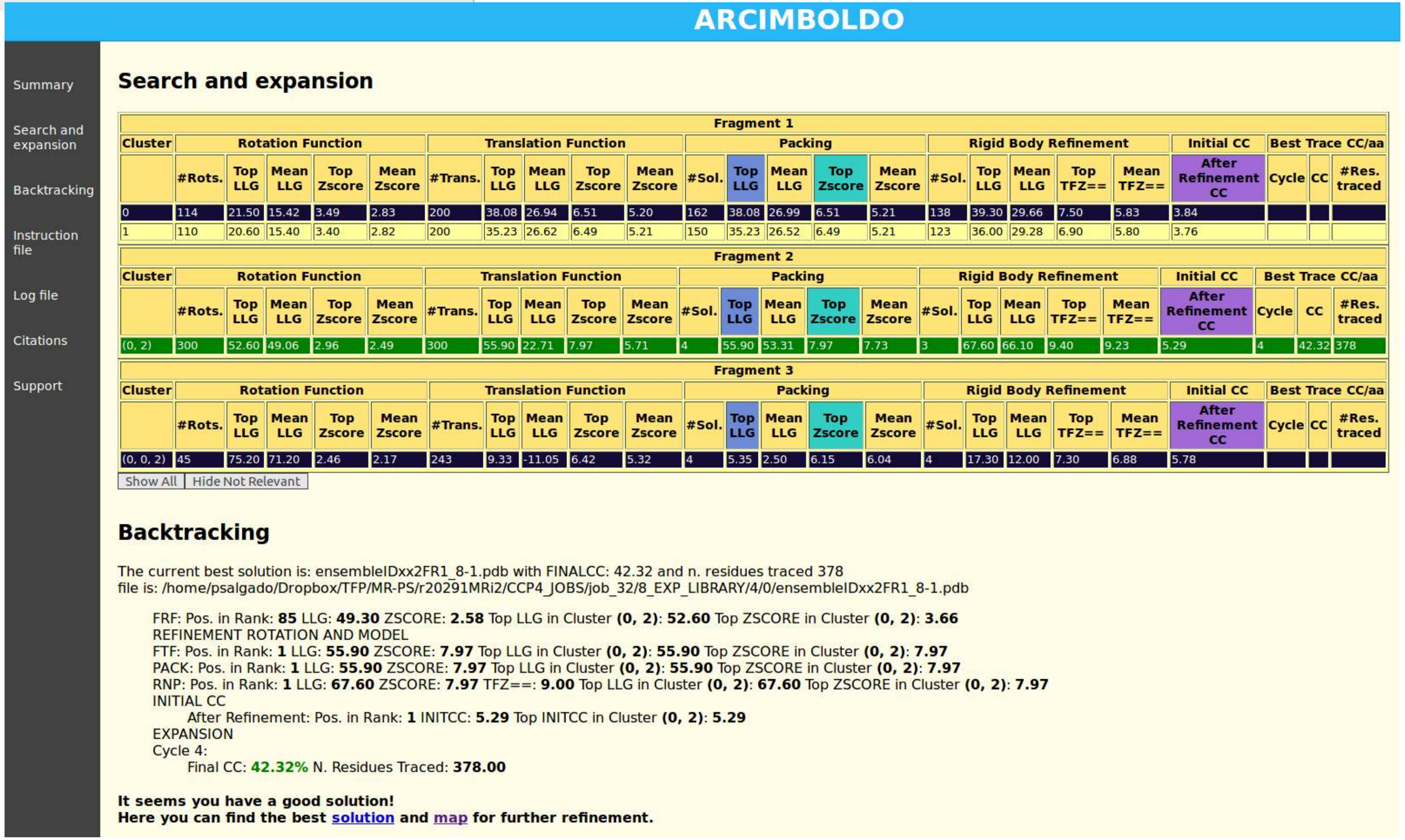
Reflections: 1 Reflections from HKLOUT_0-observed_data_asIMEAN_1

Asymmetric unit contains **3** components of molecular weight **15858.0** Daltons

Model

Use **one or more copies of a helix** assuming rmsd from target **0.2** A

Search for **3** copies of a helix containing **30** residues



Search and expansion

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	Top TFZ==	Mean TFZ==	After Refinement CC	Cycle	CC	#Res. traced
0	114	21.50	15.42	3.49	2.83	200	38.08	26.94	6.51	5.20	162	38.08	26.99	6.51	5.21	138	39.30	29.66	7.50	5.83	3.84			
1	110	20.60	15.40	3.40	2.82	200	35.23	26.62	6.49	5.21	150	35.23	26.52	6.49	5.21	123	36.00	29.28	6.90	5.80	3.76			

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	Top TFZ==	Mean TFZ==	After Refinement CC	Cycle	CC	#Res. traced
(0, 2)	300	52.60	49.06	2.96	2.49	300	55.90	22.71	7.97	5.71	4	55.90	53.31	7.97	7.73	3	67.60	66.10	9.40	9.23	5.29	4	42.32	378

Fragment 3																								
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	Top TFZ==	Mean TFZ==	After Refinement CC	Cycle	CC	#Res. traced
(0, 0, 2)	45	75.20	71.20	2.46	2.17	243	9.33	-11.05	6.42	5.32	4	5.35	2.50	6.15	6.04	4	17.30	12.00	7.30	6.88	5.78			

Backtracking

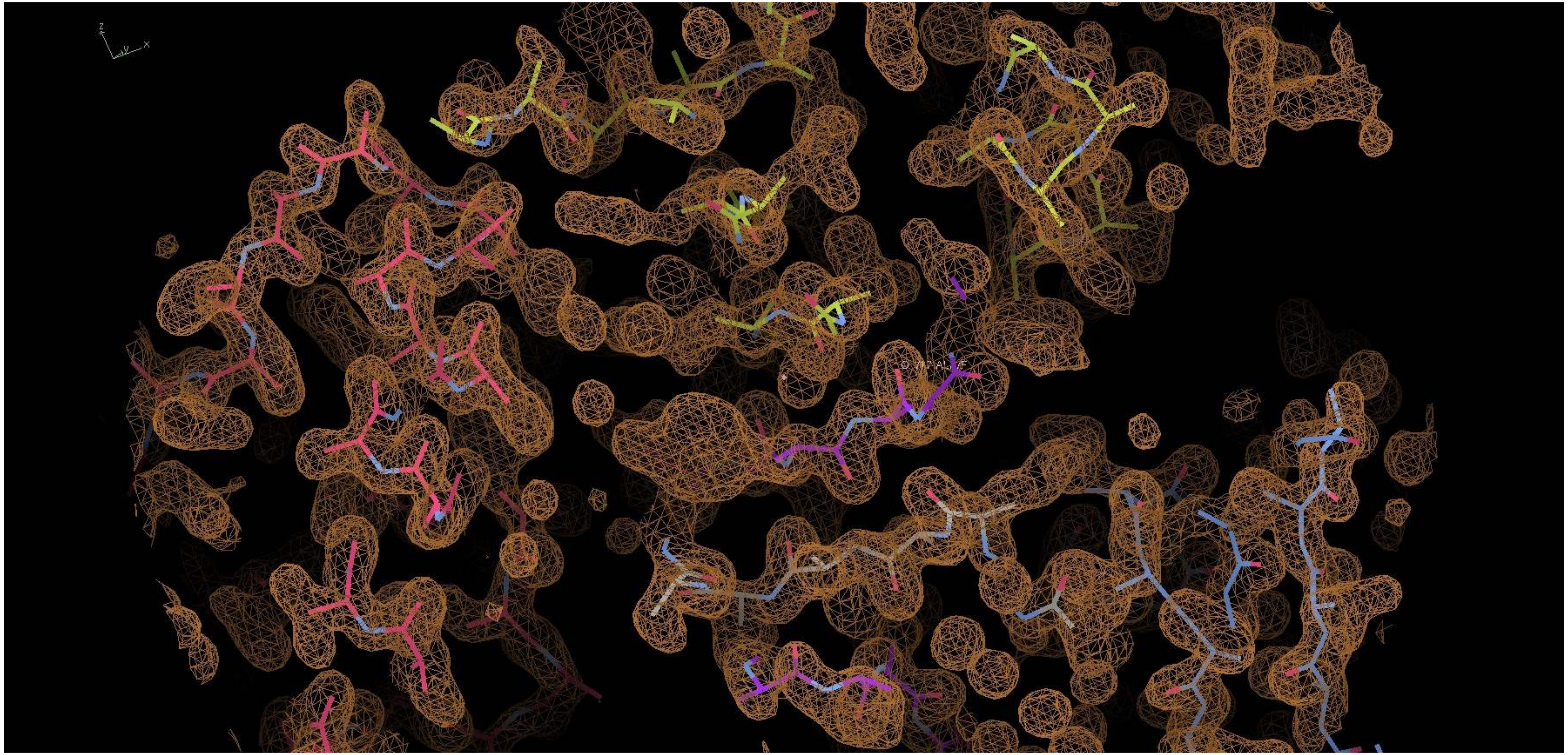
The current best solution is: ensembleIDxx2FR1_8-1.pdb with FINALCC: 42.32 and n. residues traced 378
file is: /home/psalgado/Dropbox/TFP/MR-PS/r20291MRi2/CCP4_JOBS/job_32/8_EXP_LIBRARY/4/0/ensembleIDxx2FR1_8-1.pdb

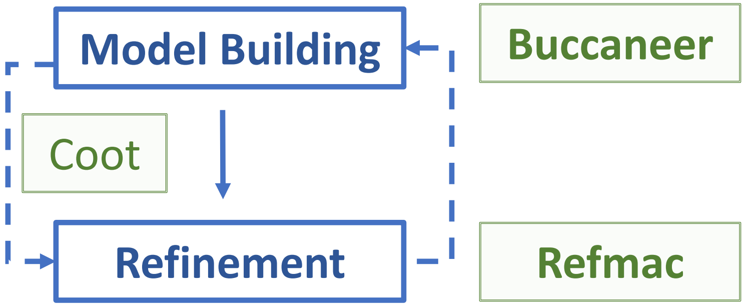
FRF: Pos. in Rank: **85** LLG: **49.30** ZSCORE: **2.58** Top LLG in Cluster **(0, 2)**: **52.60** Top ZSCORE in Cluster **(0, 2)**: **3.66**
REFINEMENT ROTATION AND MODEL
FTF: Pos. in Rank: **1** LLG: **55.90** ZSCORE: **7.97** Top LLG in Cluster **(0, 2)**: **55.90** Top ZSCORE in Cluster **(0, 2)**: **7.97**
PACK: Pos. in Rank: **1** LLG: **55.90** ZSCORE: **7.97** Top LLG in Cluster **(0, 2)**: **55.90** Top ZSCORE in Cluster **(0, 2)**: **7.97**
RNP: Pos. in Rank: **1** LLG: **67.60** ZSCORE: **7.97** TFZ==: **9.00** Top LLG in Cluster **(0, 2)**: **67.60** Top ZSCORE in Cluster **(0, 2)**: **7.97**
INITIAL CC
    After Refinement: Pos. in Rank: **1** INITCC: **5.29** Top INITCC in Cluster **(0, 2)**: **5.29**
EXPANSION
Cycle 4:
    Final CC: **42.32%** N. Residues Traced: **378.00**

It seems you have a good solution!
Here you can find the best solution and map for further refinement.

Arcimboldo

Molecular replacement





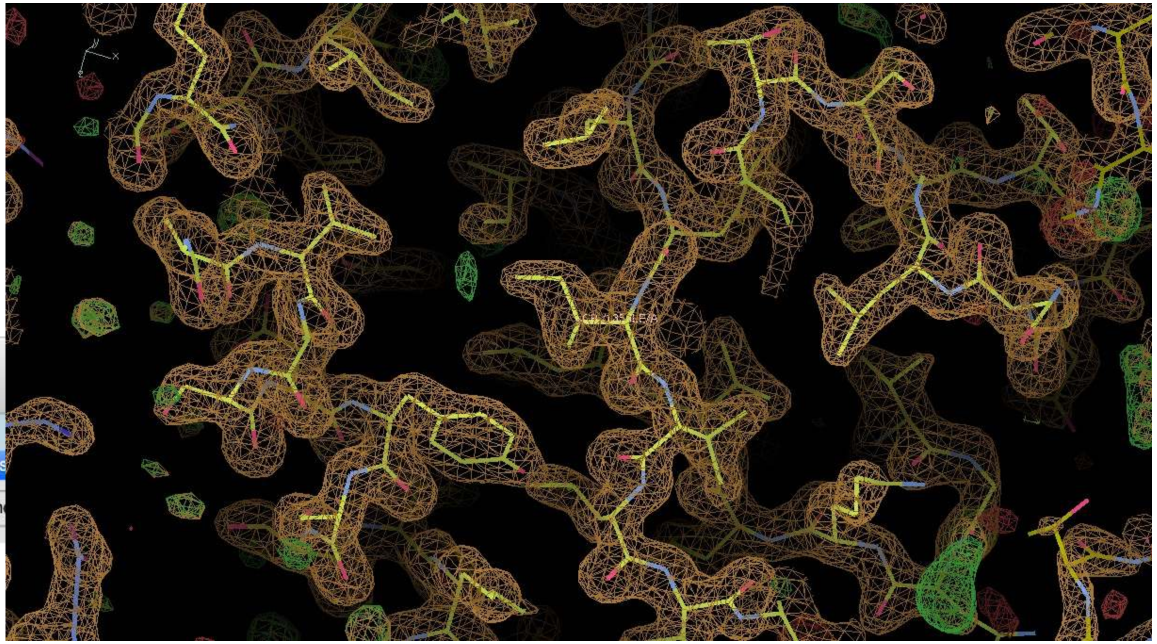
Model Building

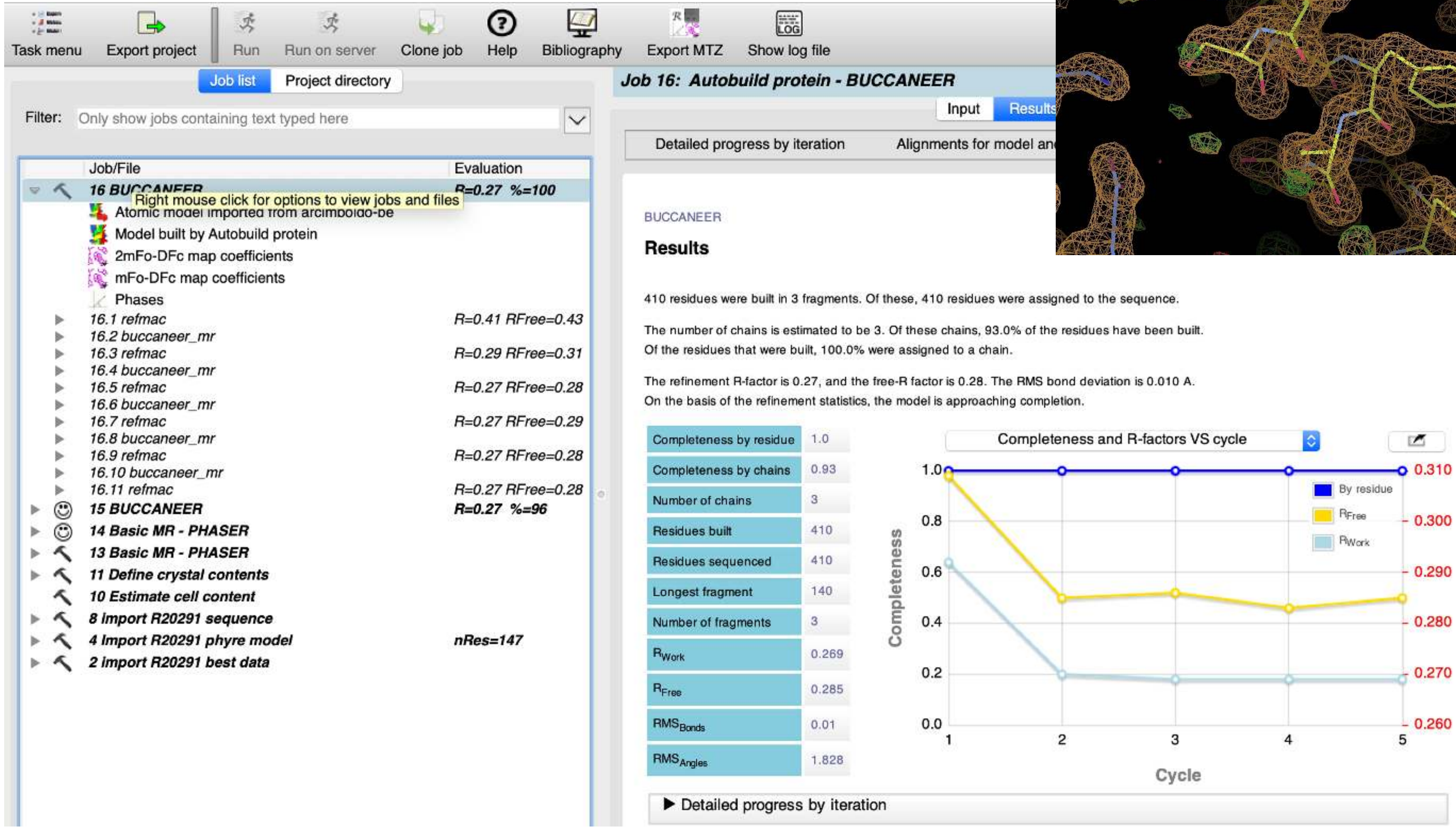
Buccaneer

Coot

Refinement

Refmac





Job/File	Evaluation
16 BUCCANEER	*R=0.27 %=100*
Atomic model imported from arcimboldo-be	
Model built by Autobuild protein	
2mFo-DFc map coefficients	
mFo-DFc map coefficients	
Phases	
16.1 refmac	*R=0.41 RFree=0.43*
16.2 buccaneer_mr	
16.3 refmac	*R=0.29 RFree=0.31*
16.4 buccaneer_mr	
16.5 refmac	*R=0.27 RFree=0.28*
16.6 buccaneer_mr	
16.7 refmac	*R=0.27 RFree=0.29*
16.8 buccaneer_mr	
16.9 refmac	*R=0.27 RFree=0.28*
16.10 buccaneer_mr	
16.11 refmac	*R=0.27 RFree=0.28*
15 BUCCANEER	*R=0.27 %=96*
14 Basic MR - PHASER	
13 Basic MR - PHASER	
11 Define crystal contents	
10 Estimate cell content	
8 import R20291 sequence	
4 Import R20291 phyre model	*nRes=147*
2 import R20291 best data	

Job 16: Autobuild protein - BUCCANEER

BUCCANEER

Results

410 residues were built in 3 fragments. Of these, 410 residues were assigned to the sequence.

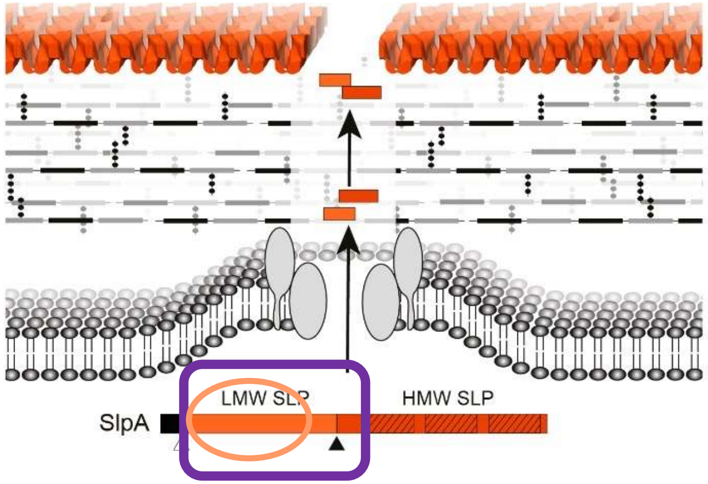
The number of chains is estimated to be 3. Of these chains, 93.0% of the residues have been built.
Of the residues that were built, 100.0% were assigned to a chain.

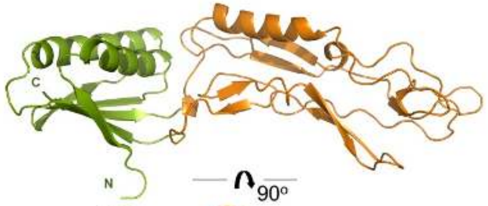
The refinement R-factor is 0.27, and the free-R factor is 0.28. The RMS bond deviation is 0.010 A.
On the basis of the refinement statistics, the model is approaching completion.

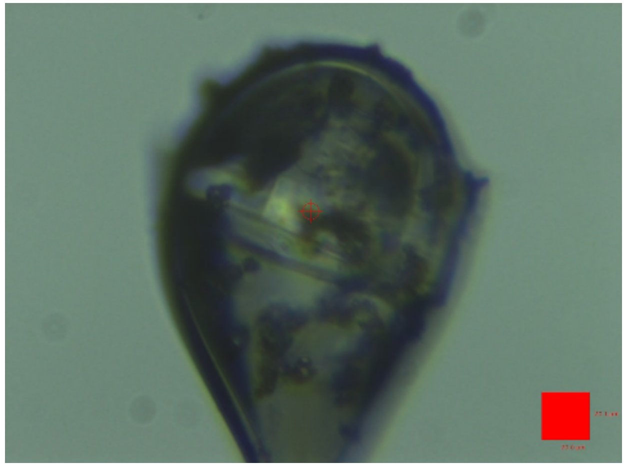
Completeness by residue	1.0
Completeness by chains	0.93
Number of chains	3
Residues built	410
Residues sequenced	410
Longest fragment	140
Number of fragments	3
R_{Work}	0.269
R_{Free}	0.285
RMS_{Bonds}	0.01
RMS_{Angles}	1.828

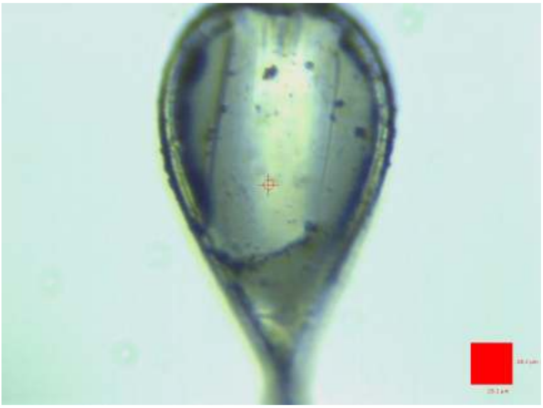
▶ Detailed progress by iteration

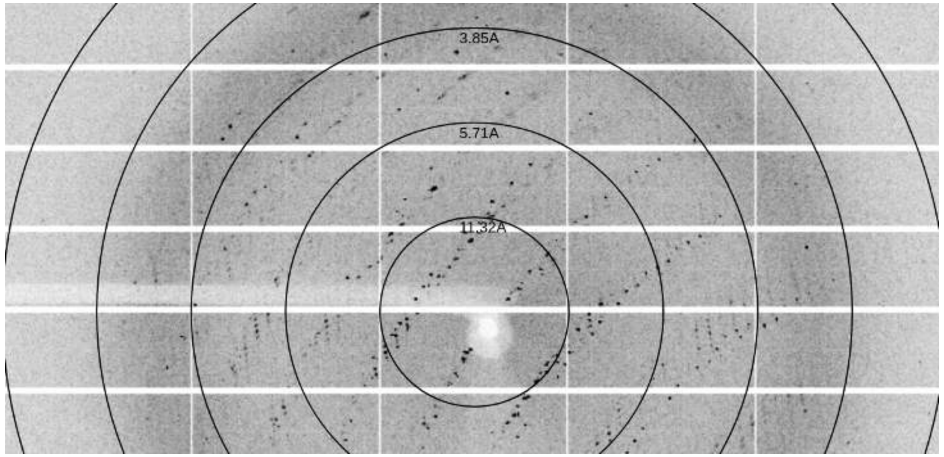
Case study 2: *C. difficile* S-layer



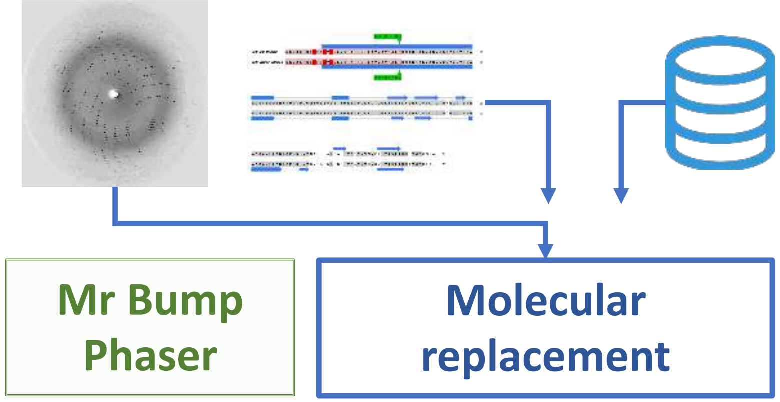






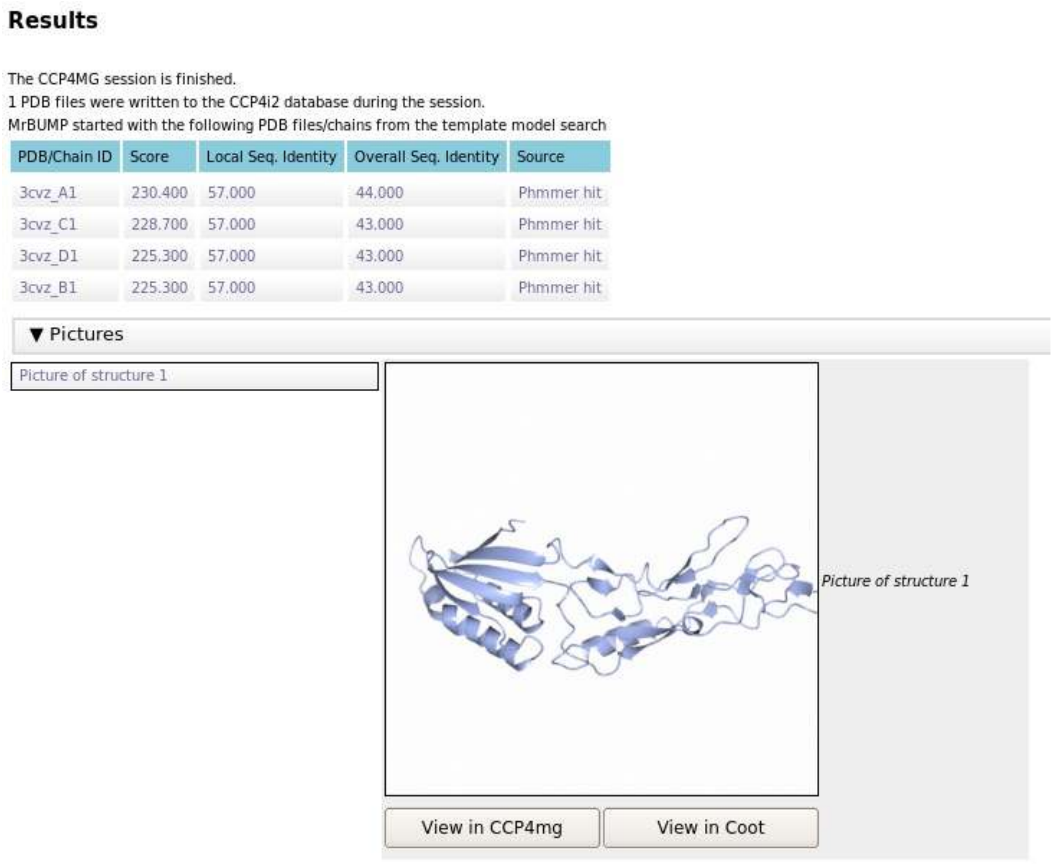


	Overall	High res.
Space group	C2	
Unit cell		
a, b, c (Å)	173.25, 29.58, 144.63	
α=γ, β (°)	90.0, 94.2	
Resolution (Å)	144.24 – 2.57	2.57-2.69
I/σI	10.2	2.2
Half-set correlation CC(1/2)	0.993	0.838
Completeness %	99	98



Mr Bump Phaser

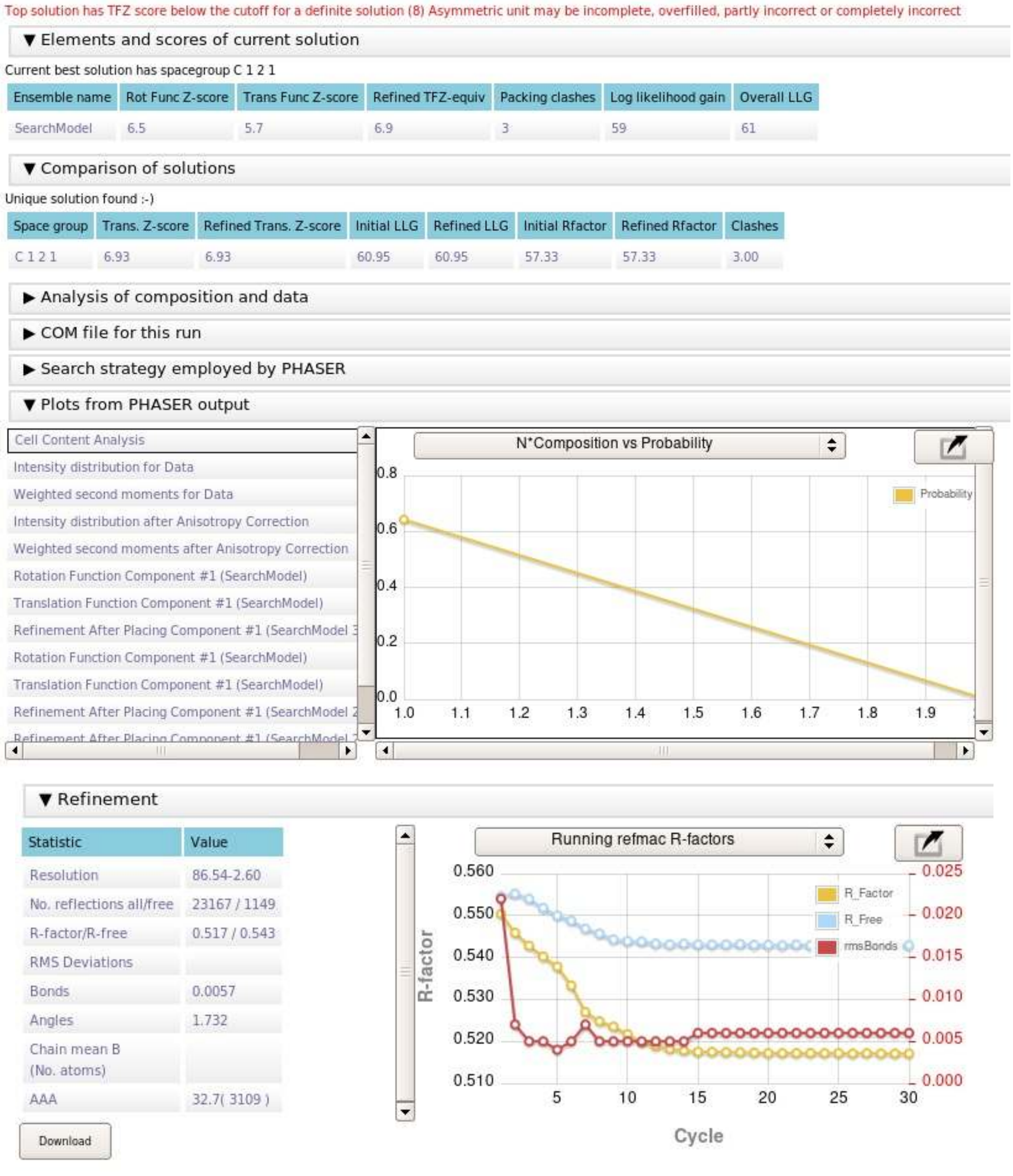
Molecular replacement



Results

The CCP4MG session is finished.
1 PDB files were written to the CCP4i2 database during the session.
MrBUMP started with the following PDB files/chains from the template model search

PDB/Chain ID	Score	Local Seq. Identity	Overall Seq. Identity	Source
3cvz_A1	230.400	57.000	44.000	Phmmer hit
3cvz_C1	228.700	57.000	43.000	Phmmer hit
3cvz_D1	225.300	57.000	43.000	Phmmer hit
3cvz_B1	225.300	57.000	43.000	Phmmer hit



Top solution has TFZ score below the cutoff for a definite solution (8) Asymmetric unit may be incomplete, overfilled, partly incorrect or completely incorrect

▼ Elements and scores of current solution

Current best solution has spacegroup C 1 2 1

Ensemble name	Rot Func Z-score	Trans Func Z-score	Refined TFZ-equiv	Packing clashes	Log likelihood gain	Overall LLG
SearchModel	6.5	5.7	6.9	3	59	61

▼ Comparison of solutions

Unique solution found :-)

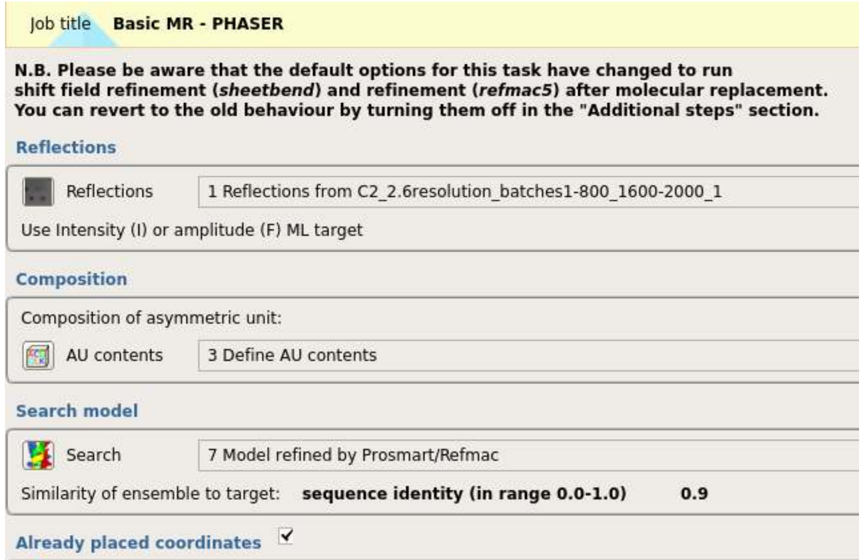
Space group	Trans. Z-score	Refined Trans. Z-score	Initial LLG	Refined LLG	Initial Rfactor	Refined Rfactor	Clashes
C 1 2 1	6.93	6.93	60.95	60.95	57.33	57.33	3.00

▼ Refinement

Statistic	Value
Resolution	86.54-2.60
No. reflections all/free	23167 / 1149
R-factor/R-free	0.517 / 0.543
RMS Deviations	
Bonds	0.0057
Angles	1.732
Chain mean B (No. atoms)	
AAA	32.7(3109)

Molecular replacement

Phaser



Job title **Basic MR - PHASER**

N.B. Please be aware that the default options for this task have changed to run shift field refinement (*sheetbend*) and refinement (*refmac5*) after molecular replacement. You can revert to the old behaviour by turning them off in the "Additional steps" section.

Reflections

Reflections: 1 Reflections from C2_2.6resolution_batches1-800_1600-2000_1

Use Intensity (I) or amplitude (F) ML target

Composition

Composition of asymmetric unit:

AU contents: 3 Define AU contents

Search model

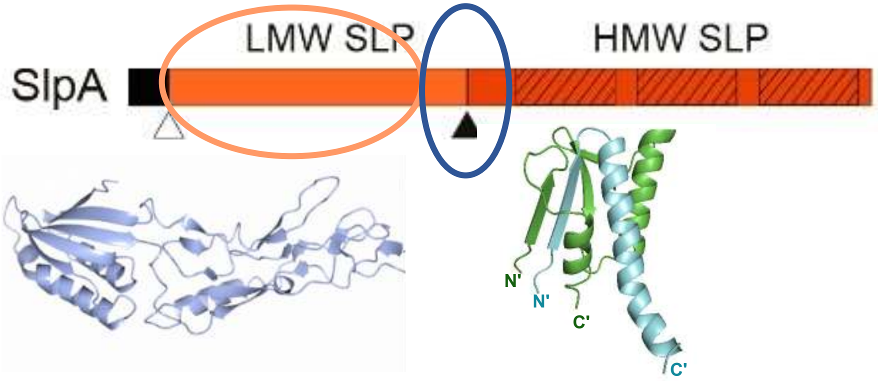
Search: 7 Model refined by Prosmart/Refmac

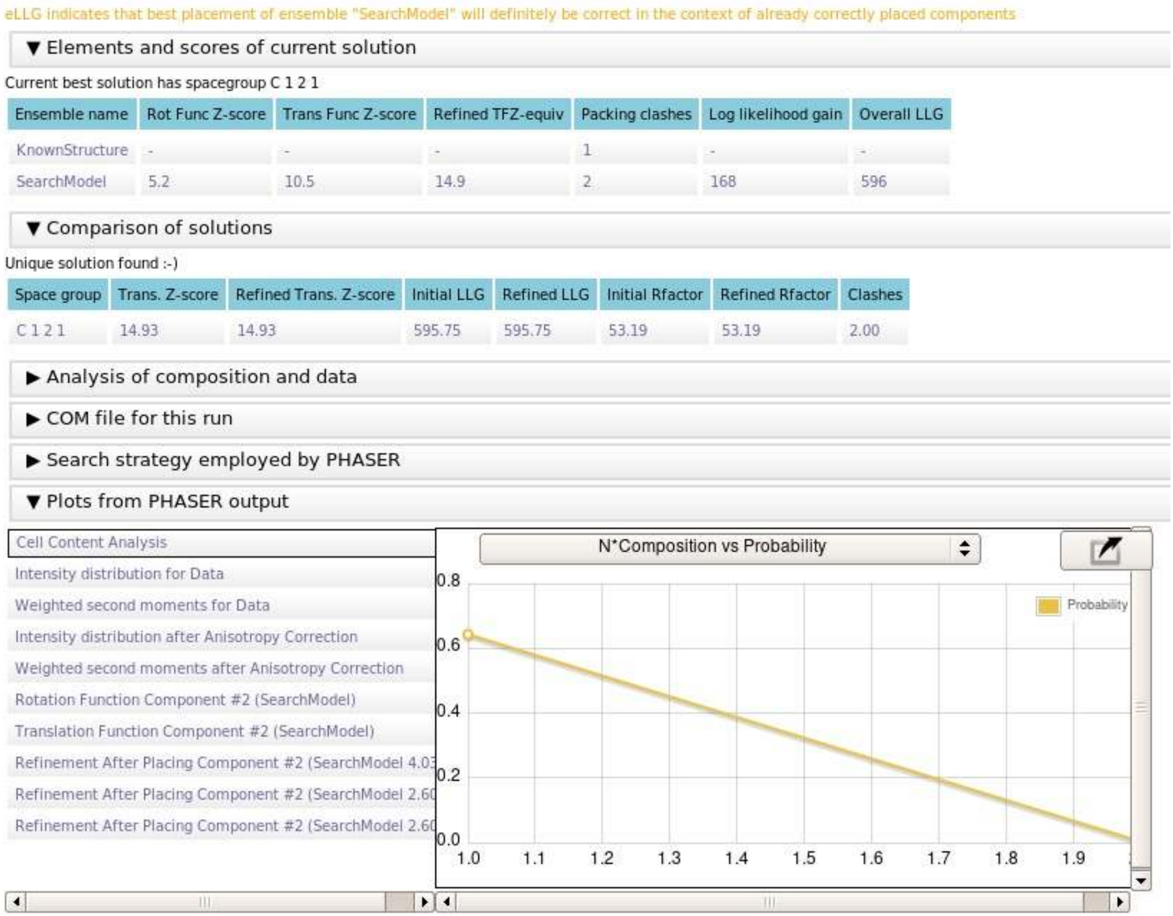
Similarity of ensemble to target: **sequence identity (in range 0.0-1.0)** **0.9**

Already placed coordinates ☑

Fixed: 7 Model refined by Pros

Similarity of fixed ensemble to target: **sequ**





eLLG indicates that best placement of ensemble "SearchModel" will definitely be correct in the context of already correctly placed components

▼ Elements and scores of current solution

Current best solution has spacegroup C 1 2 1

Ensemble name	Rot Func Z-score	Trans Func Z-score	Refined TFZ-equiv	Packing clashes	Log likelihood gain	Overall LLG
KnownStructure	-	-	-	1	-	-
SearchModel	5.2	10.5	14.9	2	168	596

▼ Comparison of solutions

Unique solution found :-)

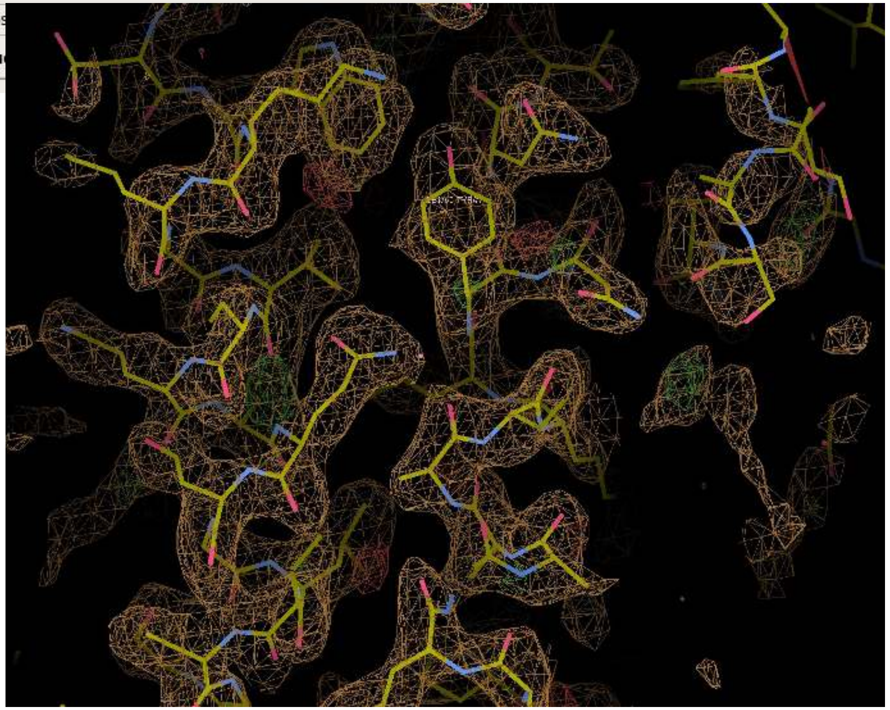
Space group	Trans. Z-score	Refined Trans. Z-score	Initial LLG	Refined LLG	Initial Rfactor	Refined Rfactor	Clashes
C 1 2 1	14.93	14.93	595.75	595.75	53.19	53.19	2.00

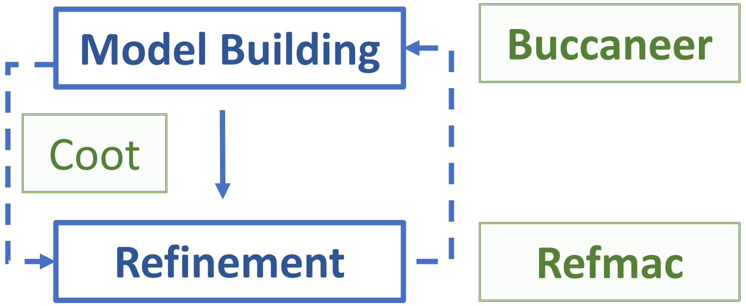
▶ Analysis of composition and data

▶ COM file for this run

▶ Search strategy employed by PHASER

▼ Plots from PHASER output





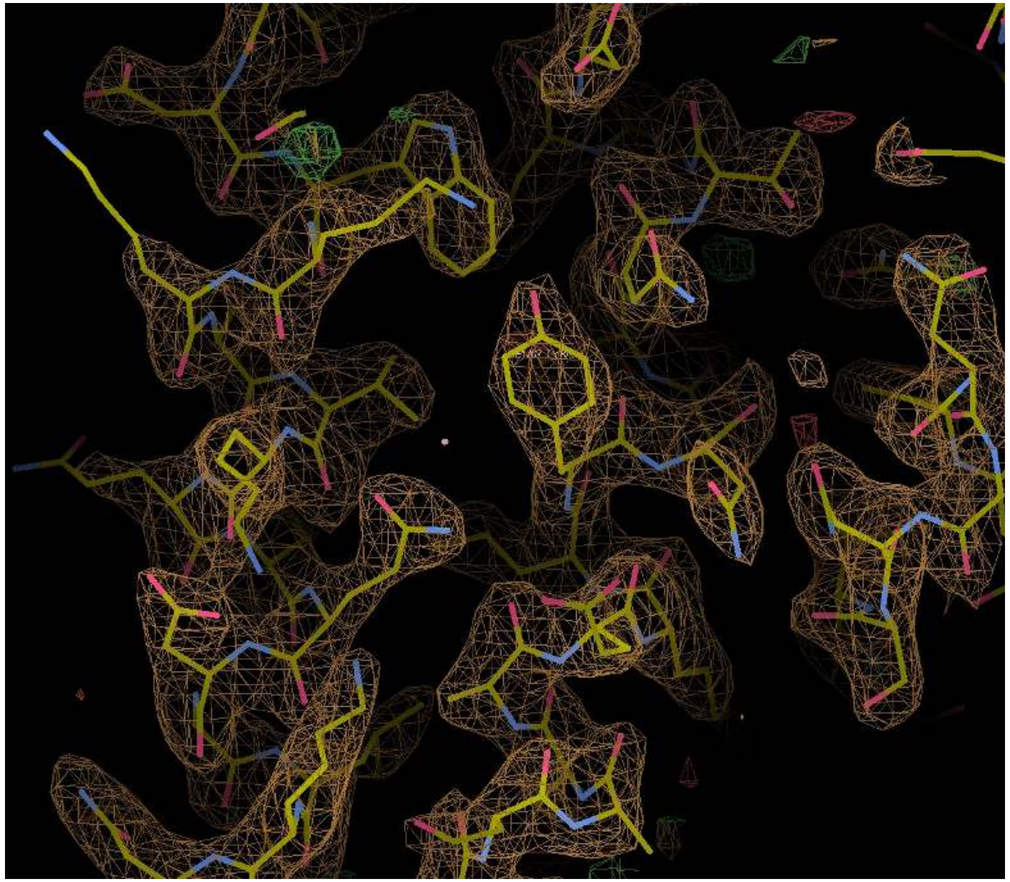
Model Building

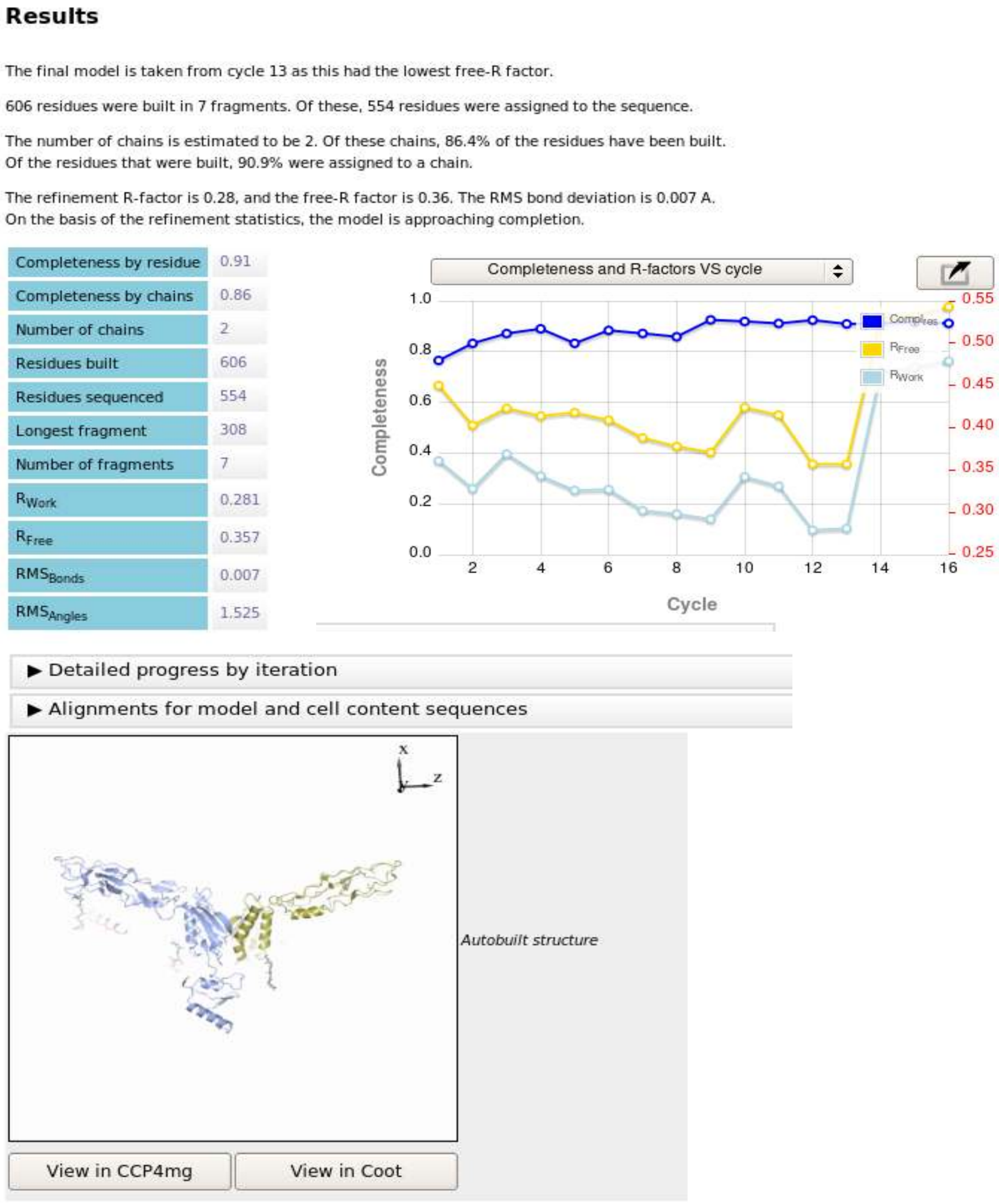
Buccaneer

Coot

Refinement

Refmac





Results

The final model is taken from cycle 13 as this had the lowest free-R factor.

606 residues were built in 7 fragments. Of these, 554 residues were assigned to the sequence.

The number of chains is estimated to be 2. Of these chains, 86.4% of the residues have been built.
Of the residues that were built, 90.9% were assigned to a chain.

The refinement R-factor is 0.28, and the free-R factor is 0.36. The RMS bond deviation is 0.007 A.
On the basis of the refinement statistics, the model is approaching completion.

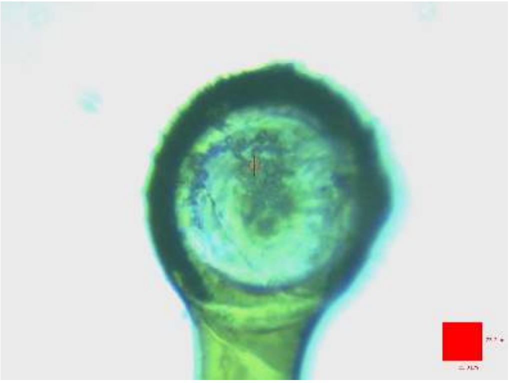
Completeness by residue	0.91
Completeness by chains	0.86
Number of chains	2
Residues built	606
Residues sequenced	554
Longest fragment	308
Number of fragments	7
R_{Work}	0.281
R_{Free}	0.357
RMS_{Bonds}	0.007
RMS_{Angles}	1.525

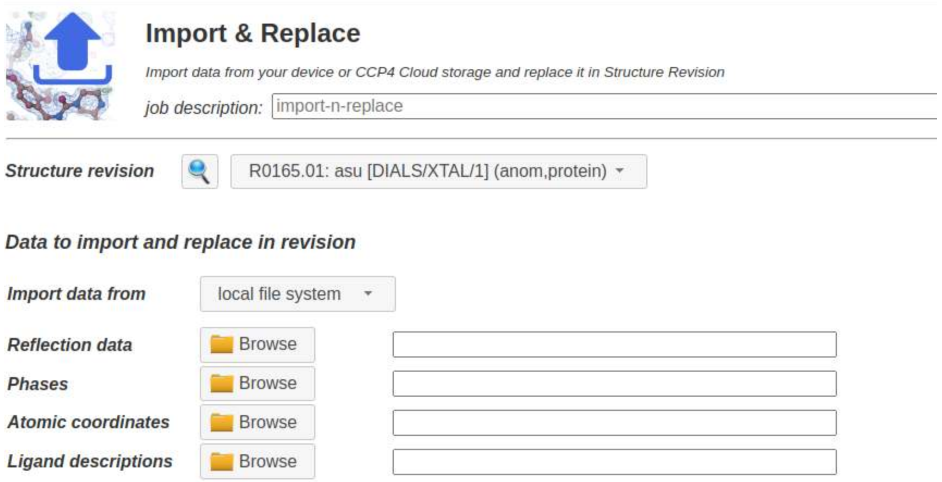
▶ Detailed progress by iteration

▶ Alignments for model and cell content sequences

Autobuilt structure

Case study 2.2 : *C. difficile* S-layer mutant using CCP4cloud





Import & Replace

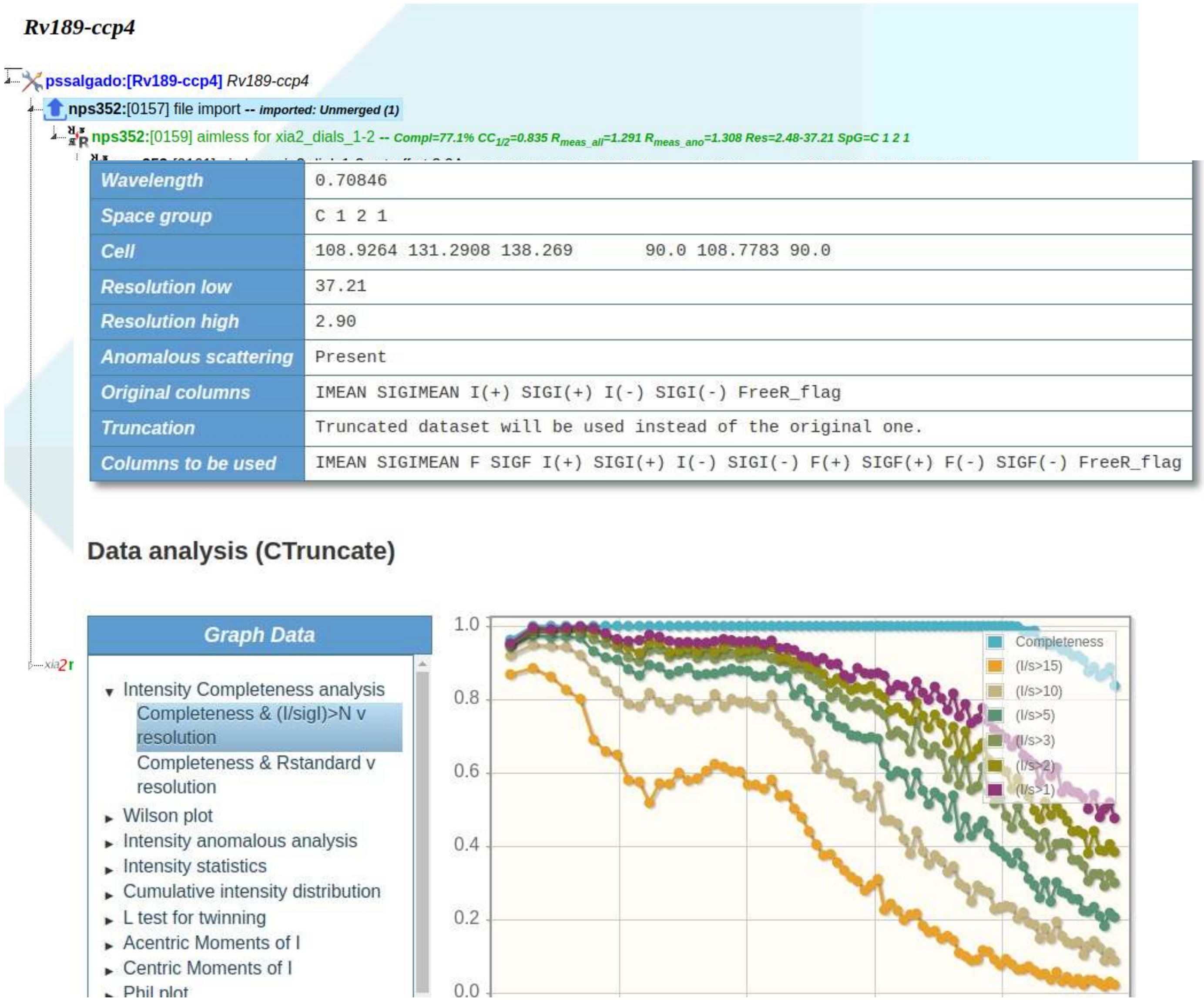
Import data from your device or CCP4 Cloud storage and replace it in Structure Revision

job description: import-n-replace

Structure revision R0165.01: asu [DIALS/XTAL/1] (anom,protein)

Data to import and replace in revision

Import data from local file system



Rv189-ccp4

pssalgado:[Rv189-ccp4] *Rv189-ccp4*

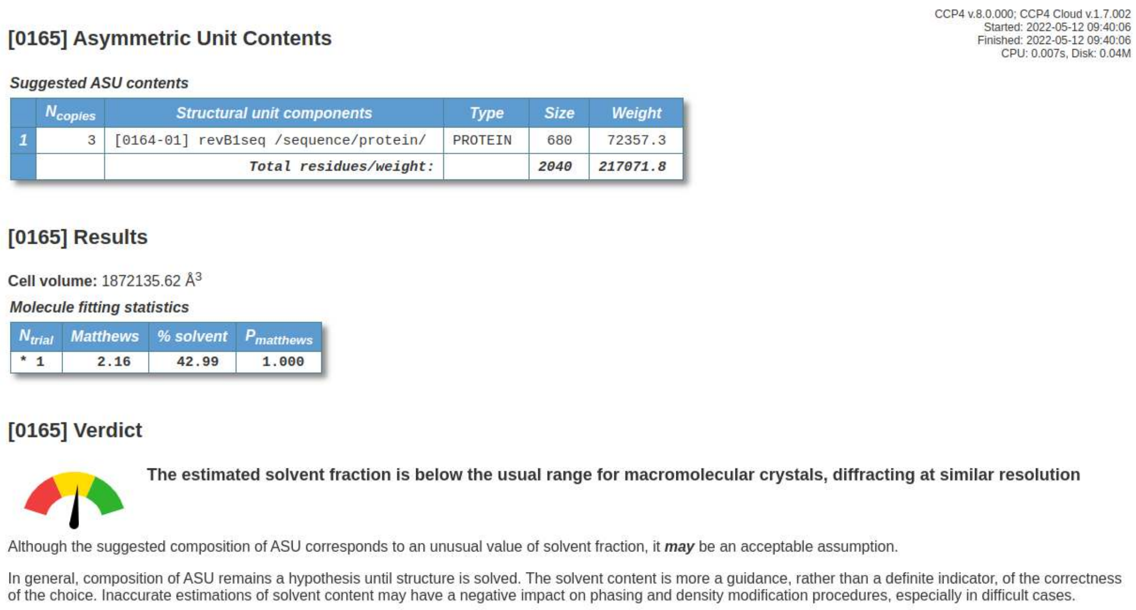
nps352:[0157] file import **-- *imported: Unmerged (1)***

nps352:[0159] aimless for xia2_dials_1-2 **-- *Compl=77.1% $CC_{1/2}$=0.835 R_{meas_all}=1.291 R_{meas_ano}=1.308 Res=2.48-37.21 SpG=C 1 2 1***

Wavelength	0.70846
Space group	C 1 2 1
Cell	108.9264 131.2908 138.269 90.0 108.7783 90.0
Resolution low	37.21
Resolution high	2.90
Anomalous scattering	Present
Original columns	IMEAN SIGIMEAN I(+) SIGI(+) I(-) SIGI(-) FreeR_flag
Truncation	Truncated dataset will be used instead of the original one.
Columns to be used	IMEAN SIGIMEAN F SIGF I(+) SIGI(+) I(-) SIGI(-) F(+) SIGF(+) F(-) SIGF(-) FreeR_flag

Data analysis (CTruncate)

Model preparation



CCP4 v.8.0.000; CCP4 Cloud v.1.7.002
Started: 2022-05-12 09:40:06
Finished: 2022-05-12 09:40:06
CPU: 0.007s, Disk: 0.04M

[0165] Asymmetric Unit Contents

Suggested ASU contents

	N_{copies}	Structural unit components	Type	Size	Weight
1	3	[0164-01] revB1seq /sequence/protein/	PROTEIN	680	72357.3
		Total residues/weight:		*2040*	*217071.8*

[0165] Results

Cell volume: 1872135.62 Å^3

Molecule fitting statistics

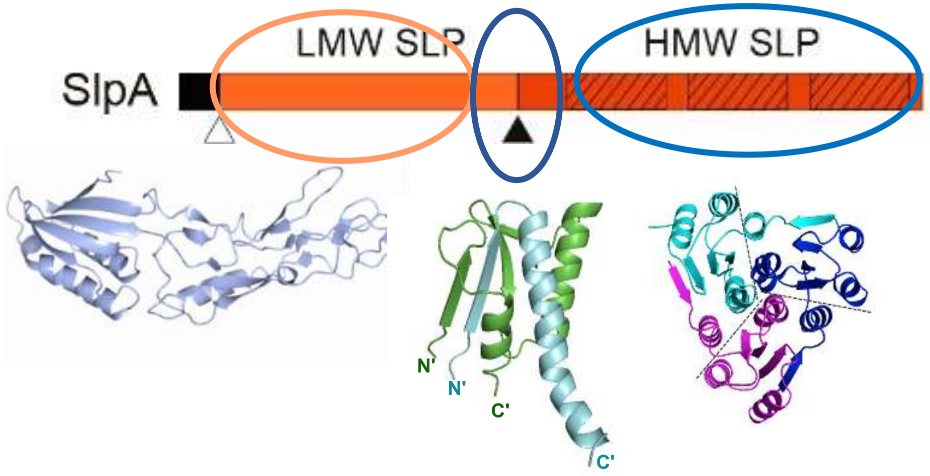
N_{trial}	Matthews	% solvent	$P_{matthews}$
* 1	2.16	42.99	1.000

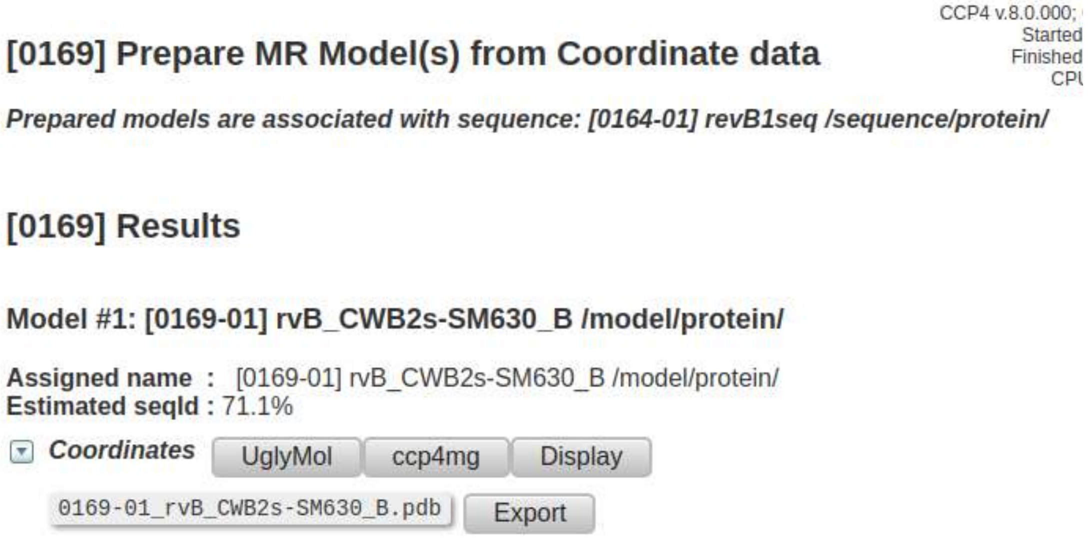
[0165] Verdict

The estimated solvent fraction is below the usual range for macromolecular crystals, diffracting at similar resolution

Although the suggested composition of ASU corresponds to an unusual value of solvent fraction, it ***may*** be an acceptable assumption.

In general, composition of ASU remains a hypothesis until structure is solved. The solvent content is more a guidance, rather than a definite indicator, of the correctness of the choice. Inaccurate estimations of solvent content may have a negative impact on phasing and density modification procedures, especially in difficult cases.





CCP4 v.8.0.000;
Started
Finished
CPU

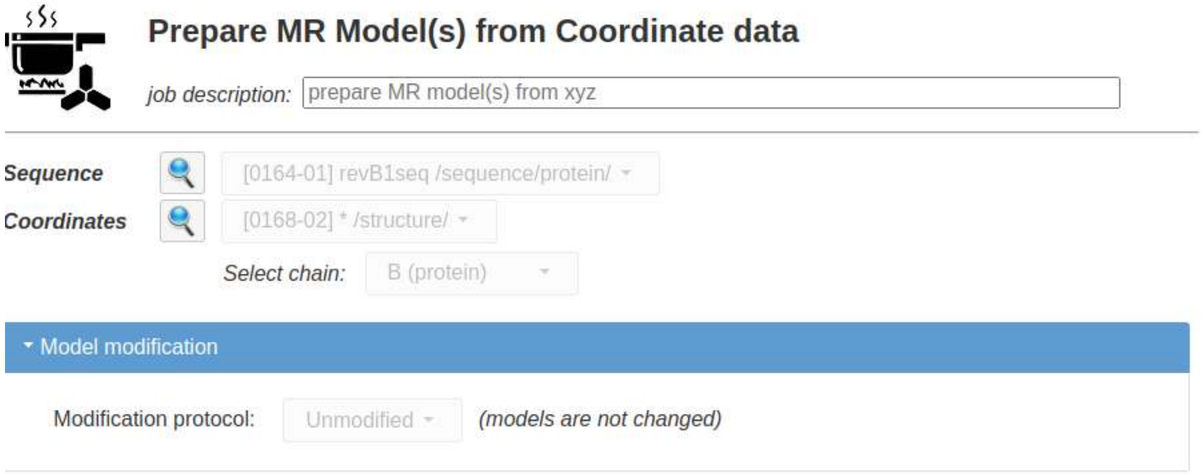
[0169] Prepare MR Model(s) from Coordinate data

Prepared models are associated with sequence: [0164-01] revB1seq /sequence/protein/

[0169] Results

Model #1: [0169-01] rvB_CWB2s-SM630_B /model/protein/

Assigned name : [0169-01] rvB_CWB2s-SM630_B /model/protein/
Estimated seqId : 71.1%



Prepare MR Model(s) from Coordinate data

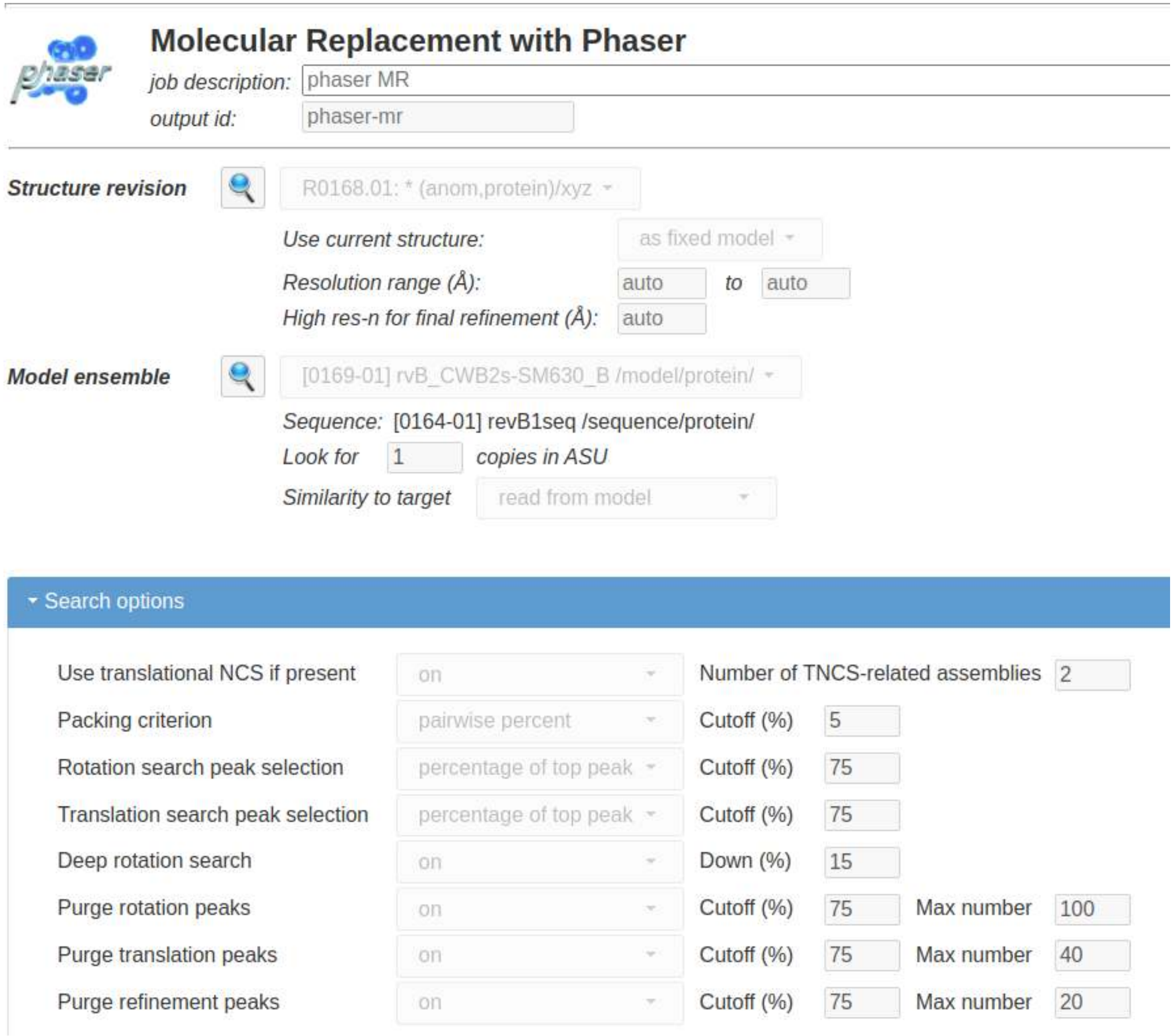
job description: prepare MR model(s) from xyz

Sequence: [0164-01] revB1seq /sequence/protein/
Coordinates: [0168-02] * /structure/
Select chain: B (protein)

Modification protocol: Unmodified *(models are not changed)*

Molecular replacement

Phaser



Molecular Replacement with Phaser

job description: phaser MR

output id: phaser-mr

Structure revision R0168.01: * (anom,protein)/xyz

Use current structure: as fixed model

Resolution range (Å): auto *to* auto

High res-n for final refinement (Å): auto

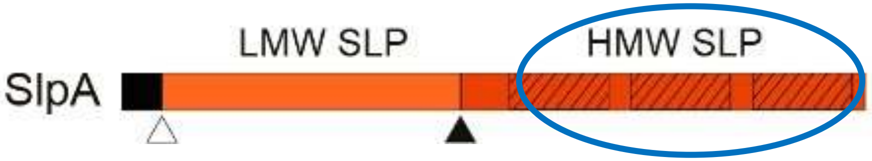
Model ensemble [0169-01] rvB_CWB2s-SM630_B /model/protein/

Sequence: [0164-01] revB1seq /sequence/protein/

Look for 1 *copies in ASU*

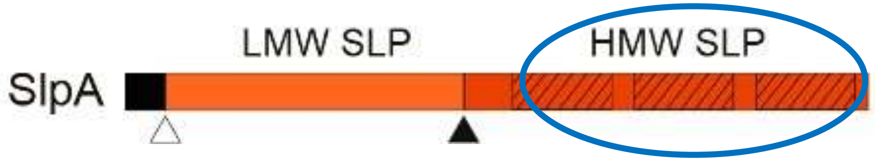
Similarity to target read from model

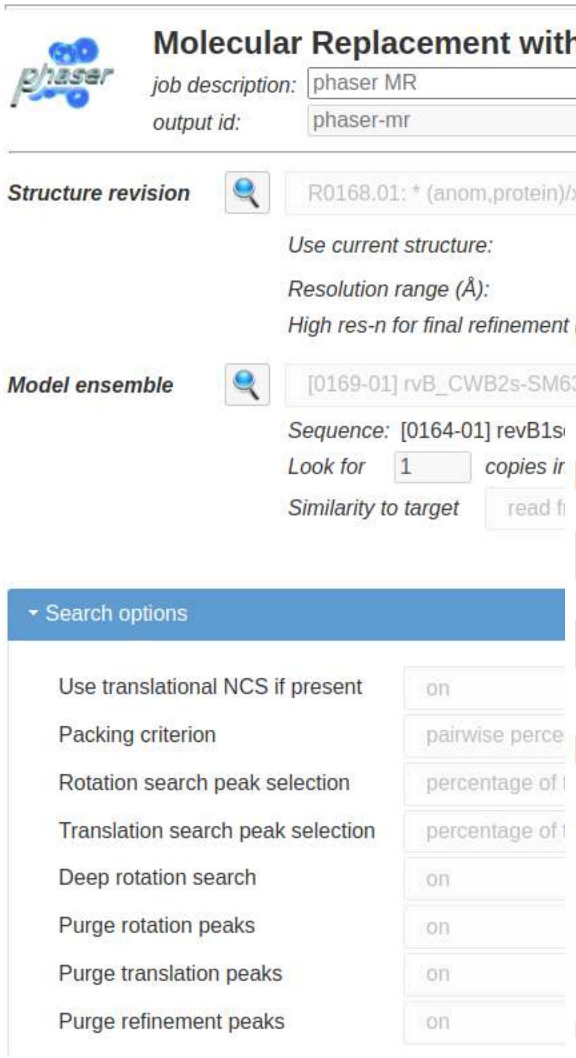
Use translational NCS if present	on	Number of TNCS-related assemblies	2		
Packing criterion	pairwise percent	Cutoff (%)	5		
Rotation search peak selection	percentage of top peak	Cutoff (%)	75		
Translation search peak selection	percentage of top peak	Cutoff (%)	75		
Deep rotation search	on	Down (%)	15		
Purge rotation peaks	on	Cutoff (%)	75	Max number	100
Purge translation peaks	on	Cutoff (%)	75	Max number	40
Purge refinement peaks	on	Cutoff (%)	75	Max number	20



Molecular replacement

Phaser





Molecular Replacement with Phaser

job description: phaser MR

output id: phaser-mr

Structure revision R0168.01: * (anom,protein)/xyz

Use current structure: as fixed model

Resolution range (Å): auto *to* auto

High res-n for final refinement (Å): auto

Model ensemble [0169-01] rvB_CWB2s-SM630_B /model/protein/

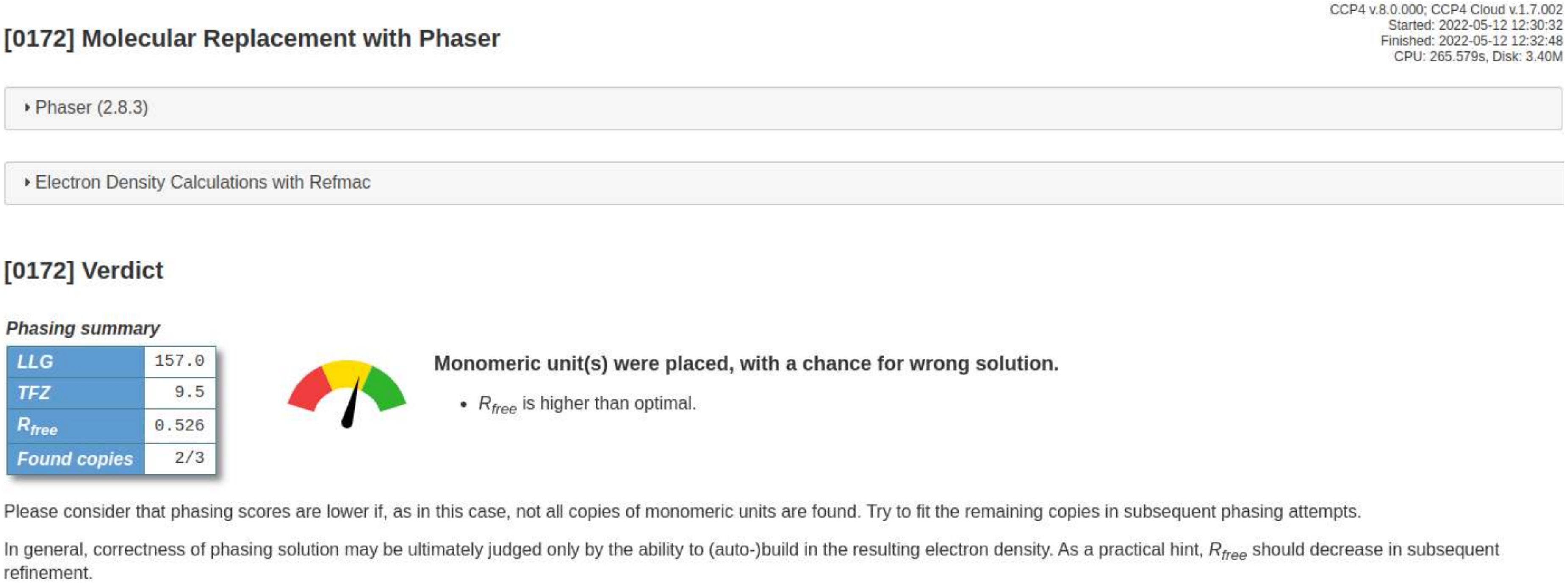
Sequence: [0164-01] revB1s

Look for 1 *copies ir*

Similarity to target read fi

Search options

Use translational NCS if present	on
Packing criterion	pairwise perce
Rotation search peak selection	percentage of t
Translation search peak selection	percentage of t
Deep rotation search	on
Purge rotation peaks	on
Purge translation peaks	on
Purge refinement peaks	on



CCP4 v.8.0.000; CCP4 Cloud v.1.7.002
Started: 2022-05-12 12:30:32
Finished: 2022-05-12 12:32:48
CPU: 265.579s, Disk: 3.40M

[0172] Molecular Replacement with Phaser

Phaser (2.8.3)

Electron Density Calculations with Refmac

[0172] Verdict

Phasing summary

LLG	157.0
TFZ	9.5
R_{free}	0.526
Found copies	2/3

Monomeric unit(s) were placed, with a chance for wrong solution.

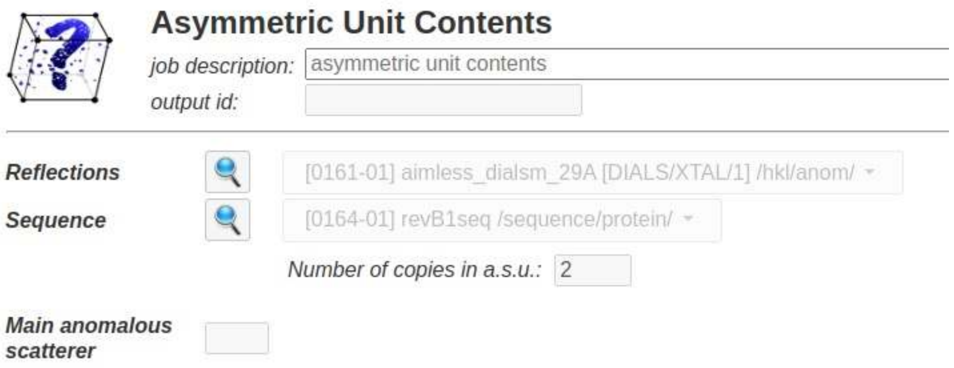
- R_{free} is higher than optimal.

Please consider that phasing scores are lower if, as in this case, not all copies of monomeric units are found. Try to fit the remaining copies in subsequent phasing attempts.

In general, correctness of phasing solution may be ultimately judged only by the ability to (auto-)build in the resulting electron density. As a practical hint, R_{free} should decrease in subsequent refinement.

Molecular replacement

Phaser



Asymmetric Unit Contents

job description: asymmetric unit contents

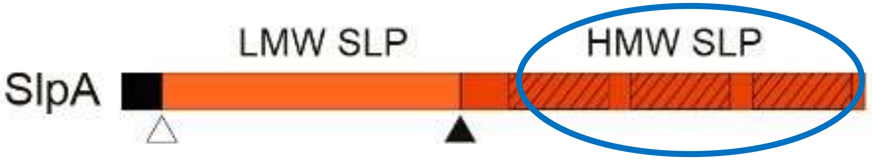
output id:

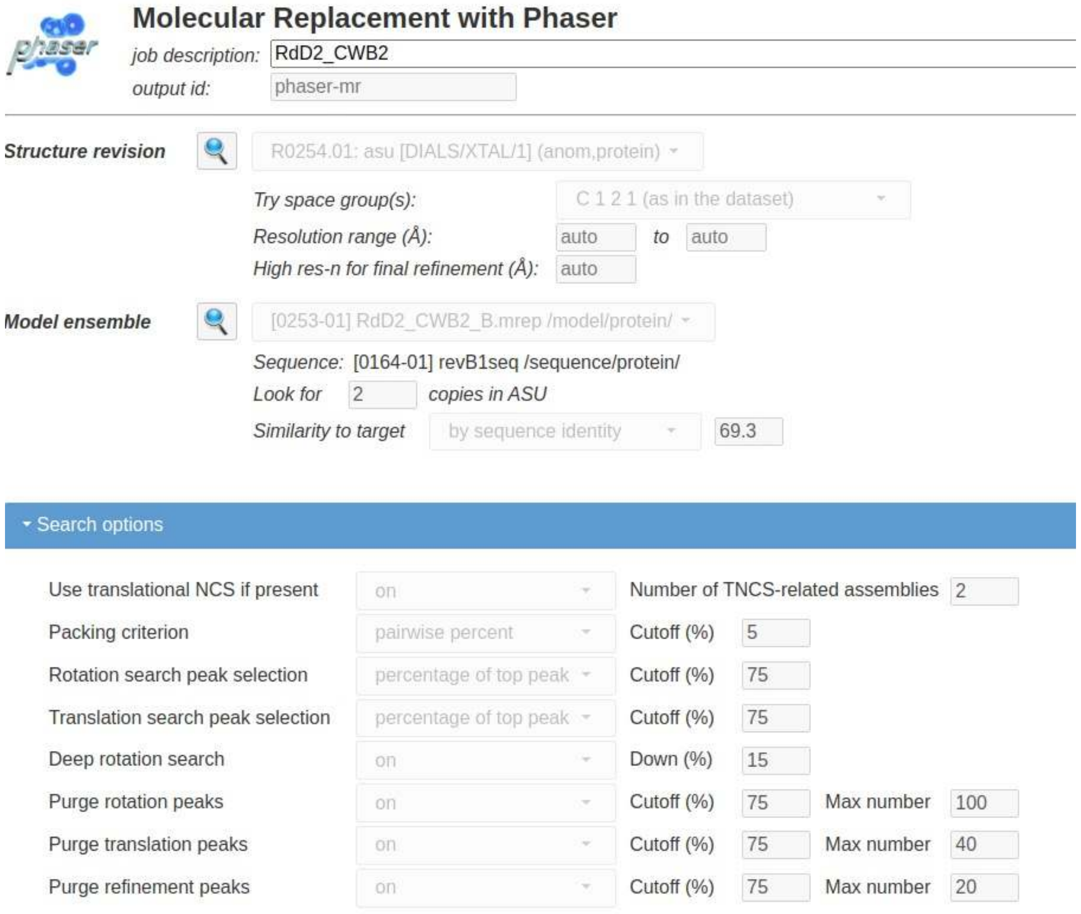
Reflections: [0161-01] aimless_dialsm_29A [DIALS/XTAL/1] /hkl/anom/

Sequence: [0164-01] revB1seq /sequence/protein/

Number of copies in a.s.u.: 2

Main anomalous scatterer:





Molecular Replacement with Phaser

job description: RdD2_CWB2

output id: phaser-mr

Structure revision: R0254.01: asu [DIALS/XTAL/1] (anom,protein)

Try space group(s): C 1 2 1 (as in the dataset)

Resolution range (Å): auto to auto

High res-n for final refinement (Å): auto

Model ensemble: [0253-01] RdD2_CWB2_B.mrep /model/protein/

Sequence: [0164-01] revB1seq /sequence/protein/

Look for 2 copies in ASU

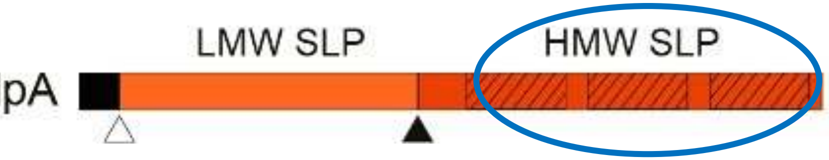
Similarity to target: by sequence identity 69.3

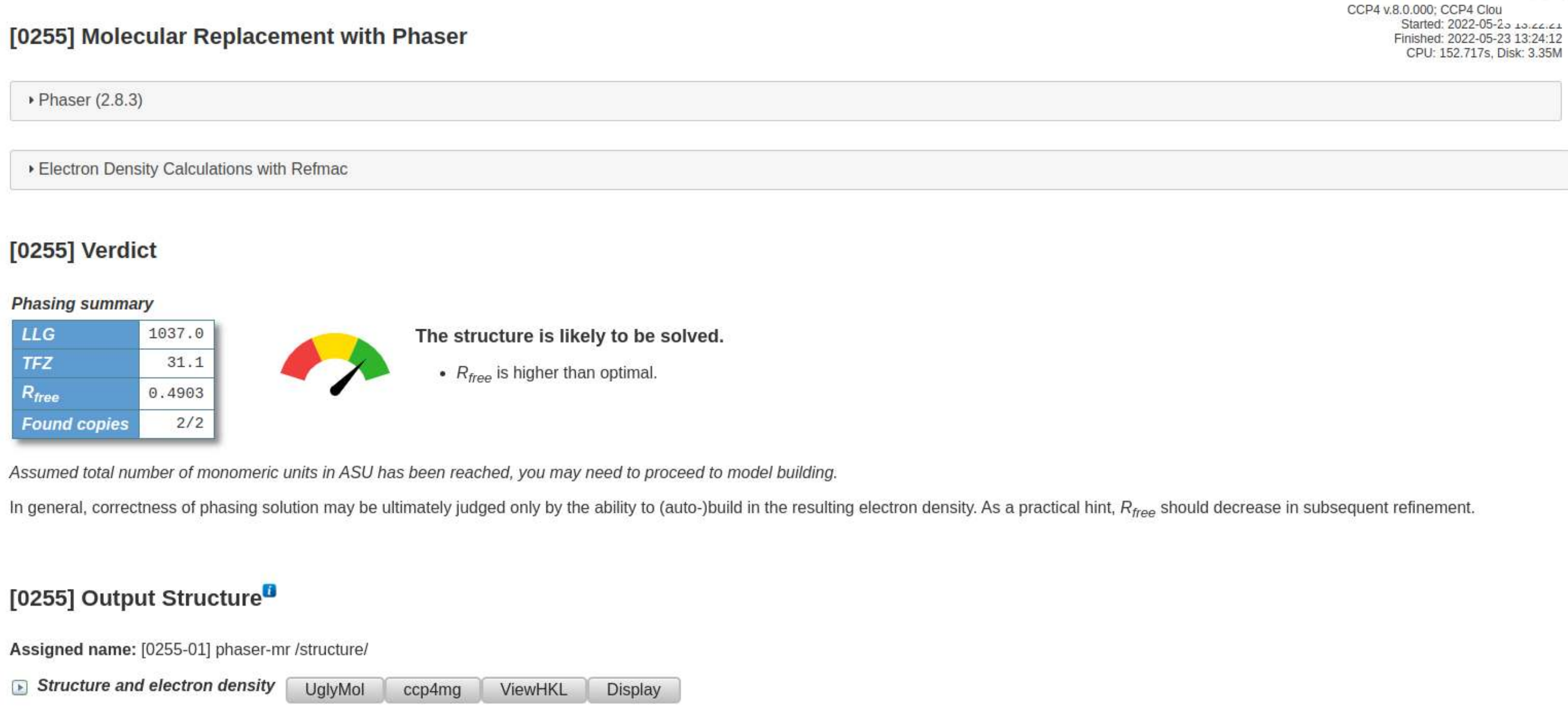
Search options

Use translational NCS if present	on	Number of TNCS-related assemblies	2		
Packing criterion	pairwise percent	Cutoff (%)	5		
Rotation search peak selection	percentage of top peak	Cutoff (%)	75		
Translation search peak selection	percentage of top peak	Cutoff (%)	75		
Deep rotation search	on	Down (%)	15		
Purge rotation peaks	on	Cutoff (%)	75	Max number	100
Purge translation peaks	on	Cutoff (%)	75	Max number	40
Purge refinement peaks	on	Cutoff (%)	75	Max number	20

Molecular replacement

Phaser




CCP4 v.8.0.000; CCP4 Clou
Started: 2022-05-23 13:22:21
Finished: 2022-05-23 13:24:12
CPU: 152.717s, Disk: 3.35M

[0255] Molecular Replacement with Phaser

‣ Phaser (2.8.3)

‣ Electron Density Calculations with Refmac

[0255] Verdict

Phasing summary

LLG	1037.0
TFZ	31.1
R_{free}	0.4903
Found copies	2/2

The structure is likely to be solved.

- R_{free} is higher than optimal.

Assumed total number of monomeric units in ASU has been reached, you may need to proceed to model building.

In general, correctness of phasing solution may be ultimately judged only by the ability to (auto-)build in the resulting electron density. As a practical hint, R_{free} should decrease in subsequent refinement.

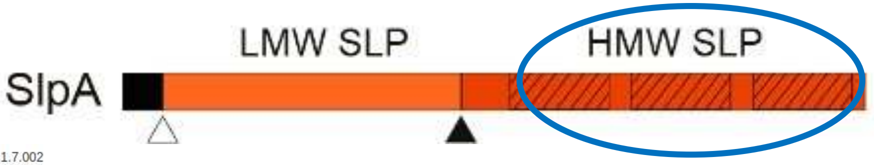
[0255] Output Structure

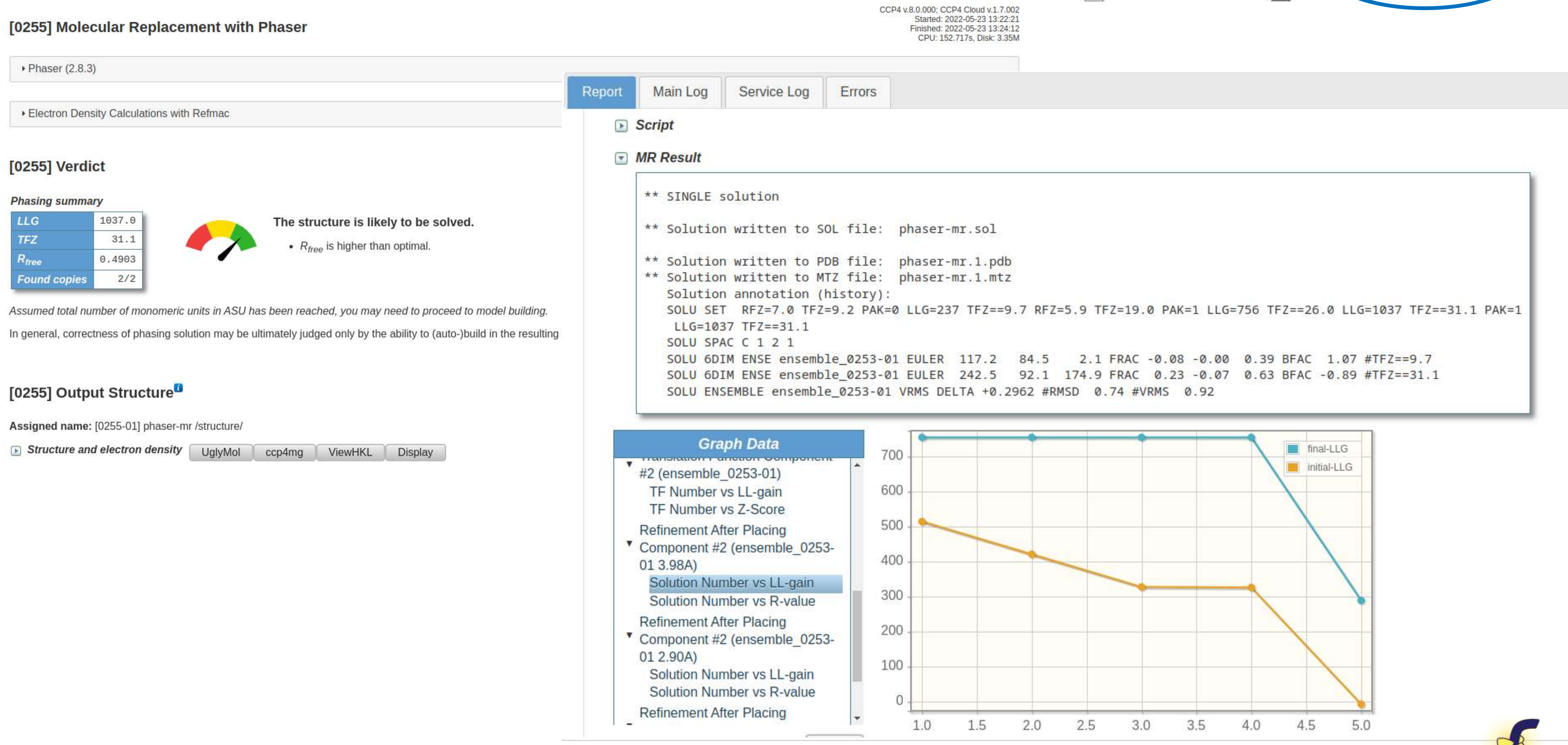
Assigned name: [0255-01] phaser-mr /structure/

Structure and electron density UglyMol ccp4mg ViewHKL Display

Molecular replacement

Phaser





[0255] Molecular Replacement with Phaser

CCP4 v.8.0.000; CCP4 Cloud v.1.7.002
Started: 2022-05-23 13:22:21
Finished: 2022-05-23 13:24:12
CPU: 152.717s, Disk: 3.35M

- Phaser (2.8.3)
- Electron Density Calculations with Refmac

[0255] Verdict

Phasing summary

LLG	1037.0
TFZ	31.1
R_{free}	0.4903
Found copies	2/2

The structure is likely to be solved.

- R_{free} is higher than optimal.

Assumed total number of monomeric units in ASU has been reached, you may need to proceed to model building.

In general, correctness of phasing solution may be ultimately judged only by the ability to (auto-)build in the resulting

[0255] Output Structure

Assigned name: [0255-01] phaser-mr /structure/

Structure and electron density

Script

MR Result

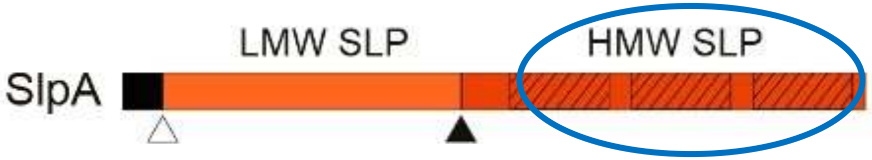
```
** SINGLE solution

** Solution written to SOL file:  phaser-mr.sol

** Solution written to PDB file:  phaser-mr.1.pdb
** Solution written to MTZ file:  phaser-mr.1.mtz
   Solution annotation (history):
   SOLU SET  RFZ=7.0 TFZ=9.2 PAK=0 LLG=237 TFZ==9.7 RFZ=5.9 TFZ=19.0 PAK=1 LLG=756 TFZ==26.0 LLG=1037 TFZ==31.1 PAK=1
    LLG=1037 TFZ==31.1
   SOLU SPAC C 1 2 1
   SOLU 6DIM ENSE ensemble_0253-01 EULER  117.2   84.5    2.1 FRAC -0.08 -0.00  0.39 BFAC  1.07 #TFZ==9.7
   SOLU 6DIM ENSE ensemble_0253-01 EULER  242.5   92.1  174.9 FRAC  0.23 -0.07  0.63 BFAC -0.89 #TFZ==31.1
   SOLU ENSEMBLE ensemble_0253-01 VRMS DELTA +0.2962 #RMSD  0.74 #VRMS  0.92
```

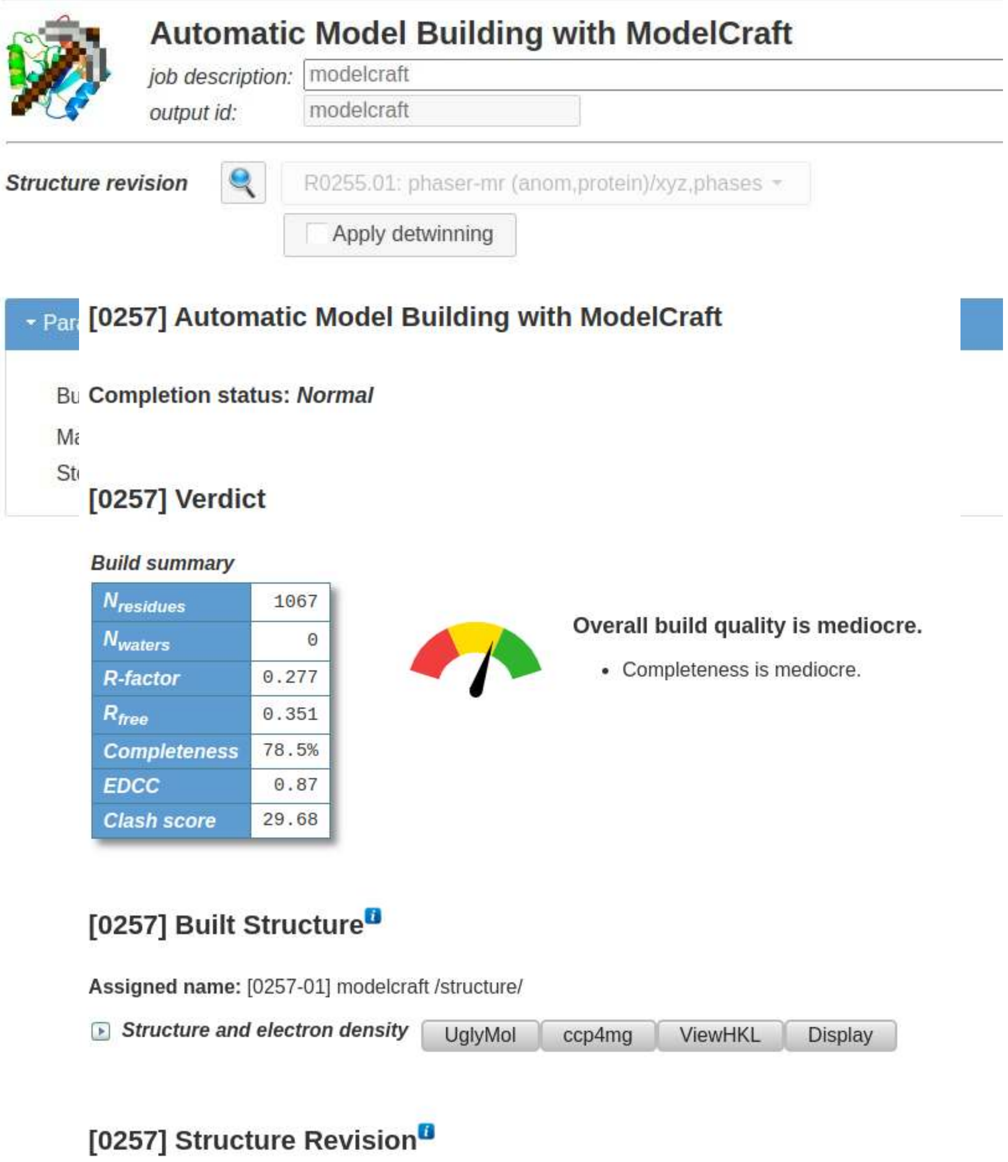
Graph Data

- #2 (ensemble_0253-01)
  - TF Number vs LL-gain
  - TF Number vs Z-Score
- Refinement After Placing Component #2 (ensemble_0253-01 3.98A)
  - Solution Number vs LL-gain
  - Solution Number vs R-value
- Refinement After Placing Component #2 (ensemble_0253-01 2.90A)
  - Solution Number vs LL-gain
  - Solution Number vs R-value
- Refinement After Placing

Model building

Modelcraft





Automatic Model Building with ModelCraft

job description: modelcraft

output id: modelcraft

Structure revision R0255.01: phaser-mr (anom,protein)/xyz,phases

☐ Apply detwinning

[0257] Automatic Model Building with ModelCraft

Completion status: ***Normal***

[0257] Verdict

Build summary

$N_{residues}$	1067
N_{waters}	0
R-factor	0.277
R_{free}	0.351
Completeness	78.5%
EDCC	0.87
Clash score	29.68

Overall build quality is mediocre.

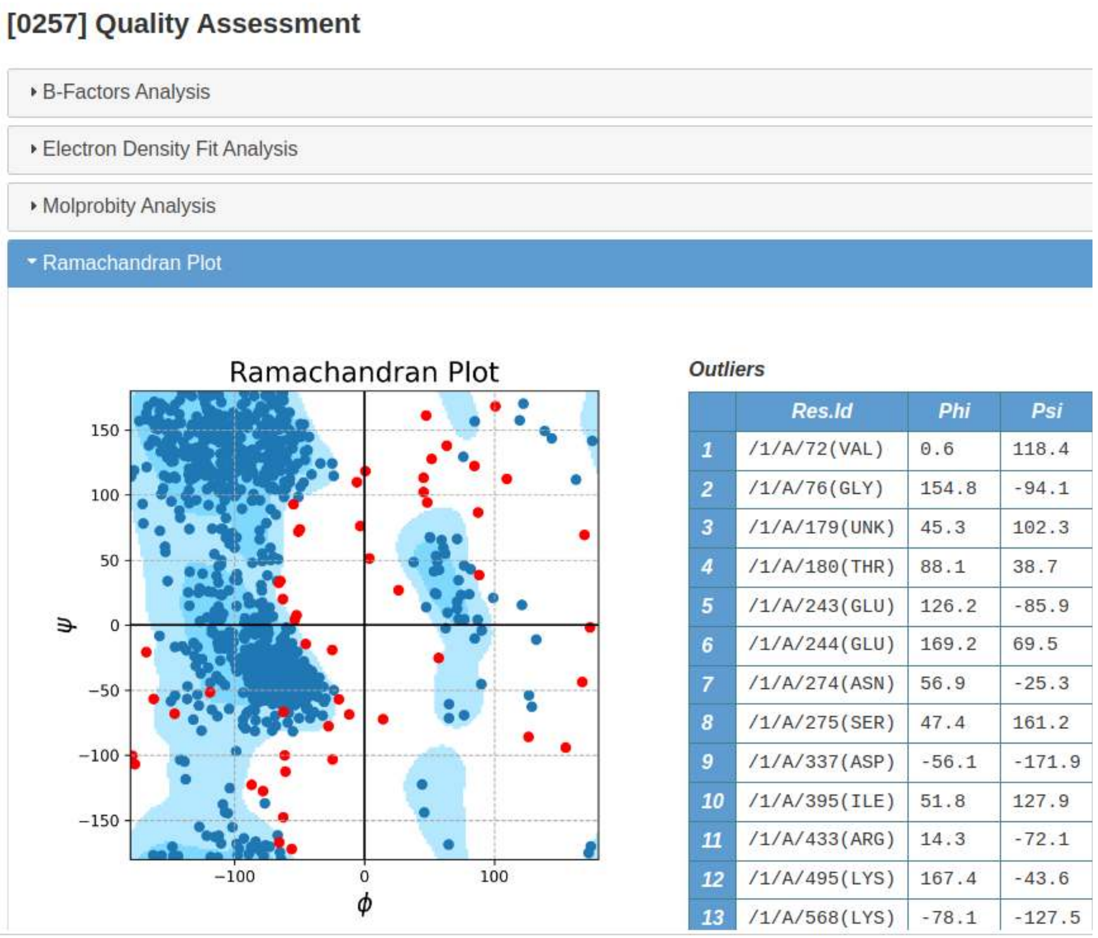
- Completeness is mediocre.

[0257] Built Structure

Assigned name: [0257-01] modelcraft /structure/

Structure and electron density

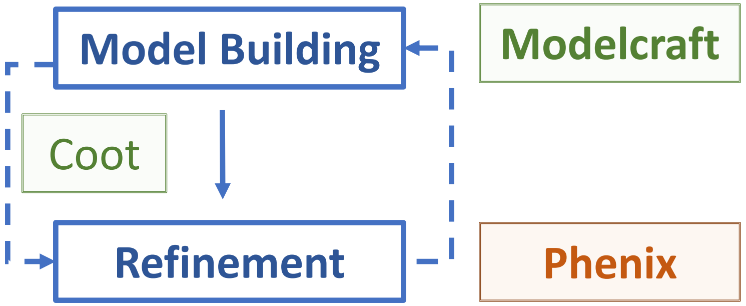
[0257] Structure Revision



[0257] Quality Assessment

Outliers

	Res.Id	Phi	Psi
1	/1/A/72(VAL)	0.6	118.4
2	/1/A/76(GLY)	154.8	-94.1
3	/1/A/179(UNK)	45.3	102.3
4	/1/A/180(THR)	88.1	38.7
5	/1/A/243(GLU)	126.2	-85.9
6	/1/A/244(GLU)	169.2	69.5
7	/1/A/274(ASN)	56.9	-25.3
8	/1/A/275(SER)	47.4	161.2
9	/1/A/337(ASP)	-56.1	-171.9
10	/1/A/395(ILE)	51.8	127.9
11	/1/A/433(ARG)	14.3	-72.1
12	/1/A/495(LYS)	167.4	-43.6
13	/1/A/568(LYS)	-78.1	-127.5



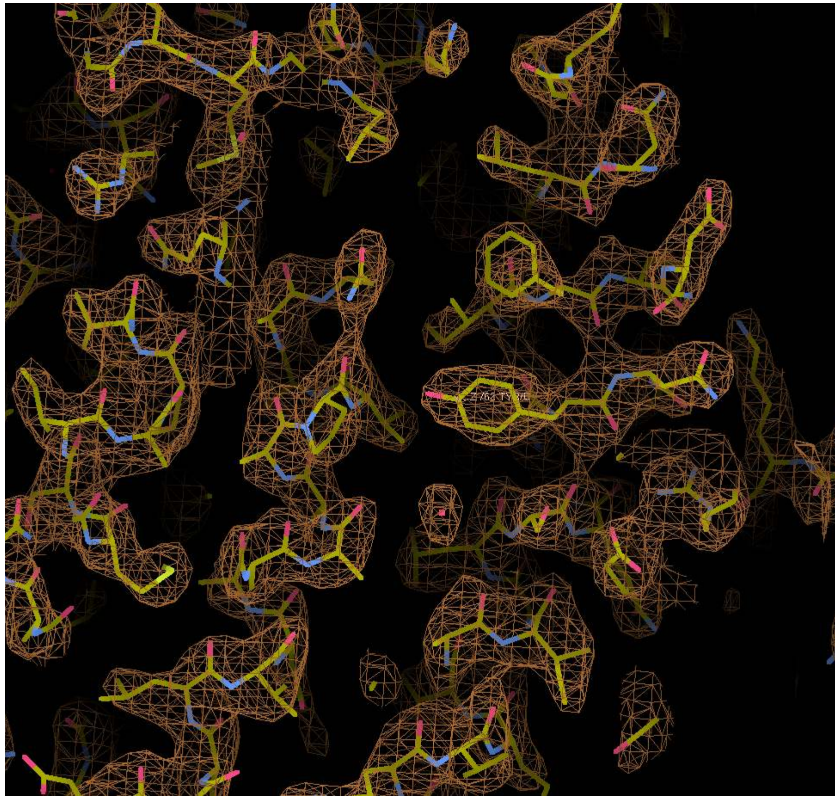
Model Building

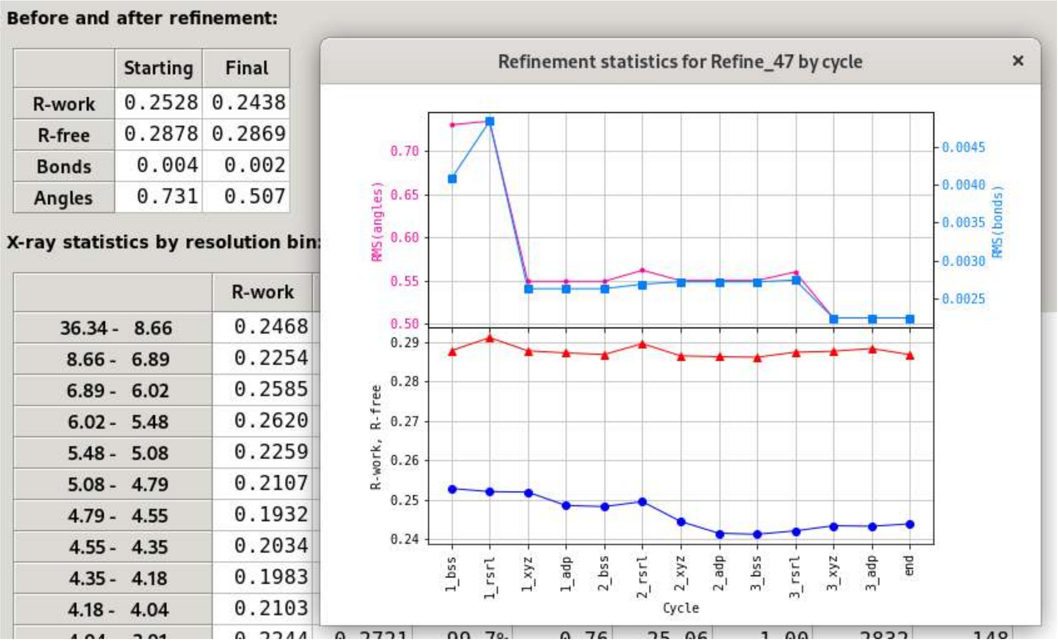
Coot

Refinement

Modelcraft

Phenix



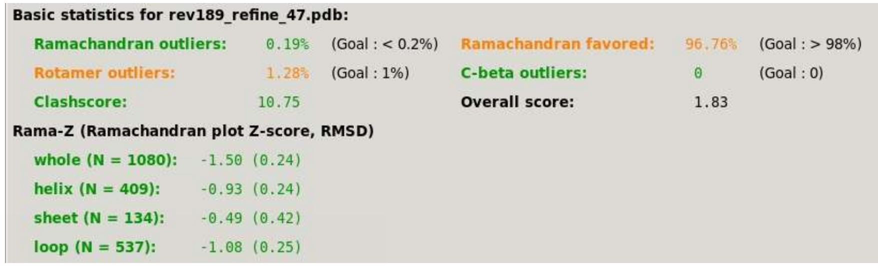


Before and after refinement:

	Starting	Final
R-work	0.2528	0.2438
R-free	0.2878	0.2869
Bonds	0.004	0.002
Angles	0.731	0.507

X-ray statistics by resolution bin:

	R-work
36.34 - 8.66	0.2468
8.66 - 6.89	0.2254
6.89 - 6.02	0.2585
6.02 - 5.48	0.2620
5.48 - 5.08	0.2259
5.08 - 4.79	0.2107
4.79 - 4.55	0.1932
4.55 - 4.35	0.2034
4.35 - 4.18	0.1983
4.18 - 4.04	0.2103

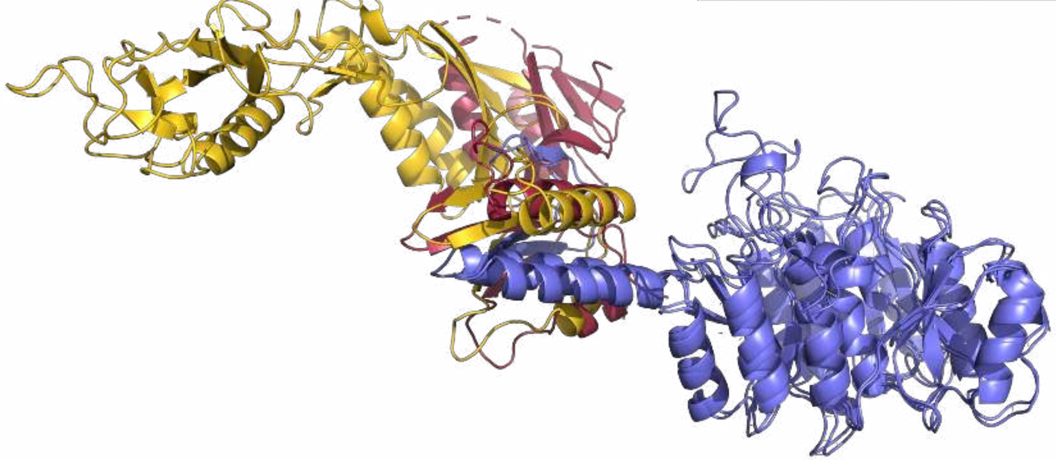


Basic statistics for rev189_refine_47.pdb:

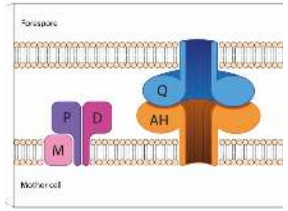
Ramachandran outliers:	0.19%	(Goal : < 0.2%)	Ramachandran favored:	96.76%	(Goal : > 98%)
Rotamer outliers:	1.28%	(Goal : 1%)	C-beta outliers:	0	(Goal : 0)
Clashscore:	10.75		Overall score:	1.83	

Rama-Z (Ramachandran plot Z-score, RMSD)

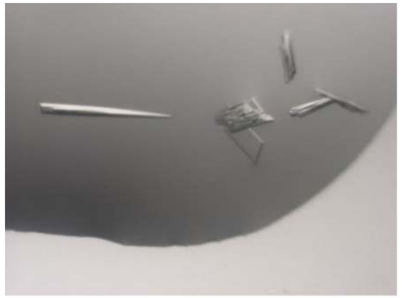
whole (N = 1080):	-1.50 (0.24)
helix (N = 409):	-0.93 (0.24)
sheet (N = 134):	-0.49 (0.42)
loop (N = 537):	-1.08 (0.25)

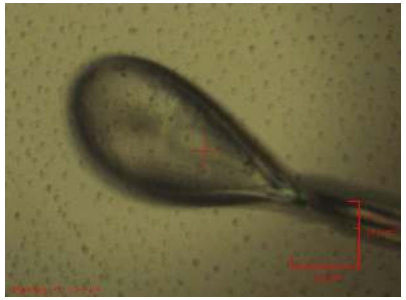


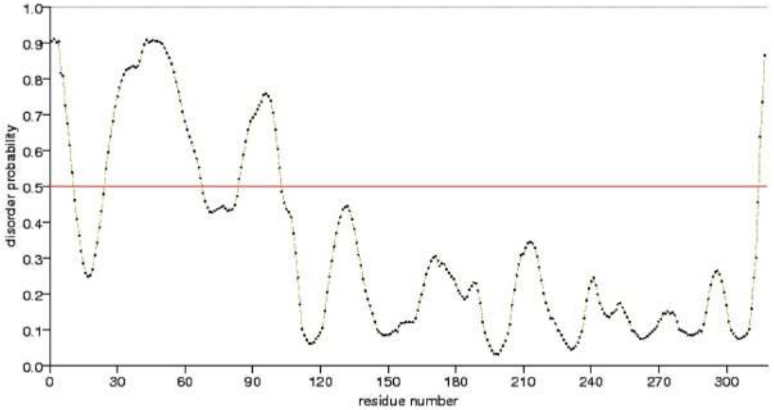
Case study 3: *C. difficile* PG hydrolase

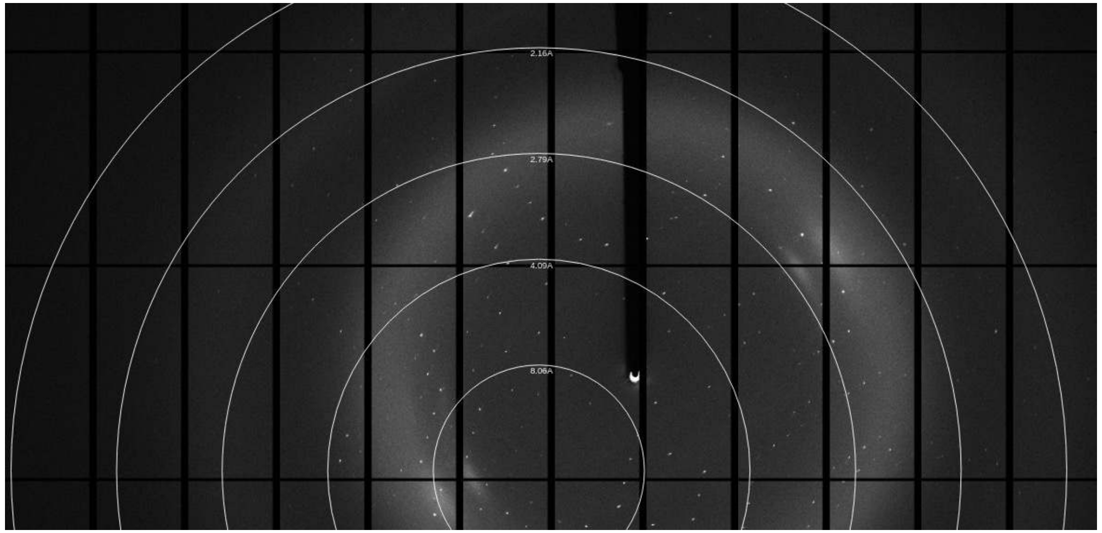


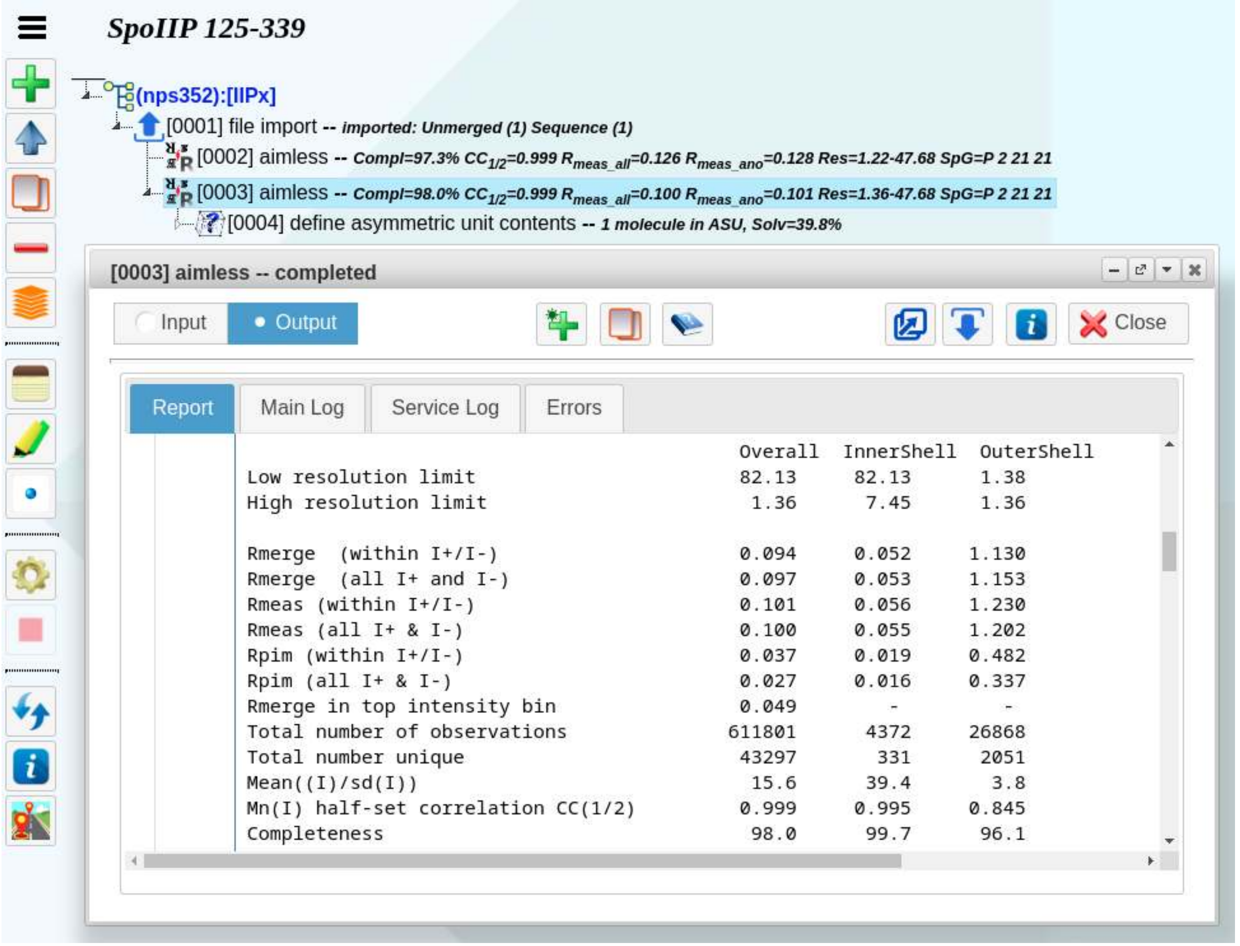
```
1    GAMGNQDDFL KFLVNSSYPE AKVEGNDTEN KKNNKNKETS KENKEESKEE   50
51   NTKSKDASKV DNKKESEKEY IKLYVGKENV PDIESKNSDT TETNTTSSSD  100
101  YKDDLRVTKE NPRILIYHTH GCETYSNSPD GNYHSRDKKN SVMEVGSALT  150
151  SALDSKGWGV VHTTKYHDYP SYNNSYASSL KTIQSILPKY NSVDIAIDLH  200
201  RDARDLTNPA TKEKDHLKYT TMINGERVSK FFFVVGGKNT NRKQLRALAE  250
251  DITAFAEKKY PGLVSPIVEK DYARFNQFAV KNHMLVEIGN NATSVEESKA  300
301  TTKYLAEILD EYFKQKN                                      350
```









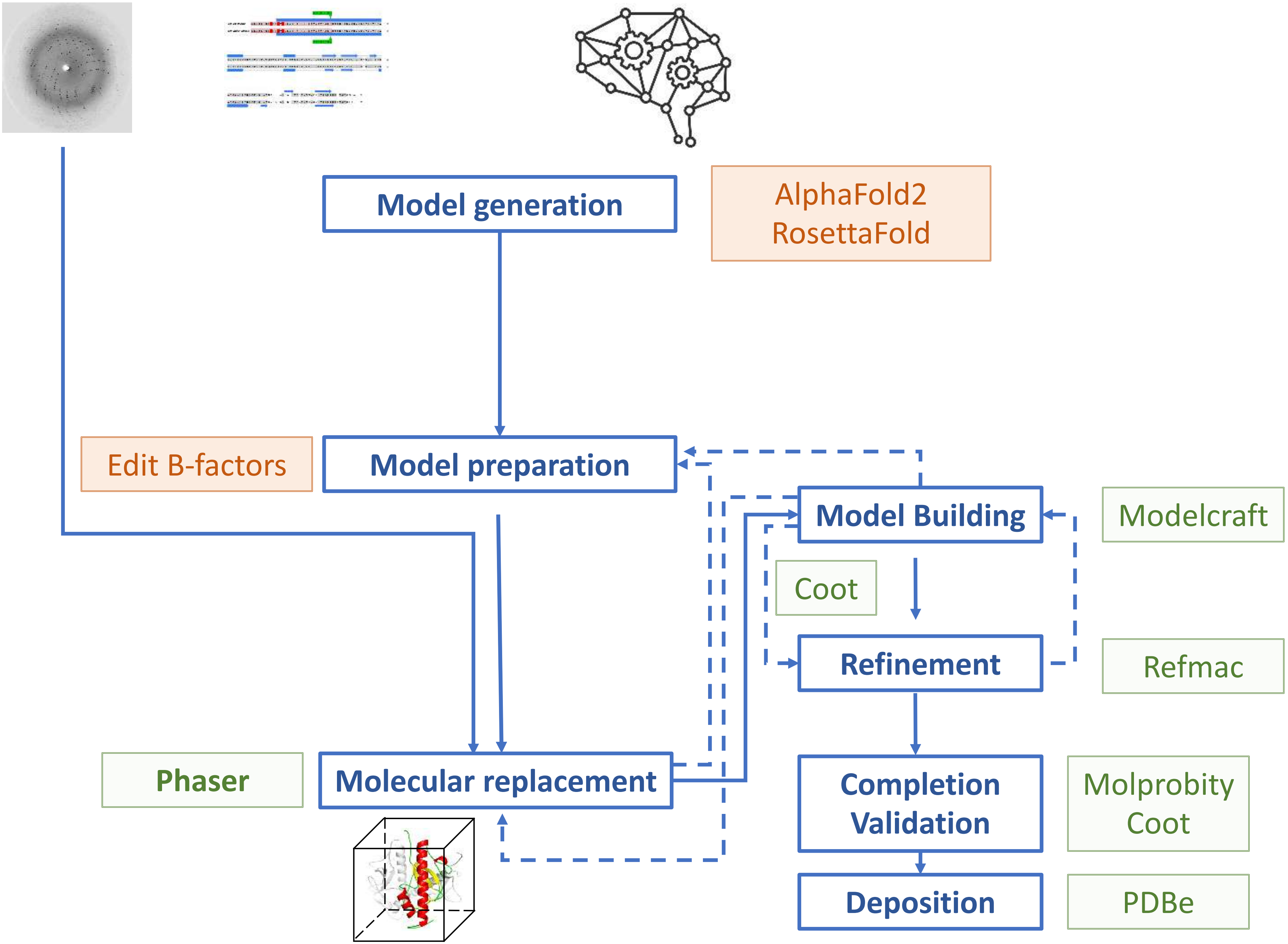


SpoIIP 125-339

(nps352):[IIPx]
[0001] file import -- *imported: Unmerged (1) Sequence (1)*
[0002] aimless -- *Compl=97.3% $CC_{1/2}$=0.999 R_{meas_all}=0.126 R_{meas_ano}=0.128 Res=1.22-47.68 SpG=P 2 21 21*
[0003] aimless -- *Compl=98.0% $CC_{1/2}$=0.999 R_{meas_all}=0.100 R_{meas_ano}=0.101 Res=1.36-47.68 SpG=P 2 21 21*
[0004] define asymmetric unit contents -- *1 molecule in ASU, Solv=39.8%*

[0003] aimless -- completed

```
                                           Overall  InnerShell  OuterShell
Low resolution limit                         82.13     82.13       1.38
High resolution limit                         1.36      7.45       1.36

Rmerge  (within I+/I-)                       0.094     0.052      1.130
Rmerge  (all I+ and I-)                      0.097     0.053      1.153
Rmeas (within I+/I-)                         0.101     0.056      1.230
Rmeas (all I+ & I-)                          0.100     0.055      1.202
Rpim (within I+/I-)                          0.037     0.019      0.482
Rpim (all I+ & I-)                           0.027     0.016      0.337
Rmerge in top intensity bin                  0.049       -          -
Total number of observations                611801      4372      26868
Total number unique                          43297       331       2051
Mean((I)/sd(I))                               15.6      39.4        3.8
Mn(I) half-set correlation CC(1/2)           0.999     0.995      0.845
Completeness                                  98.0      99.7       96.1
```



Model generation

AlphaFold2
RosettaFold

Edit B-factors

Model preparation

Model Building

Modelcraft

Coot

Refinement

Refmac

Phaser

Molecular replacement

Completion
Validation

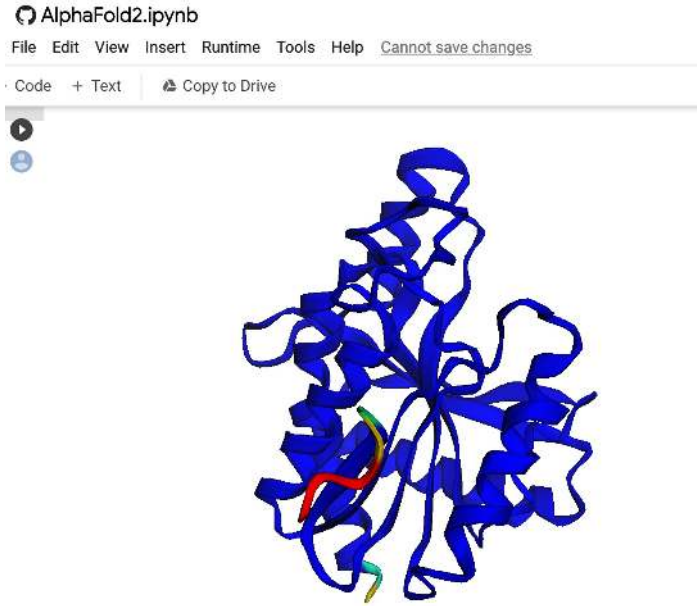
Molprobity
Coot

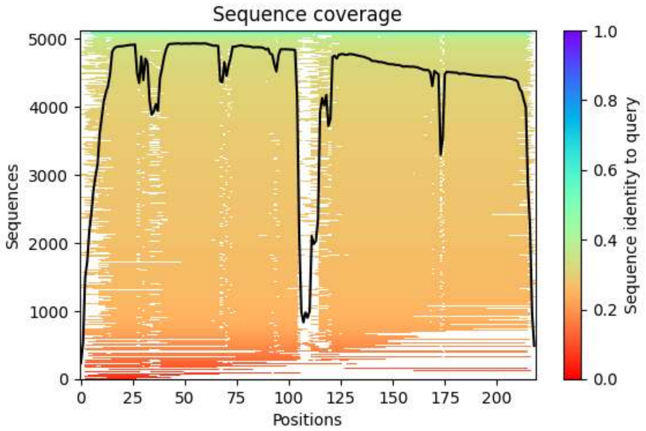
Deposition

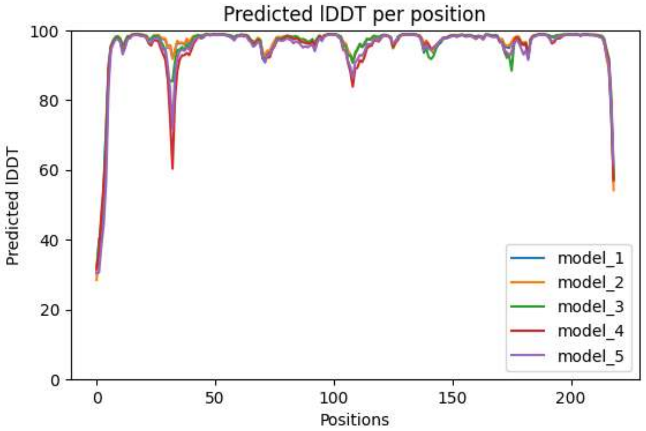
PDBe

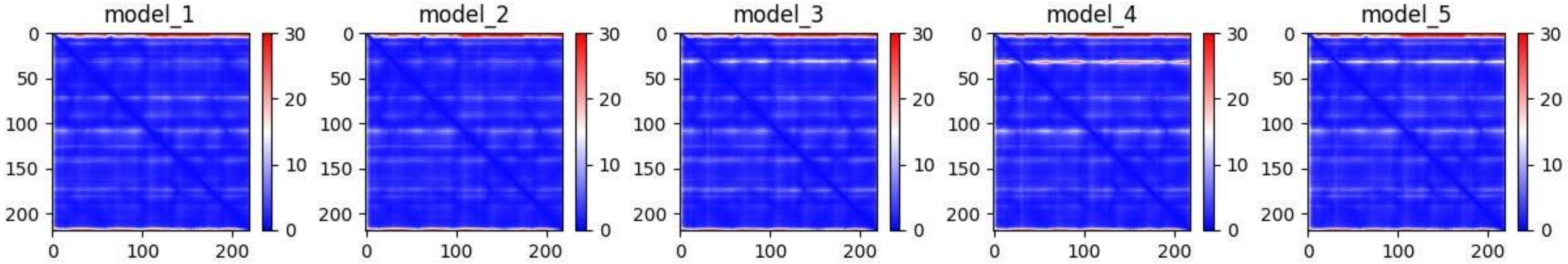
Model generation

AlphaFold2



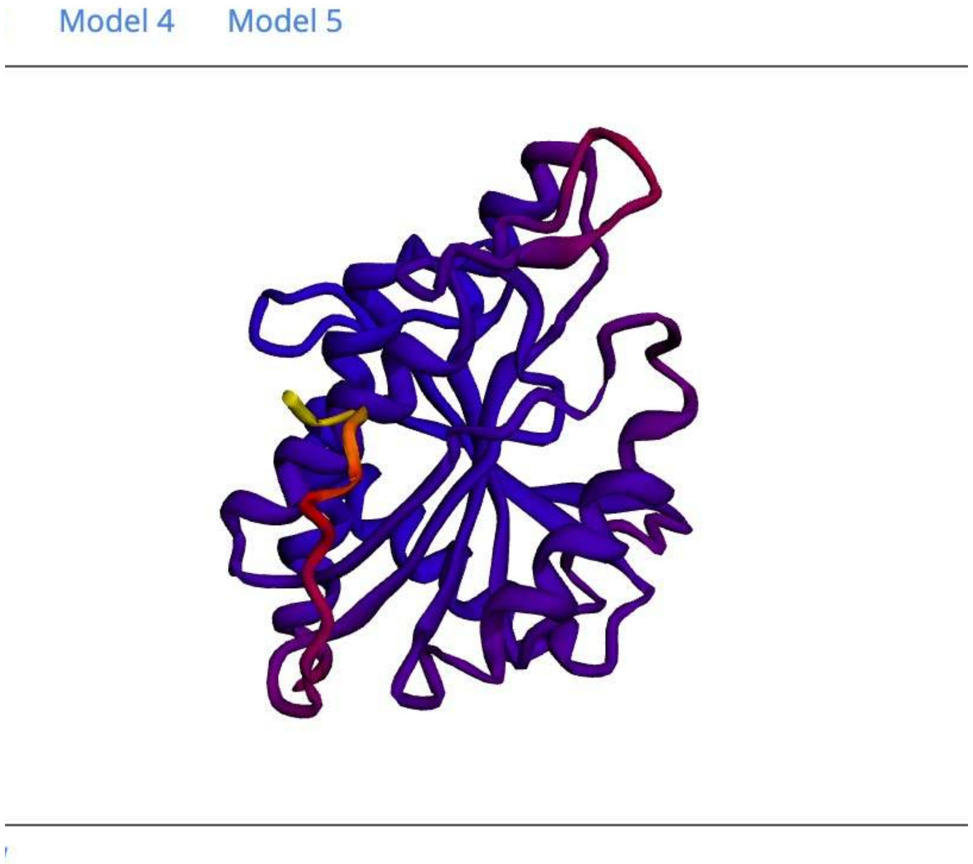


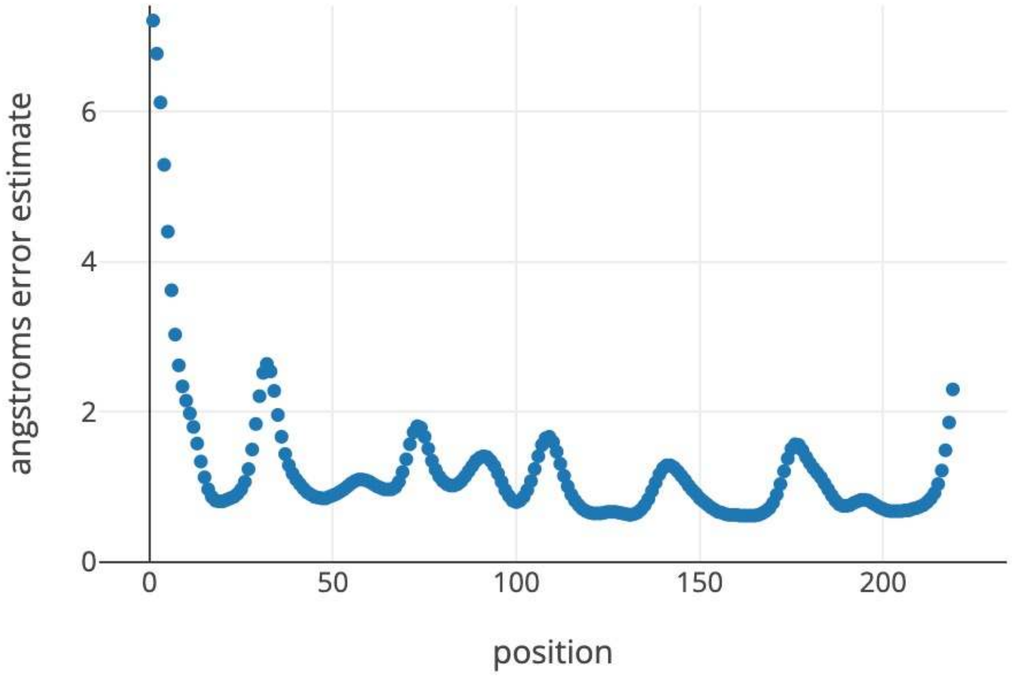




Model generation — RosettaFold

Model preparation — Edit B-factors

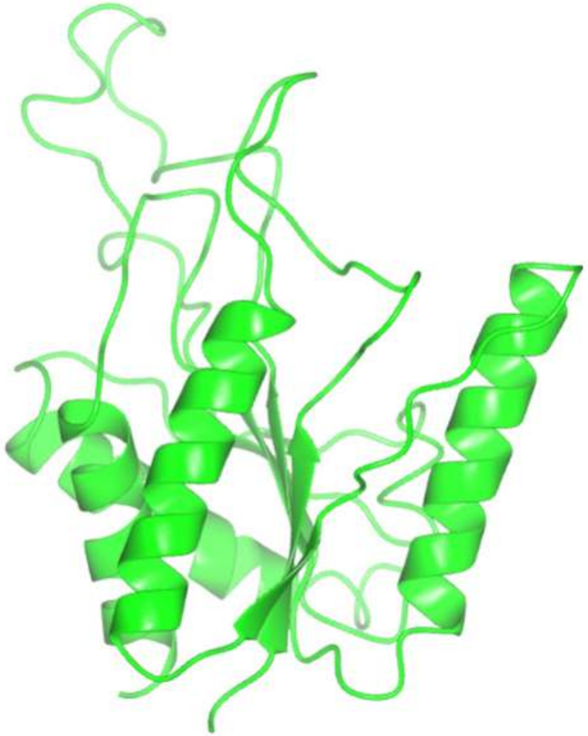




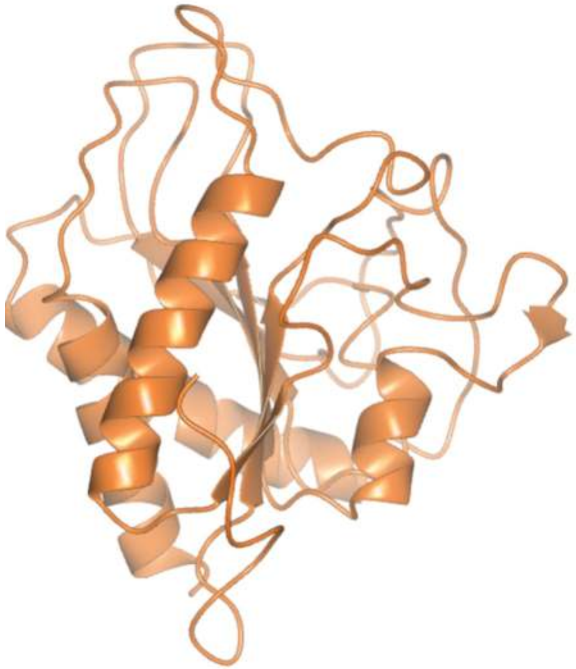
Model generation

Model preparation

SWISS-MODEL



Phyre2



SpoIIP 125-339

(nps352):[IIPx]

[0001] file import -- ***imported: Unmerged (1) Sequence (1)***

[0002] aimless -- ***Compl=97.3% $CC_{1/2}$=0.999 R_{meas_all}=0.126 R_{meas_ano}=0.128 Res=1.22-47.68 SpG=P 2 21 21***

[0003] aimless -- ***Compl=98.0% $CC_{1/2}$=0.999 R_{meas_all}=0.100 R_{meas_ano}=0.101 Res=1.36-47.68 SpG=P 2 21 21***

[0004] define asymmetric unit contents -- ***1 molecule in ASU, Solv=39.8%***

[0007] file import -- ***imported: XYZ (4)***

[0007] file import -- completed

CCP4 v.8.0.016; CCP4 Cloud v.1.7.016
Started: 2023-11-23 13:41:56
Finished: 2023-11-23 13:41:56
CPU: 00.000, Disk: 0.73M

[0007] File(s) Upload and Import

Import Summary

Imported file	*Type*	*Generated dataset(s)*
`domain_model_147888_1.pdb`	XYZ	`[0007-01] domain_model_147888_1 /xyz/protein/` `B-factors corrected assuming AlphaFold model`
`swissmodel1.pdb`	XYZ	`[0007-02] swissmodel1 /xyz/protein/` `B-factors corrected assuming Rosetta model`
`phyre2_model.pdb`	XYZ	`[0007-03] phyre2_model /xyz/protein/`
`sIIPx_3c10b_unrelaxed_model_3_rank_1.pdb`	XYZ	`[0007-04] sIIPx_3c10b_unrelaxed_model_3_rank_1 /xyz/protein/` `B-factors corrected assuming AlphaFold model`

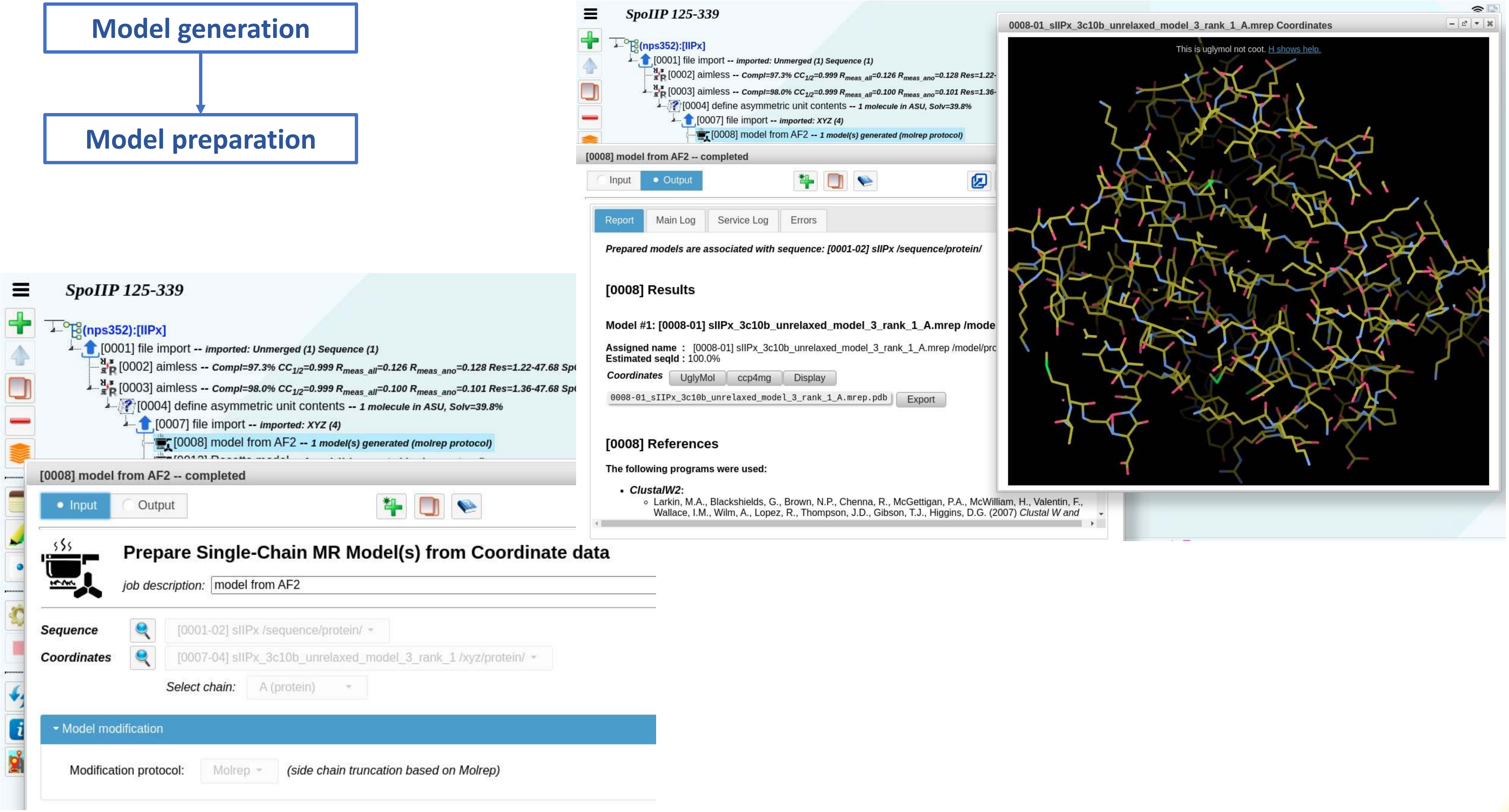
‣ XYZ Coordinates

[0007] References

Model generation

↓

Model preparation



SpoIIP 125-339

(nps352):[IIPx]
[0001] file import -- ***imported: Unmerged (1) Sequence (1)***
[0002] aimless -- ***Compl=97.3% $CC_{1/2}$=0.999 R_{meas_all}=0.126 R_{meas_ano}=0.128 Res=1.22-47.68 Sp***
[0003] aimless -- ***Compl=98.0% $CC_{1/2}$=0.999 R_{meas_all}=0.100 R_{meas_ano}=0.101 Res=1.36-47.68 Sp***
[0004] define asymmetric unit contents -- ***1 molecule in ASU, Solv=39.8%***
[0007] file import -- ***imported: XYZ (4)***
[0008] model from AF2 -- ***1 model(s) generated (molrep protocol)***

[0008] model from AF2 -- completed

Prepare Single-Chain MR Model(s) from Coordinate data

job description: model from AF2

Sequence [0001-02] sIIPx /sequence/protein/

Coordinates [0007-04] sIIPx_3c10b_unrelaxed_model_3_rank_1 /xyz/protein/

Select chain: A (protein)

Model modification

Modification protocol: Molrep *(side chain truncation based on Molrep)*

SpoIIP 125-339

(nps352):[IIPx]
[0001] file import -- ***imported: Unmerged (1) Sequence (1)***
[0002] aimless -- ***Compl=97.3% $CC_{1/2}$=0.999 R_{meas_all}=0.126 R_{meas_ano}=0.128 Res=1.22-***
[0003] aimless -- ***Compl=98.0% $CC_{1/2}$=0.999 R_{meas_all}=0.100 R_{meas_ano}=0.101 Res=1.36-***
[0004] define asymmetric unit contents -- ***1 molecule in ASU, Solv=39.8%***
[0007] file import -- ***imported: XYZ (4)***
[0008] model from AF2 -- ***1 model(s) generated (molrep protocol)***

[0008] model from AF2 -- completed

Prepared models are associated with sequence: [0001-02] sIIPx /sequence/protein/

[0008] Results

Model #1: [0008-01] sIIPx_3c10b_unrelaxed_model_3_rank_1_A.mrep /mode

Assigned name : [0008-01] sIIPx_3c10b_unrelaxed_model_3_rank_1_A.mrep /model/pro
Estimated seqId : 100.0%

Coordinates

`0008-01_sIIPx_3c10b_unrelaxed_model_3_rank_1_A.mrep.pdb`

[0008] References

The following programs were used:

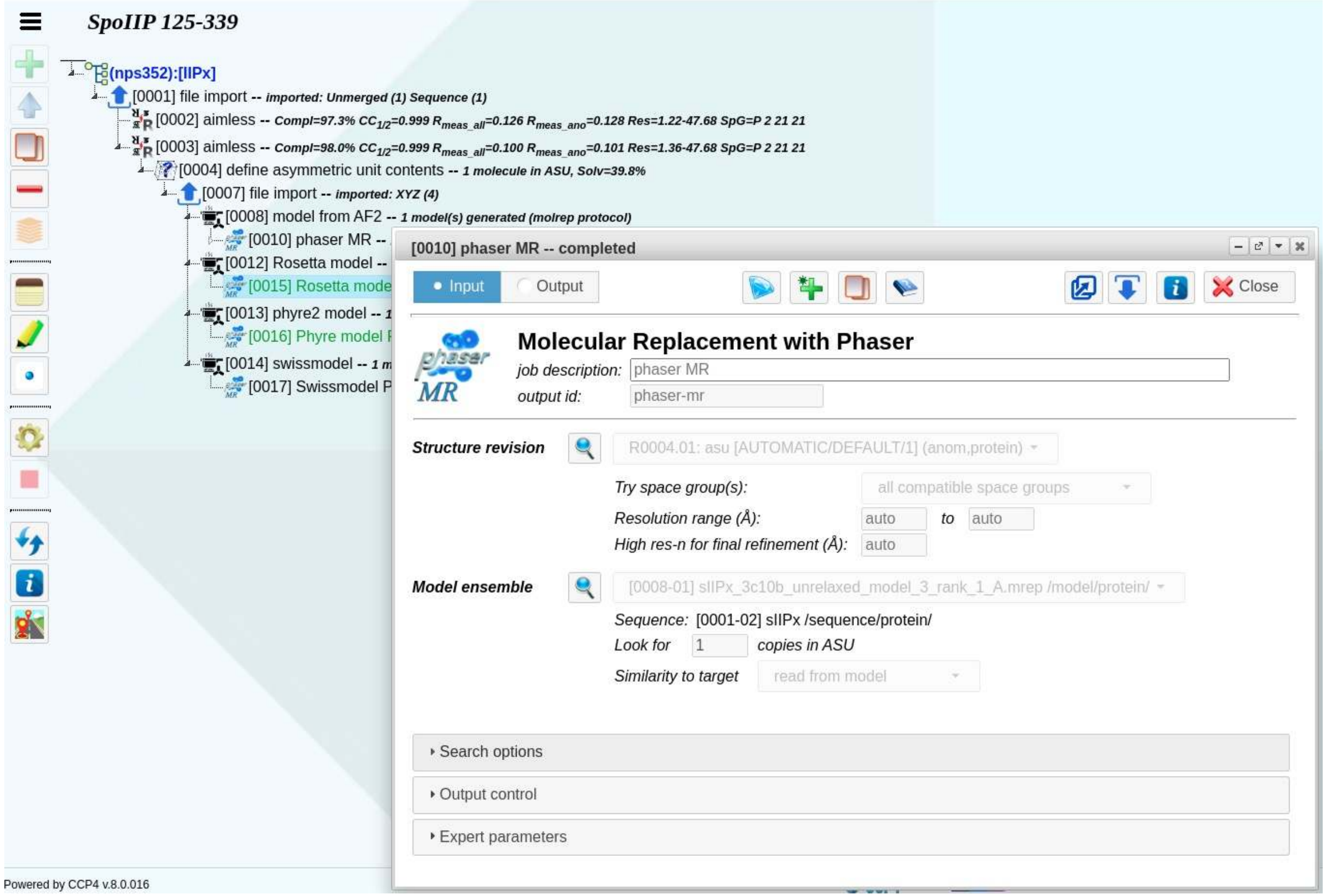
- ***ClustalW2*:**
  - Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G. (2007) *Clustal W and*

0008-01_sIIPx_3c10b_unrelaxed_model_3_rank_1_A.mrep Coordinates

This is uglymol not coot. H shows help.

Molecular replacement

Phaser



SpoIIP 125-339

(nps352):[IIPx]

- [0001] file import -- ***imported: Unmerged (1) Sequence (1)***
  - [0002] aimless -- ***Compl=97.3% $CC_{1/2}$=0.999 R_{meas_all}=0.126 R_{meas_ano}=0.128 Res=1.22-47.68 SpG=P 2 21 21***
  - [0003] aimless -- ***Compl=98.0% $CC_{1/2}$=0.999 R_{meas_all}=0.100 R_{meas_ano}=0.101 Res=1.36-47.68 SpG=P 2 21 21***
    - [0004] define asymmetric unit contents -- ***1 molecule in ASU, Solv=39.8%***
      - [0007] file import -- ***imported: XYZ (4)***
        - [0008] model from AF2 -- ***1 model(s) generated (molrep protocol)***
          - [0010] phaser MR --
        - [0012] Rosetta model --
          - [0015] Rosetta mode
        - [0013] phyre2 model -- *1*
          - [0016] Phyre model F
        - [0014] swissmodel -- *1 m*
          - [0017] Swissmodel P

Powered by CCP4 v.8.0.016

[0010] phaser MR -- completed

● Input ○ Output

Molecular Replacement with Phaser

job description: phaser MR

output id: phaser-mr

Structure revision R0004.01: asu [AUTOMATIC/DEFAULT/1] (anom,protein)

Try space group(s): all compatible space groups

Resolution range (Å): auto *to* auto

High res-n for final refinement (Å): auto

Model ensemble [0008-01] sIIPx_3c10b_unrelaxed_model_3_rank_1_A.mrep /model/protein/

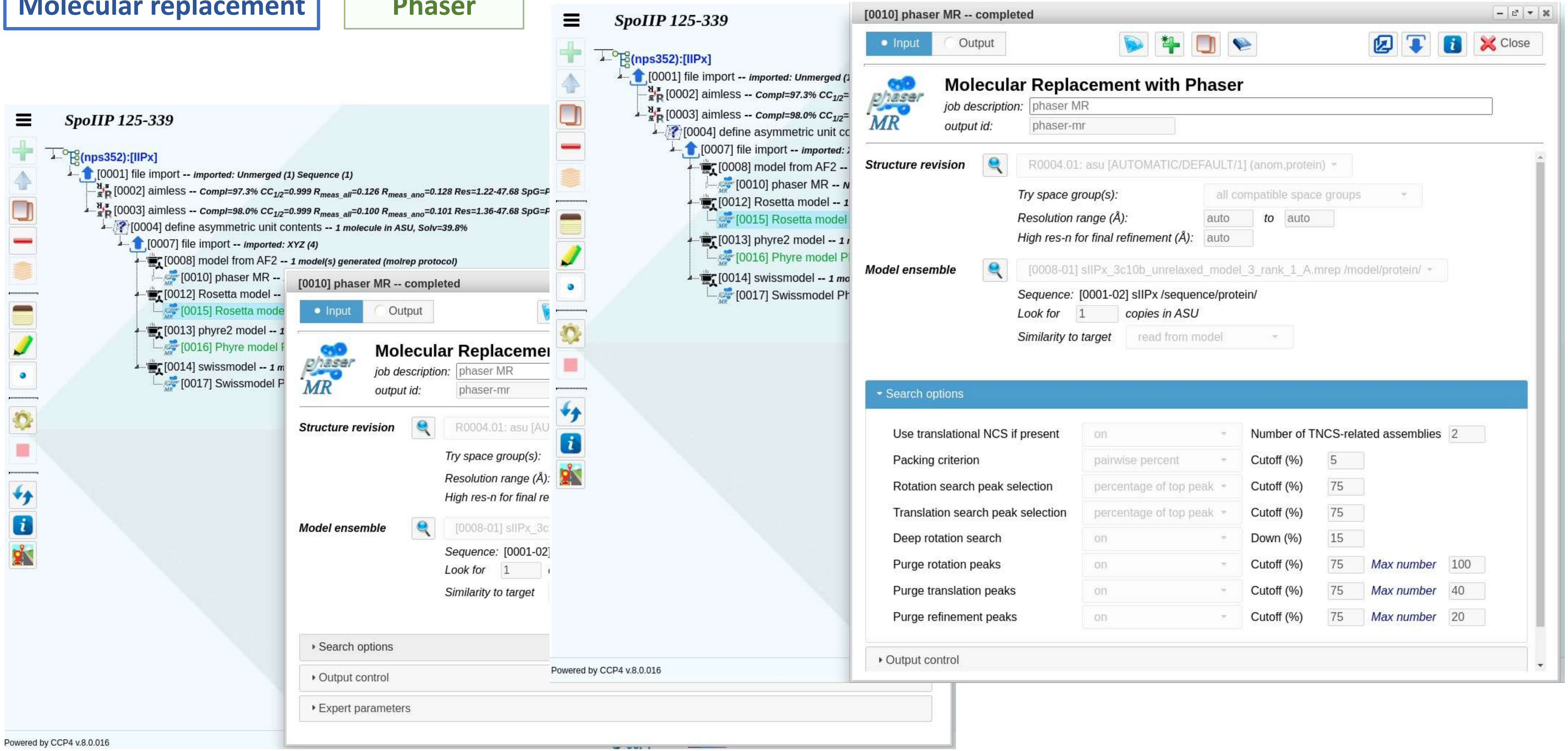
Sequence: [0001-02] sIIPx /sequence/protein/

Look for 1 *copies in ASU*

Similarity to target read from model

Molecular replacement

Phaser



SpoIIP 125-339

(nps352):[IIPx]
- [0001] file import -- *imported: Unmerged (1) Sequence (1)*
  - [0002] aimless -- *Compl=97.3% $CC_{1/2}$=0.999 R_{meas_all}=0.126 R_{meas_ano}=0.128 Res=1.22-47.68 SpG=P*
  - [0003] aimless -- *Compl=98.0% $CC_{1/2}$=0.999 R_{meas_all}=0.100 R_{meas_ano}=0.101 Res=1.36-47.68 SpG=P*
    - [0004] define asymmetric unit contents -- *1 molecule in ASU, Solv=39.8%*
      - [0007] file import -- *imported: XYZ (4)*
        - [0008] model from AF2 -- *1 model(s) generated (molrep protocol)*
          - [0010] phaser MR --
        - [0012] Rosetta model --
          - [0015] Rosetta model
        - [0013] phyre2 model -- *1*
          - [0016] Phyre model P
        - [0014] swissmodel -- *1 m*
          - [0017] Swissmodel P

[0010] phaser MR -- completed

● Input ○ Output

Molecular Replacement with Phaser

job description: phaser MR

output id: phaser-mr

Structure revision: R0004.01: asu [AUTOMATIC/DEFAULT/1] (anom,protein)

Try space group(s): all compatible space groups

Resolution range (Å): auto *to* auto

High res-n for final refinement (Å): auto

Model ensemble: [0008-01] sIIPx_3c10b_unrelaxed_model_3_rank_1_A.mrep /model/protein/

Sequence: [0001-02] sIIPx /sequence/protein/

Look for 1 *copies in ASU*

Similarity to target read from model

Search options

Use translational NCS if present	on	Number of TNCS-related assemblies	2		
Packing criterion	pairwise percent	Cutoff (%)	5		
Rotation search peak selection	percentage of top peak	Cutoff (%)	75		
Translation search peak selection	percentage of top peak	Cutoff (%)	75		
Deep rotation search	on	Down (%)	15		
Purge rotation peaks	on	Cutoff (%)	75	*Max number*	100
Purge translation peaks	on	Cutoff (%)	75	*Max number*	40
Purge refinement peaks	on	Cutoff (%)	75	*Max number*	20

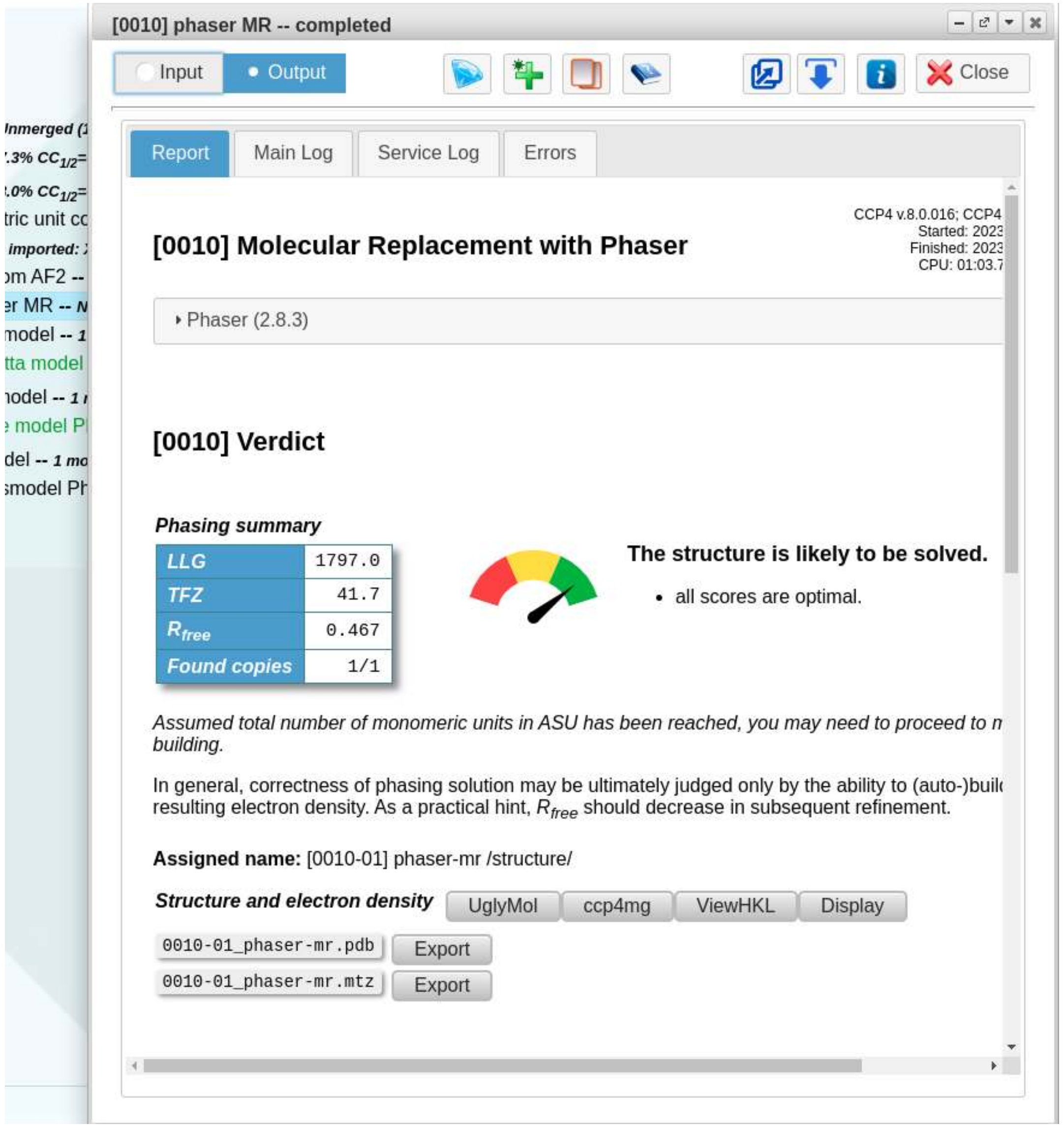
Output control

Expert parameters

Powered by CCP4 v.8.0.016

Molecular replacement

Phaser



CCP4 v.8.0.016; CCP4
Started: 2023
Finished: 2023
CPU: 01:03.7

[0010] Molecular Replacement with Phaser

‣ Phaser (2.8.3)

[0010] Verdict

Phasing summary

LLG	1797.0
TFZ	41.7
R_{free}	0.467
Found copies	1/1

The structure is likely to be solved.

- all scores are optimal.

Assumed total number of monomeric units in ASU has been reached, you may need to proceed to m building.

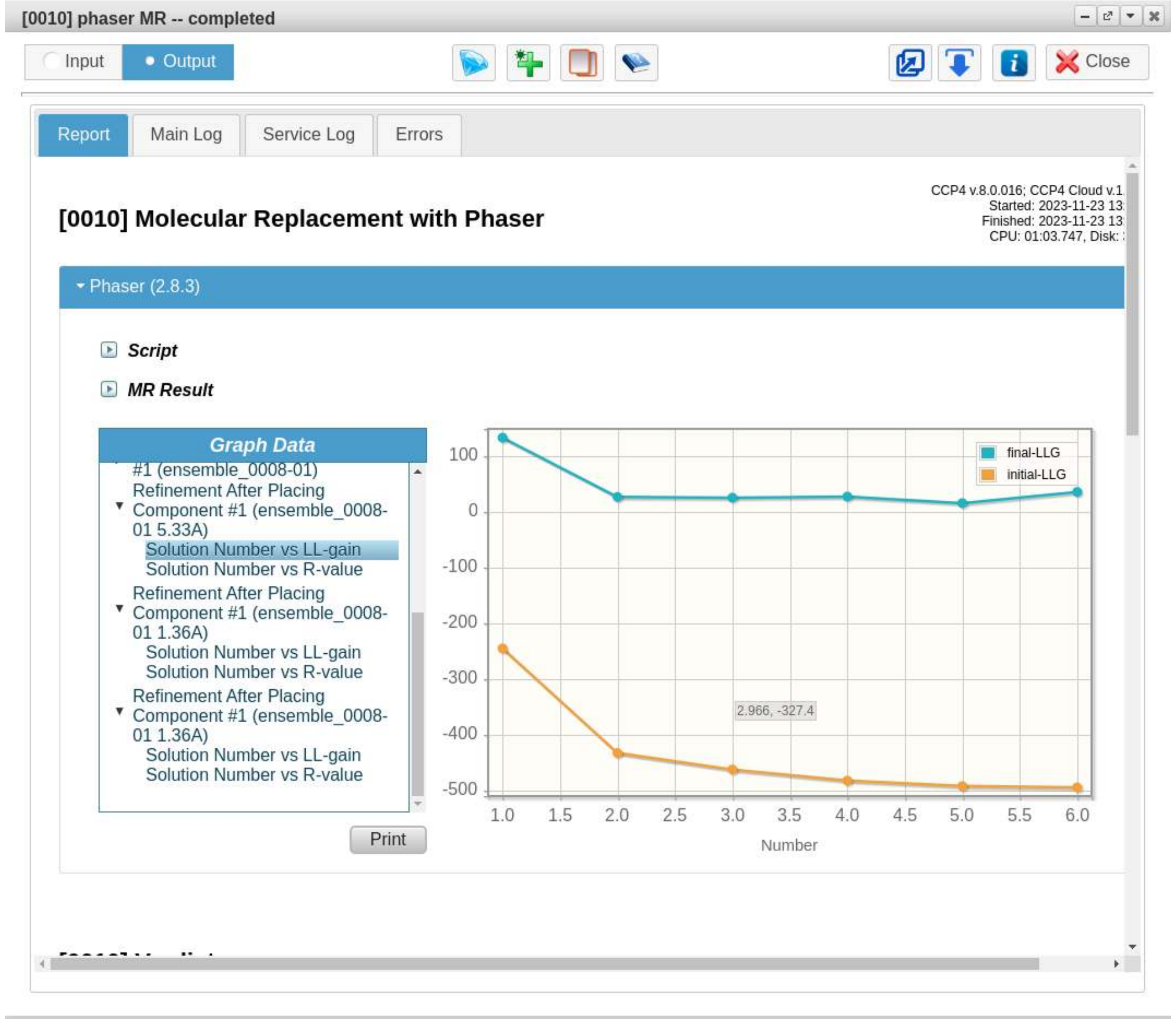
In general, correctness of phasing solution may be ultimately judged only by the ability to (auto-)buil resulting electron density. As a practical hint, R_{free} should decrease in subsequent refinement.

Assigned name: [0010-01] phaser-mr /structure/

Structure and electron density

`0010-01_phaser-mr.pdb`

`0010-01_phaser-mr.mtz`



CCP4 v.8.0.016; CCP4 Cloud v.1
Started: 2023-11-23 13
Finished: 2023-11-23 13
CPU: 01:03.747, Disk:

[0010] Molecular Replacement with Phaser

▾ Phaser (2.8.3)

Script

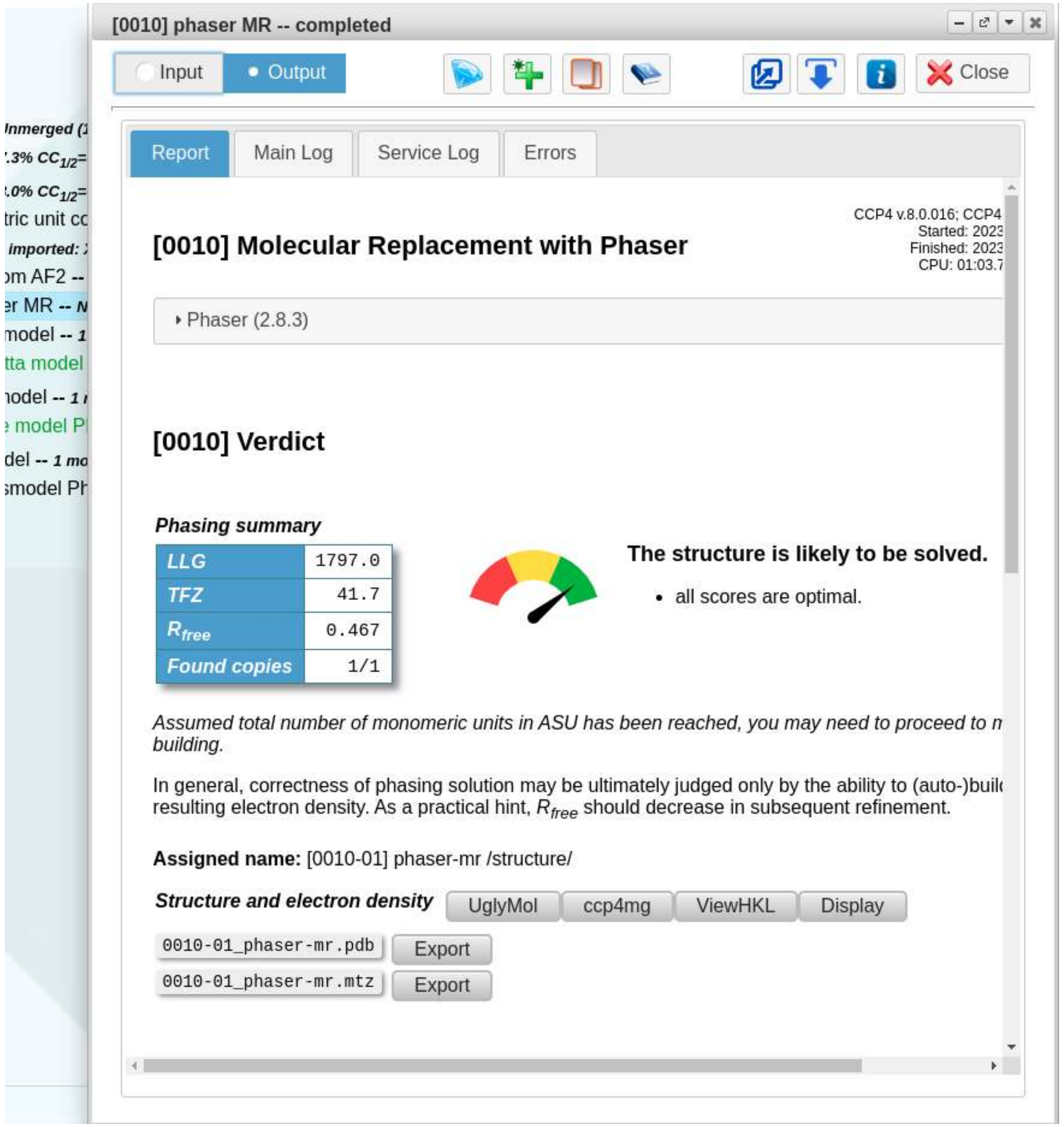
MR Result

Graph Data

- #1 (ensemble_0008-01)
- ▾ Refinement After Placing Component #1 (ensemble_0008-01 5.33A)
  - Solution Number vs LL-gain
  - Solution Number vs R-value
- ▾ Refinement After Placing Component #1 (ensemble_0008-01 1.36A)
  - Solution Number vs LL-gain
  - Solution Number vs R-value
- ▾ Refinement After Placing Component #1 (ensemble_0008-01 1.36A)
  - Solution Number vs LL-gain
  - Solution Number vs R-value

Molecular replacement

Phaser



CCP4 v.8.0.016; CCP4
Started: 2023
Finished: 2023
CPU: 01:03.7

[0010] Molecular Replacement with Phaser

‣ Phaser (2.8.3)

[0010] Verdict

Phasing summary

LLG	1797.0
TFZ	41.7
R_{free}	0.467
Found copies	1/1

The structure is likely to be solved.

- all scores are optimal.

Assumed total number of monomeric units in ASU has been reached, you may need to proceed to m building.

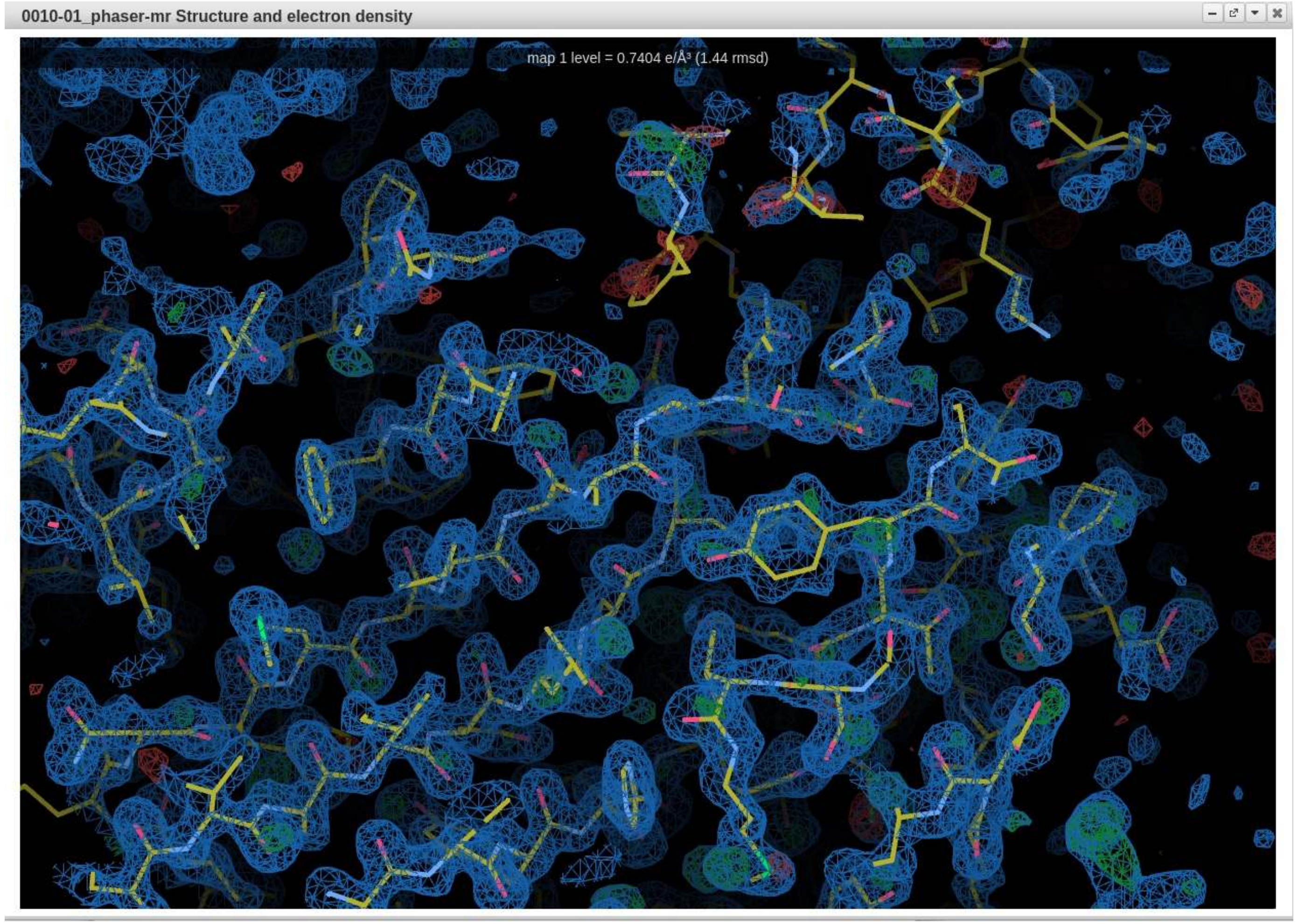
In general, correctness of phasing solution may be ultimately judged only by the ability to (auto-)buil resulting electron density. As a practical hint, R_{free} should decrease in subsequent refinement.

Assigned name: [0010-01] phaser-mr /structure/

Structure and electron density UglyMol ccp4mg ViewHKL Display

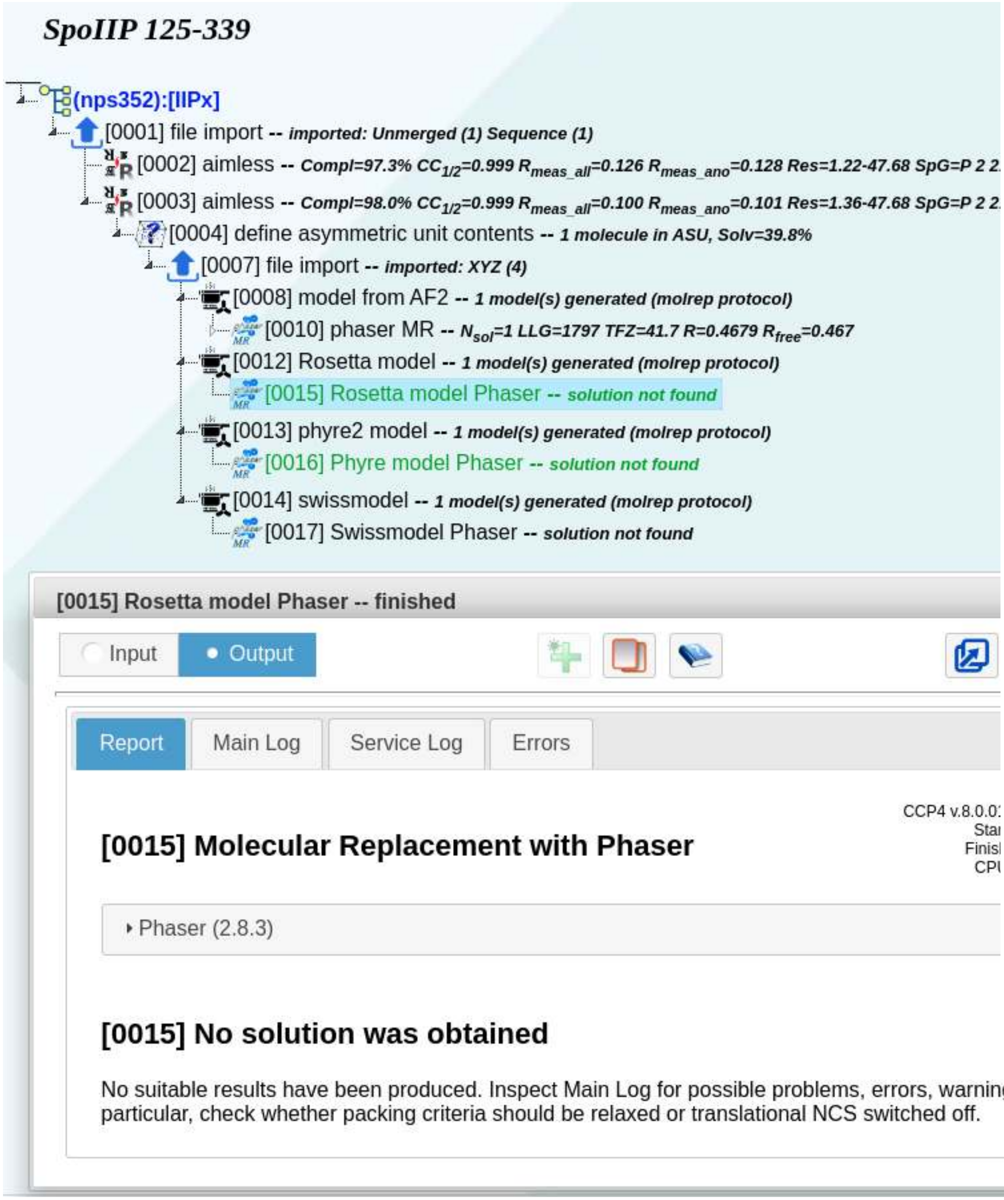
`0010-01_phaser-mr.pdb` Export

`0010-01_phaser-mr.mtz` Export



Molecular replacement

Phaser



SpoIIP 125-339

- (nps352):[IIPx]
  - [0001] file import -- *imported: Unmerged (1) Sequence (1)*
    - [0002] aimless -- *Compl=97.3% $CC_{1/2}$=0.999 R_{meas_all}=0.126 R_{meas_ano}=0.128 Res=1.22-47.68 SpG=P 2 2*
    - [0003] aimless -- *Compl=98.0% $CC_{1/2}$=0.999 R_{meas_all}=0.100 R_{meas_ano}=0.101 Res=1.36-47.68 SpG=P 2 2*
      - [0004] define asymmetric unit contents -- *1 molecule in ASU, Solv=39.8%*
        - [0007] file import -- *imported: XYZ (4)*
          - [0008] model from AF2 -- *1 model(s) generated (molrep protocol)*
            - [0010] phaser MR -- *N_{sol}=1 LLG=1797 TFZ=41.7 R=0.4679 R_{free}=0.467*
          - [0012] Rosetta model -- *1 model(s) generated (molrep protocol)*
            - [0015] Rosetta model Phaser -- *solution not found*
          - [0013] phyre2 model -- *1 model(s) generated (molrep protocol)*
            - [0016] Phyre model Phaser -- *solution not found*
          - [0014] swissmodel -- *1 model(s) generated (molrep protocol)*
            - [0017] Swissmodel Phaser -- *solution not found*

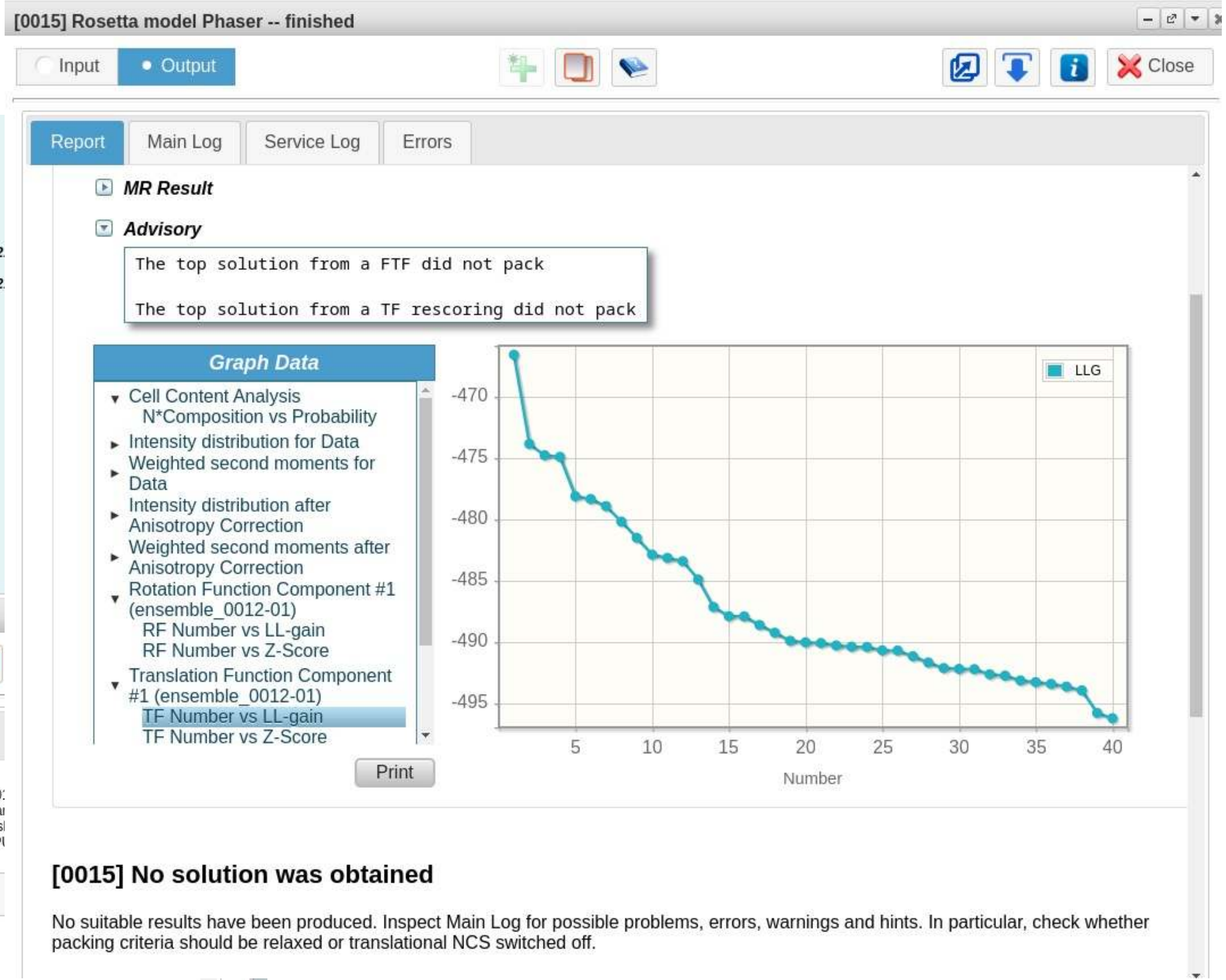
[0015] Rosetta model Phaser -- finished

CCP4 v.8.0.0
Star
Finis
CPU

[0015] Molecular Replacement with Phaser

[0015] No solution was obtained

No suitable results have been produced. Inspect Main Log for possible problems, errors, warnings and hints. In particular, check whether packing criteria should be relaxed or translational NCS switched off.



[0015] Rosetta model Phaser -- finished

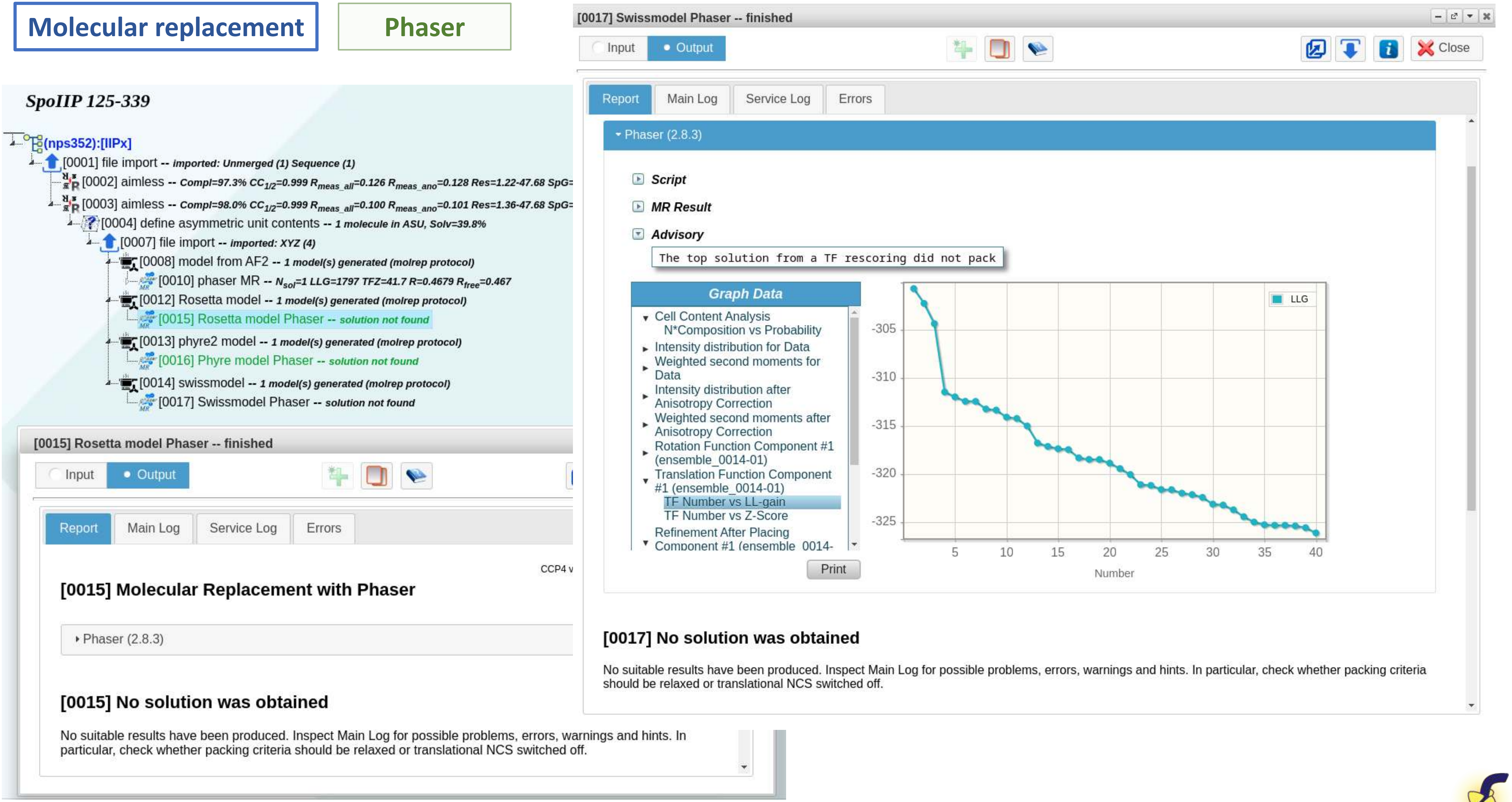
```
The top solution from a FTF did not pack

The top solution from a TF rescoring did not pack
```

[0015] No solution was obtained

No suitable results have been produced. Inspect Main Log for possible problems, errors, warnings and hints. In particular, check whether packing criteria should be relaxed or translational NCS switched off.

Molecular replacement

Phaser



SpoIIP 125-339

(nps352):[IIPx]
[0001] file import -- ***imported: Unmerged (1) Sequence (1)***
[0002] aimless -- ***Compl=97.3% $CC_{1/2}$=0.999 R_{meas_all}=0.126 R_{meas_ano}=0.128 Res=1.22-47.68 SpG=***
[0003] aimless -- ***Compl=98.0% $CC_{1/2}$=0.999 R_{meas_all}=0.100 R_{meas_ano}=0.101 Res=1.36-47.68 SpG=***
[0004] define asymmetric unit contents -- ***1 molecule in ASU, Solv=39.8%***
[0007] file import -- ***imported: XYZ (4)***
[0008] model from AF2 -- ***1 model(s) generated (molrep protocol)***
[0010] phaser MR -- ***N_{sol}=1 LLG=1797 TFZ=41.7 R=0.4679 R_{free}=0.467***
[0012] Rosetta model -- ***1 model(s) generated (molrep protocol)***
[0015] Rosetta model Phaser -- ***solution not found***
[0013] phyre2 model -- ***1 model(s) generated (molrep protocol)***
[0016] Phyre model Phaser -- ***solution not found***
[0014] swissmodel -- ***1 model(s) generated (molrep protocol)***
[0017] Swissmodel Phaser -- ***solution not found***

[0015] Rosetta model Phaser -- finished

[0015] Molecular Replacement with Phaser

▸ Phaser (2.8.3)

[0015] No solution was obtained

No suitable results have been produced. Inspect Main Log for possible problems, errors, warnings and hints. In particular, check whether packing criteria should be relaxed or translational NCS switched off.

[0017] Swissmodel Phaser -- finished

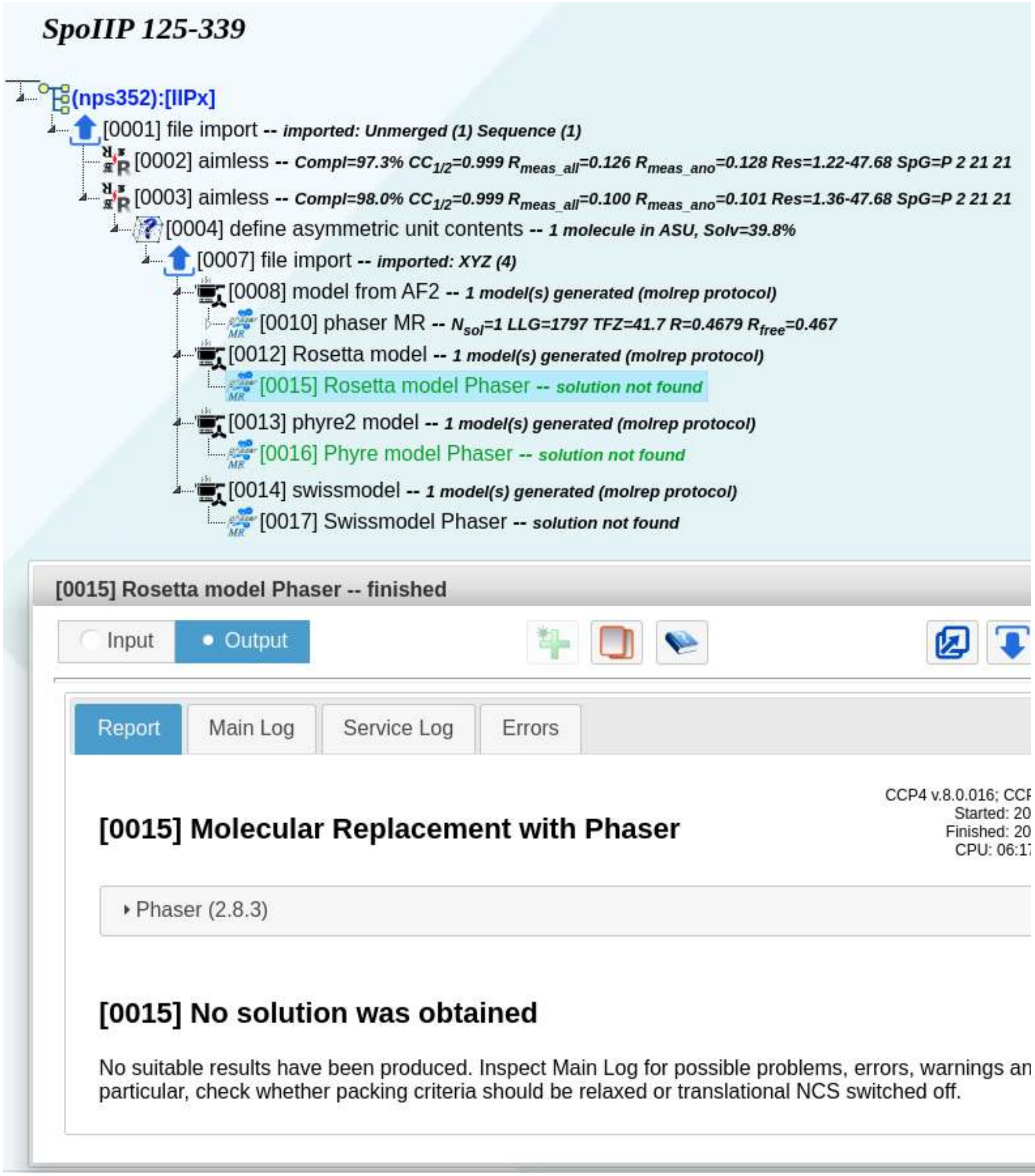
```
The top solution from a TF rescoring did not pack
```

[0017] No solution was obtained

No suitable results have been produced. Inspect Main Log for possible problems, errors, warnings and hints. In particular, check whether packing criteria should be relaxed or translational NCS switched off.

Molecular replacement

Phaser



SpoIIP 125-339

```
(nps352):[IIPx]
[0001] file import -- imported: Unmerged (1) Sequence (1)
  [0002] aimless -- Compl=97.3% CC1/2=0.999 Rmeas_all=0.126 Rmeas_ano=0.128 Res=1.22-47.68 SpG=P 2 21 21
  [0003] aimless -- Compl=98.0% CC1/2=0.999 Rmeas_all=0.100 Rmeas_ano=0.101 Res=1.36-47.68 SpG=P 2 21 21
    [0004] define asymmetric unit contents -- 1 molecule in ASU, Solv=39.8%
      [0007] file import -- imported: XYZ (4)
        [0008] model from AF2 -- 1 model(s) generated (molrep protocol)
          [0010] phaser MR -- Nsol=1 LLG=1797 TFZ=41.7 R=0.4679 Rfree=0.467
        [0012] Rosetta model -- 1 model(s) generated (molrep protocol)
          [0015] Rosetta model Phaser -- solution not found
        [0013] phyre2 model -- 1 model(s) generated (molrep protocol)
          [0016] Phyre model Phaser -- solution not found
        [0014] swissmodel -- 1 model(s) generated (molrep protocol)
          [0017] Swissmodel Phaser -- solution not found
```

[0015] Rosetta model Phaser -- finished

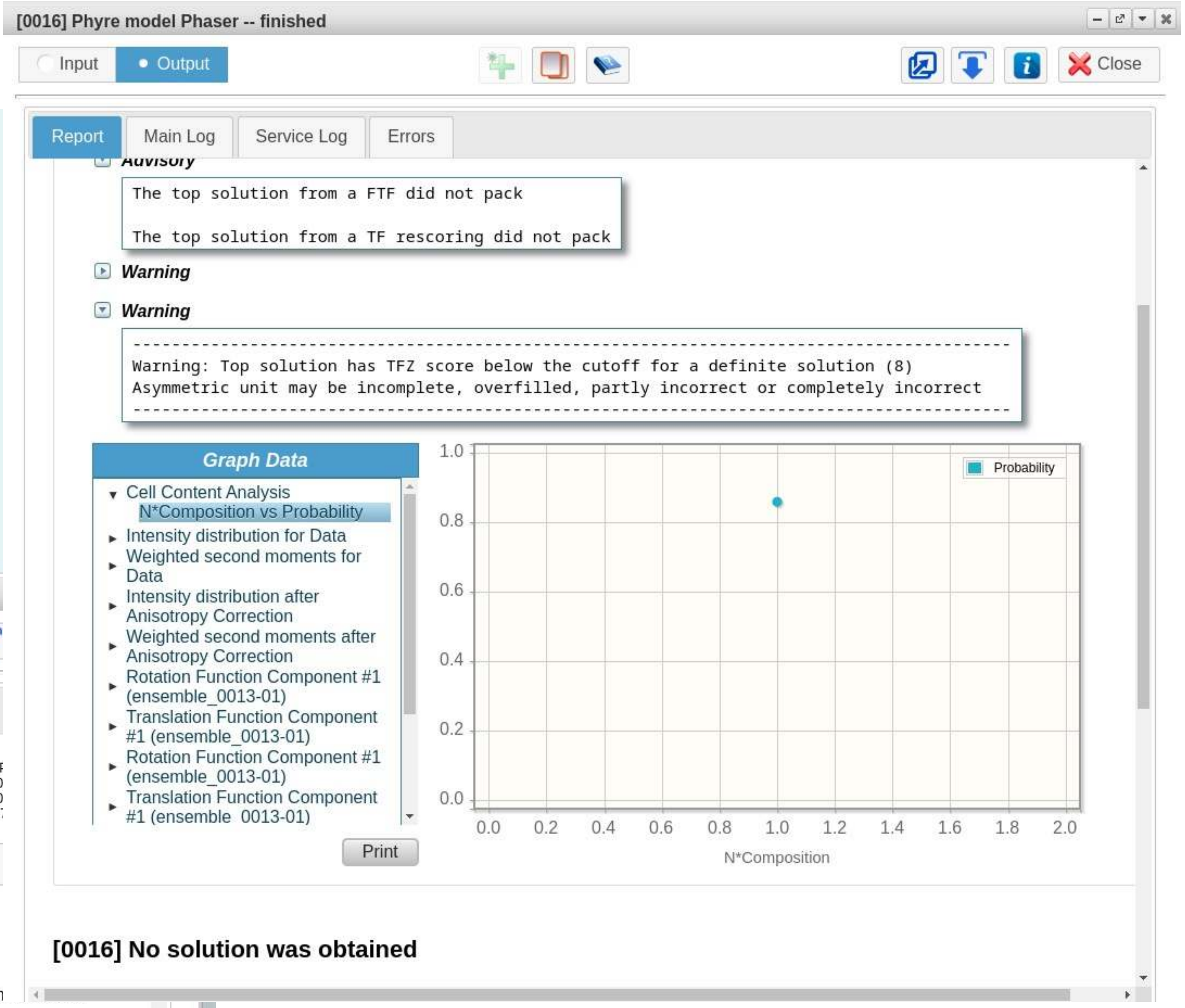
CCP4 v.8.0.016; CCF
Started: 20
Finished: 20
CPU: 06:1

[0015] Molecular Replacement with Phaser

‣ Phaser (2.8.3)

[0015] No solution was obtained

No suitable results have been produced. Inspect Main Log for possible problems, errors, warnings an particular, check whether packing criteria should be relaxed or translational NCS switched off.



[0016] Phyre model Phaser -- finished

Advisory

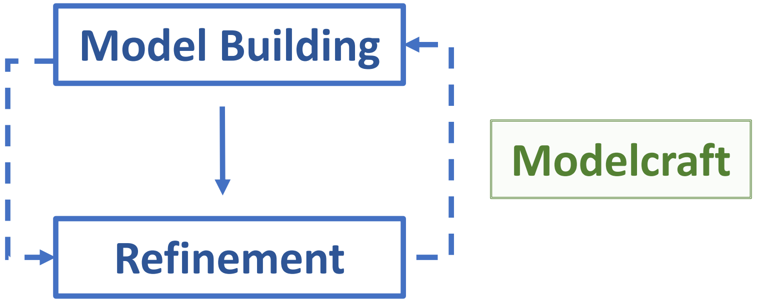
```
The top solution from a FTF did not pack

The top solution from a TF rescoring did not pack
```

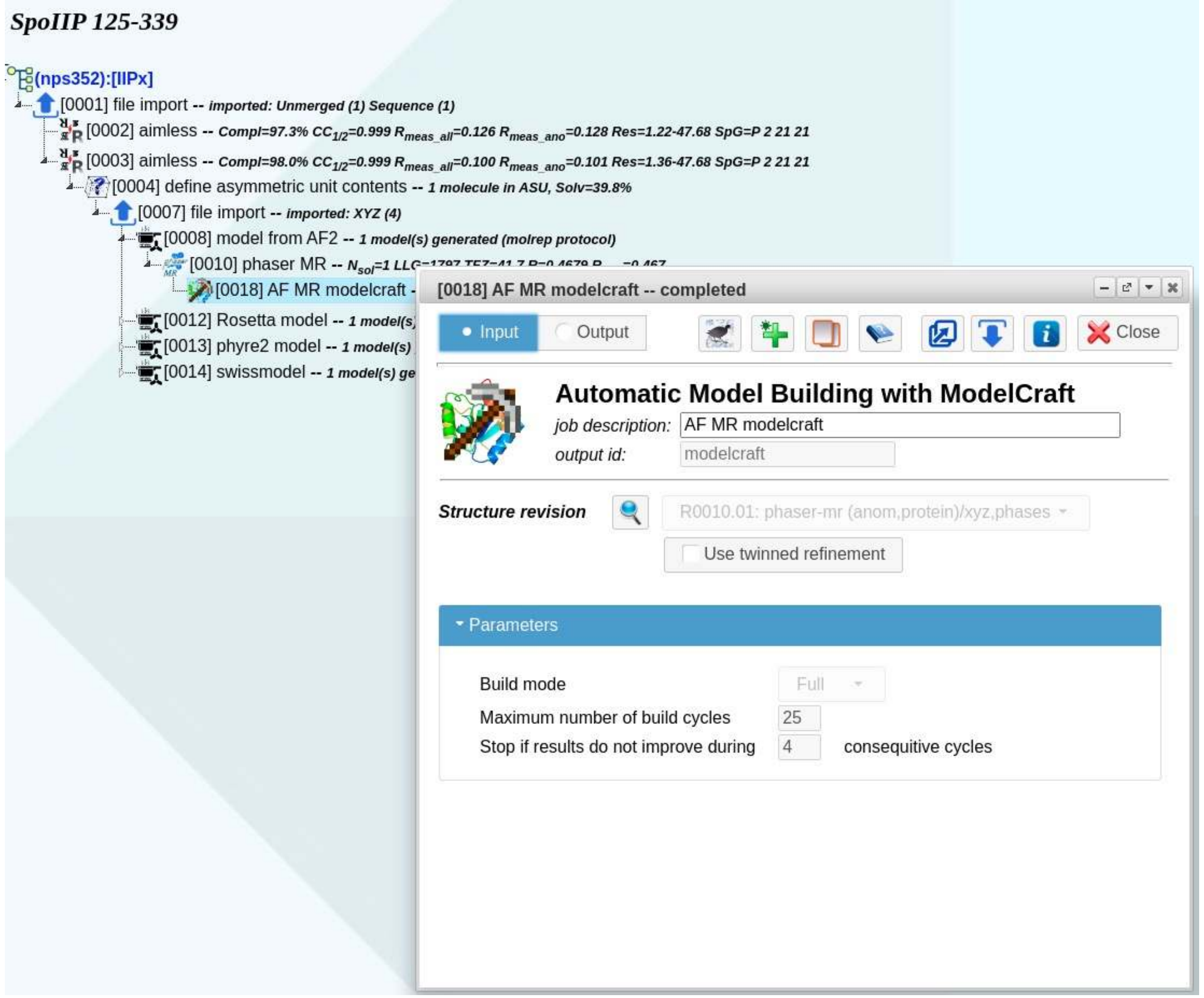
Warning

Warning

```
------------------------------------------------------------------------------------------
Warning: Top solution has TFZ score below the cutoff for a definite solution (8)
Asymmetric unit may be incomplete, overfilled, partly incorrect or completely incorrect
------------------------------------------------------------------------------------------
```

[0016] No solution was obtained





SpoIIP 125-339

(nps352):[IIPx]
[0001] file import -- *imported: Unmerged (1) Sequence (1)*
[0002] aimless -- *Compl=97.3% $CC_{1/2}$=0.999 R_{meas_all}=0.126 R_{meas_ano}=0.128 Res=1.22-47.68 SpG=P 2 21 21*
[0003] aimless -- *Compl=98.0% $CC_{1/2}$=0.999 R_{meas_all}=0.100 R_{meas_ano}=0.101 Res=1.36-47.68 SpG=P 2 21 21*
[0004] define asymmetric unit contents -- *1 molecule in ASU, Solv=39.8%*
[0007] file import -- *imported: XYZ (4)*
[0008] model from AF2 -- *1 model(s) generated (molrep protocol)*
[0010] phaser MR -- *N_{sol}=1 LLG*
[0018] AF MR modelcraft
[0012] Rosetta model -- *1 model(s)*
[0013] phyre2 model -- *1 model(s)*
[0014] swissmodel -- *1 model(s) ge*

[0018] AF MR modelcraft -- completed

Automatic Model Building with ModelCraft

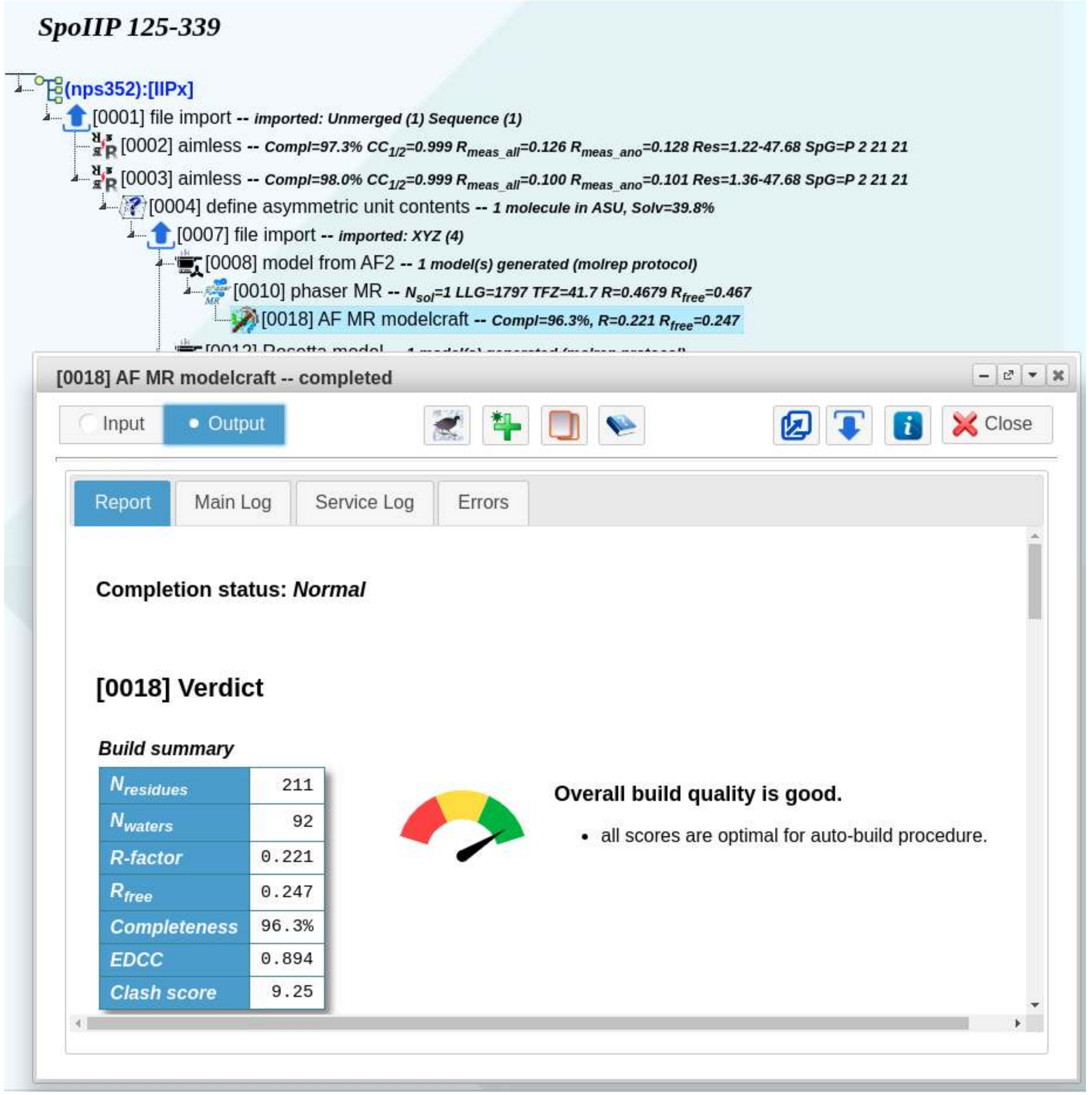
job description: AF MR modelcraft
output id: modelcraft

Structure revision R0010.01: phaser-mr (anom,protein)/xyz,phases

☐ Use twinned refinement

Parameters

Build mode: Full
Maximum number of build cycles: 25
Stop if results do not improve during 4 consequitive cycles



SpoIIP 125-339

(nps352):[IIPx]
[0001] file import -- *imported: Unmerged (1) Sequence (1)*
[0002] aimless -- *Compl=97.3% $CC_{1/2}$=0.999 R_{meas_all}=0.126 R_{meas_ano}=0.128 Res=1.22-47.68 SpG=P 2 21 21*
[0003] aimless -- *Compl=98.0% $CC_{1/2}$=0.999 R_{meas_all}=0.100 R_{meas_ano}=0.101 Res=1.36-47.68 SpG=P 2 21 21*
[0004] define asymmetric unit contents -- *1 molecule in ASU, Solv=39.8%*
[0007] file import -- *imported: XYZ (4)*
[0008] model from AF2 -- *1 model(s) generated (molrep protocol)*
[0010] phaser MR -- *N_{sol}=1 LLG=1797 TFZ=41.7 R=0.4679 R_{free}=0.467*
[0018] AF MR modelcraft -- *Compl=96.3%, R=0.221 R_{free}=0.247*

[0018] AF MR modelcraft -- completed

Completion status: *Normal*

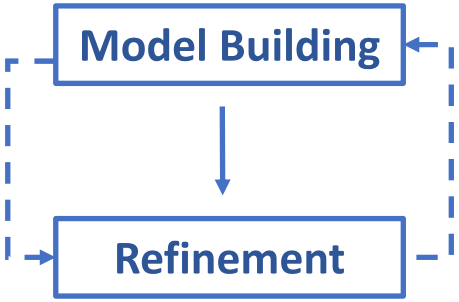
[0018] Verdict

Build summary

$N_{residues}$	211
N_{waters}	92
R-factor	0.221
R_{free}	0.247
Completeness	96.3%
EDCC	0.894
Clash score	9.25

Overall build quality is good.

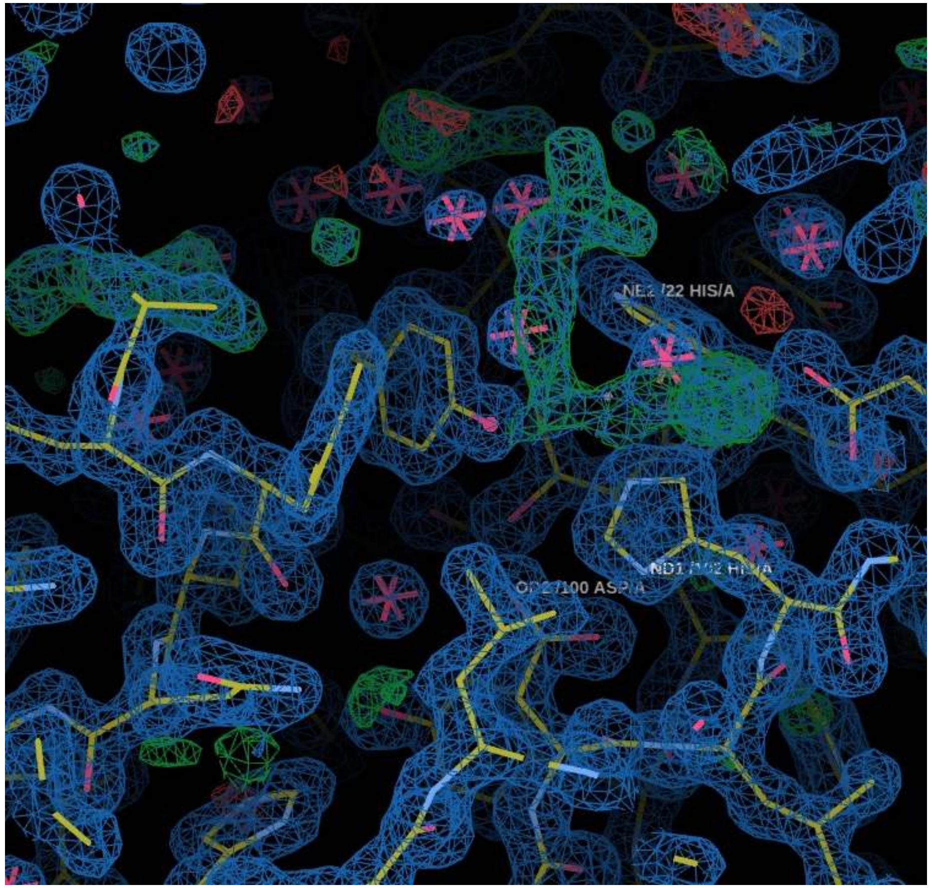
- all scores are optimal for auto-build procedure.

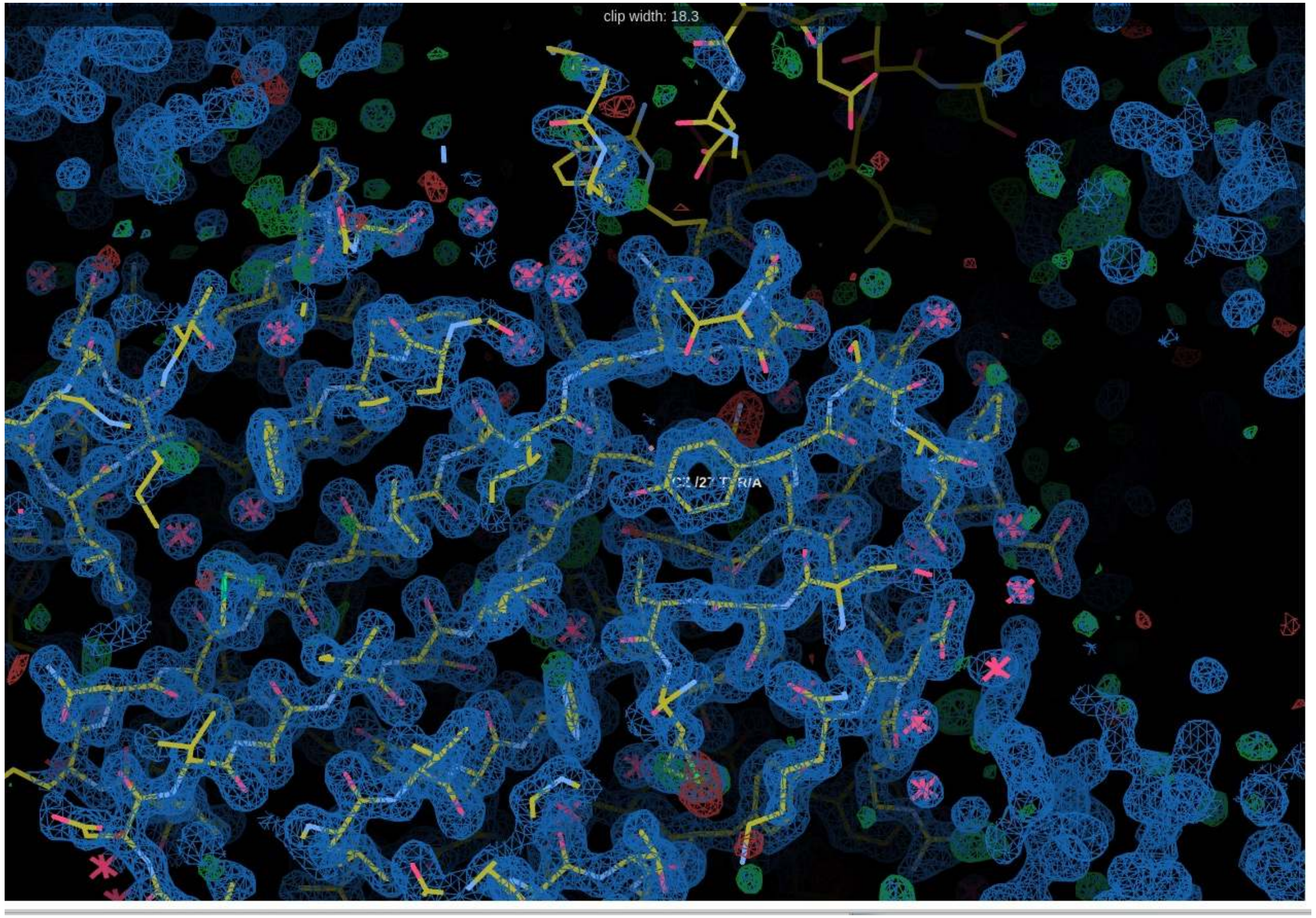


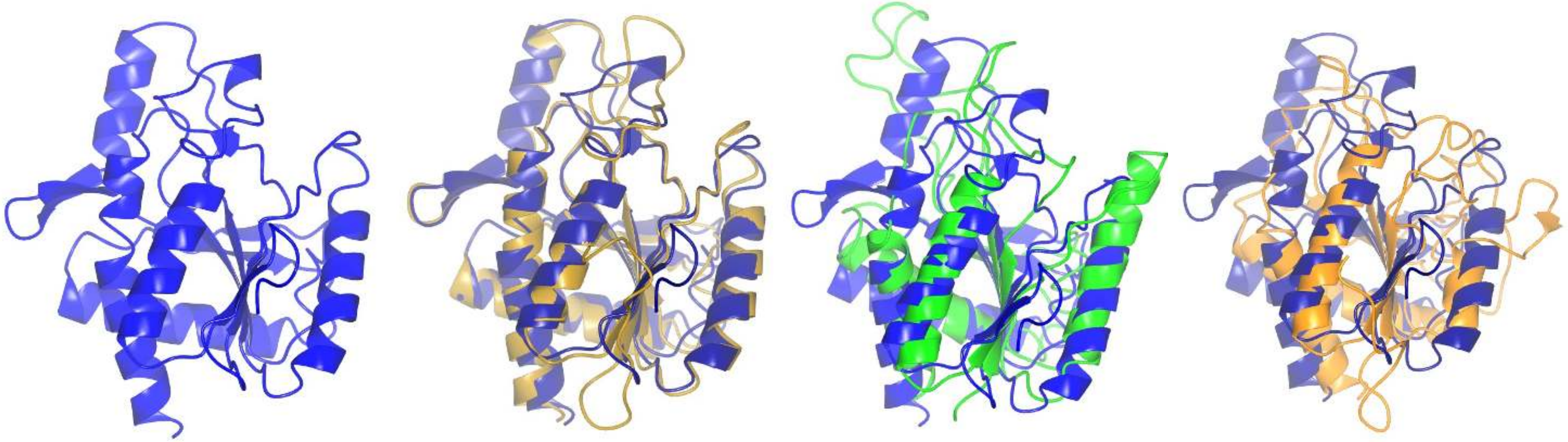
Model Building

Refinement

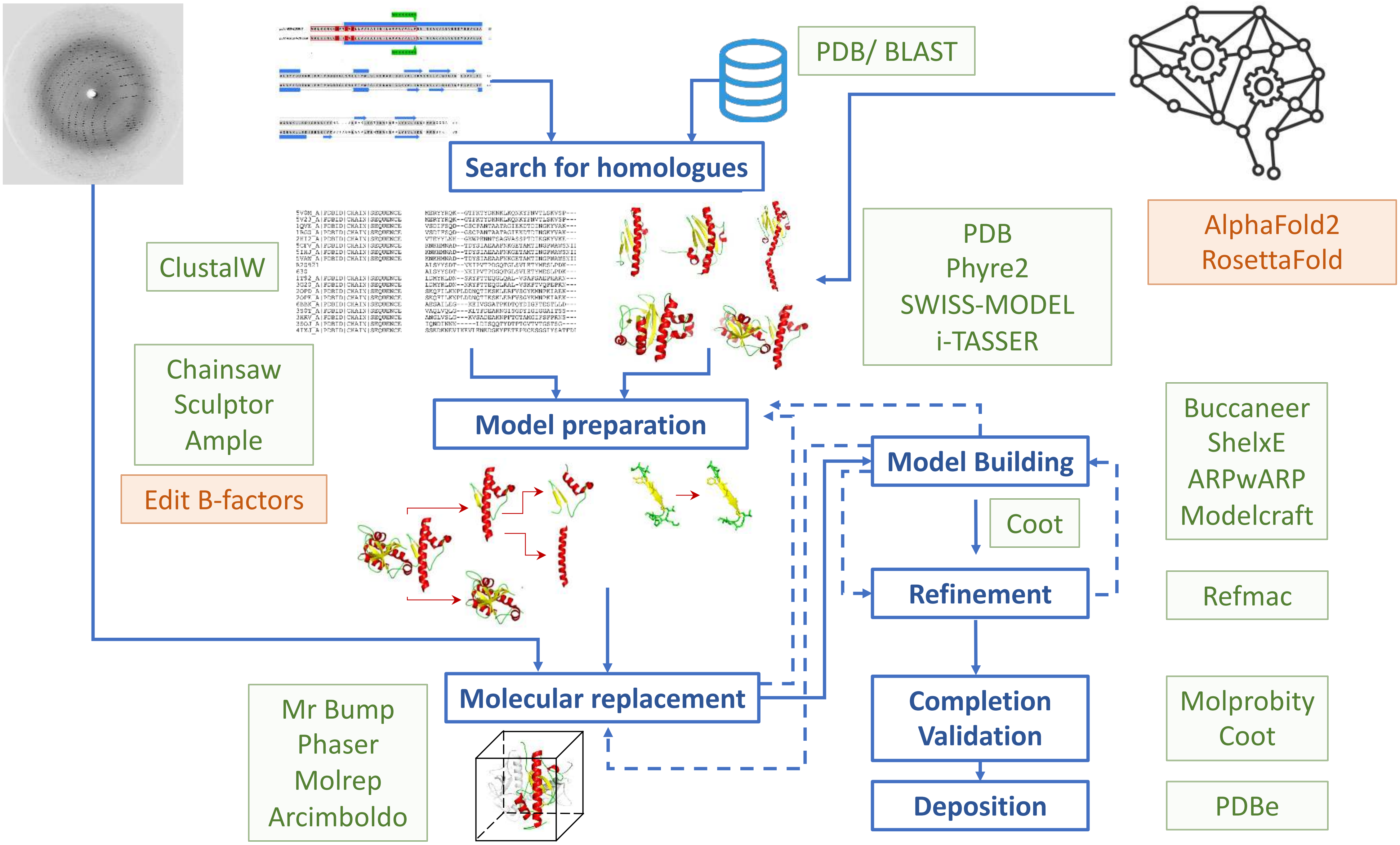
Modelcraft







AlphaFold2	RosettaFold	SWISS-MODEL	Phyre2
Phaser: TFZ – 42.1 R = 29% Rfree = 32%	Phaser: TFZ – 6.8 No solution	Phaser: TFZ – 3.6 No solution (seq ID <20%)	Phaser: TFZ – 3.7 No solution (seq ID <20%)
Modelcraft: 211 aa R = 22% Rfree = 25%	-	-	-
Refmac: R = 21% Rfree = 23%	-	-	-



PDB/ BLAST

Search for homologues

AlphaFold2
RosettaFold

ClustalW

PDB
Phyre2
SWISS-MODEL
i-TASSER

Chainsaw
Sculptor
Ample

Edit B-factors

Model preparation

Model Building

Buccaneer
ShelxE
ARPwARP
Modelcraft

Coot

Refinement

Refmac

Molecular replacement

Mr Bump
Phaser
Molrep
Arcimboldo

Completion
Validation

Molprobity
Coot

Deposition

PDBe

SALGADO LAB

Adam Crawshaw
Carys Davies
Anna Barwinska-Sendra
Paola Lanzoni
Marcin Dembek
Abbie Kelly
Gilly Wang
Charlotte Roughton
Victoria Burge

Newcastle University

Newcastle Structural Biology Laboratory

Arnaud Baslé

diamond

Kamel El Omari
Armin Wagner

Imperial College London

Neil Fairweather
Ed Couchman

The University Of Sheffield.

Rob Fagan
Joe Kirk
Per Bullough
Oishik Banerji
Jason Wilson

University of Glasgow

Gill Douce
Filipa Vaz

BBSRC Doctoral Training Partnerships

MRC Medical Research Council

wellcome

MoSMed Molecular Sciences for Medicine