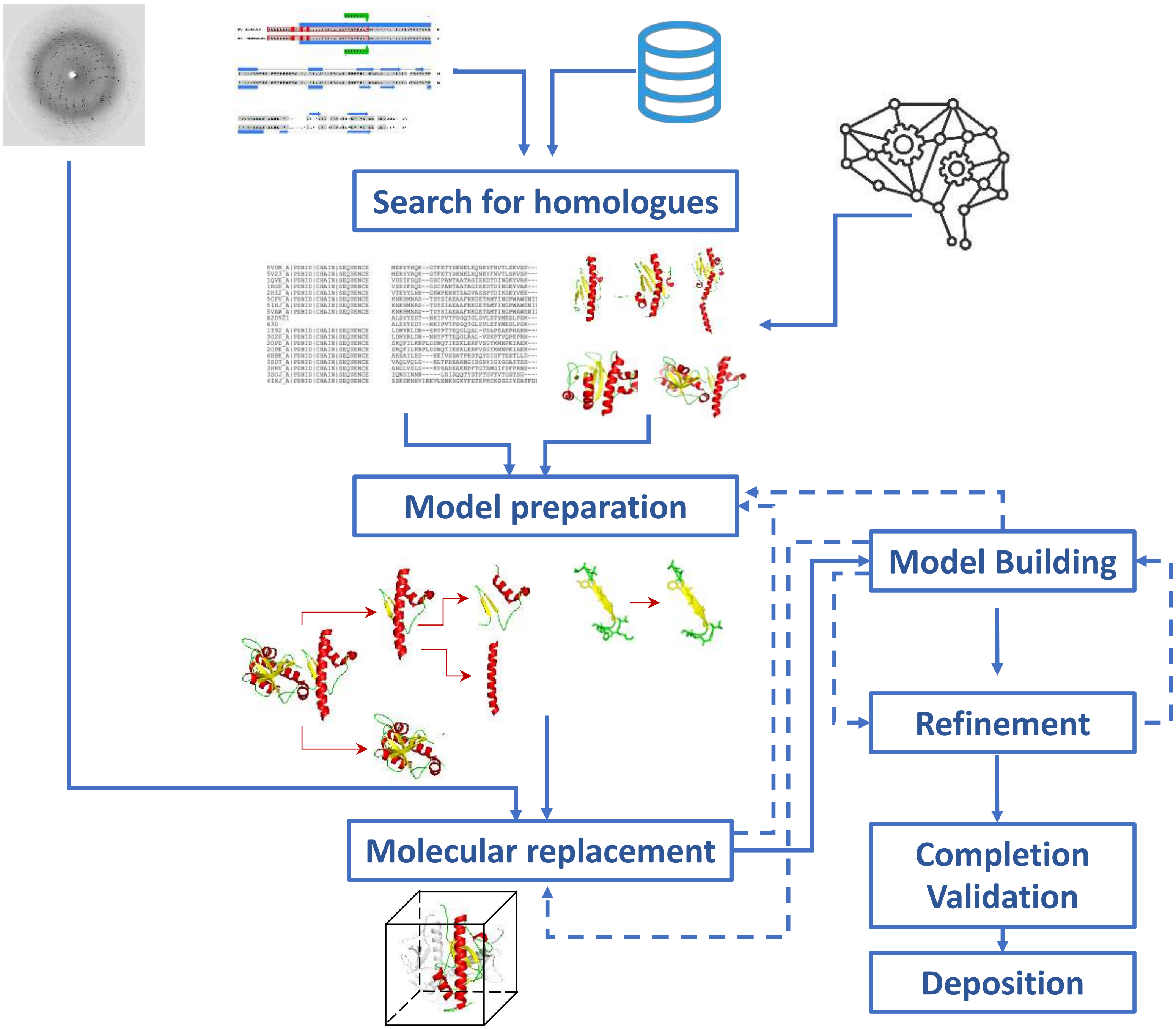
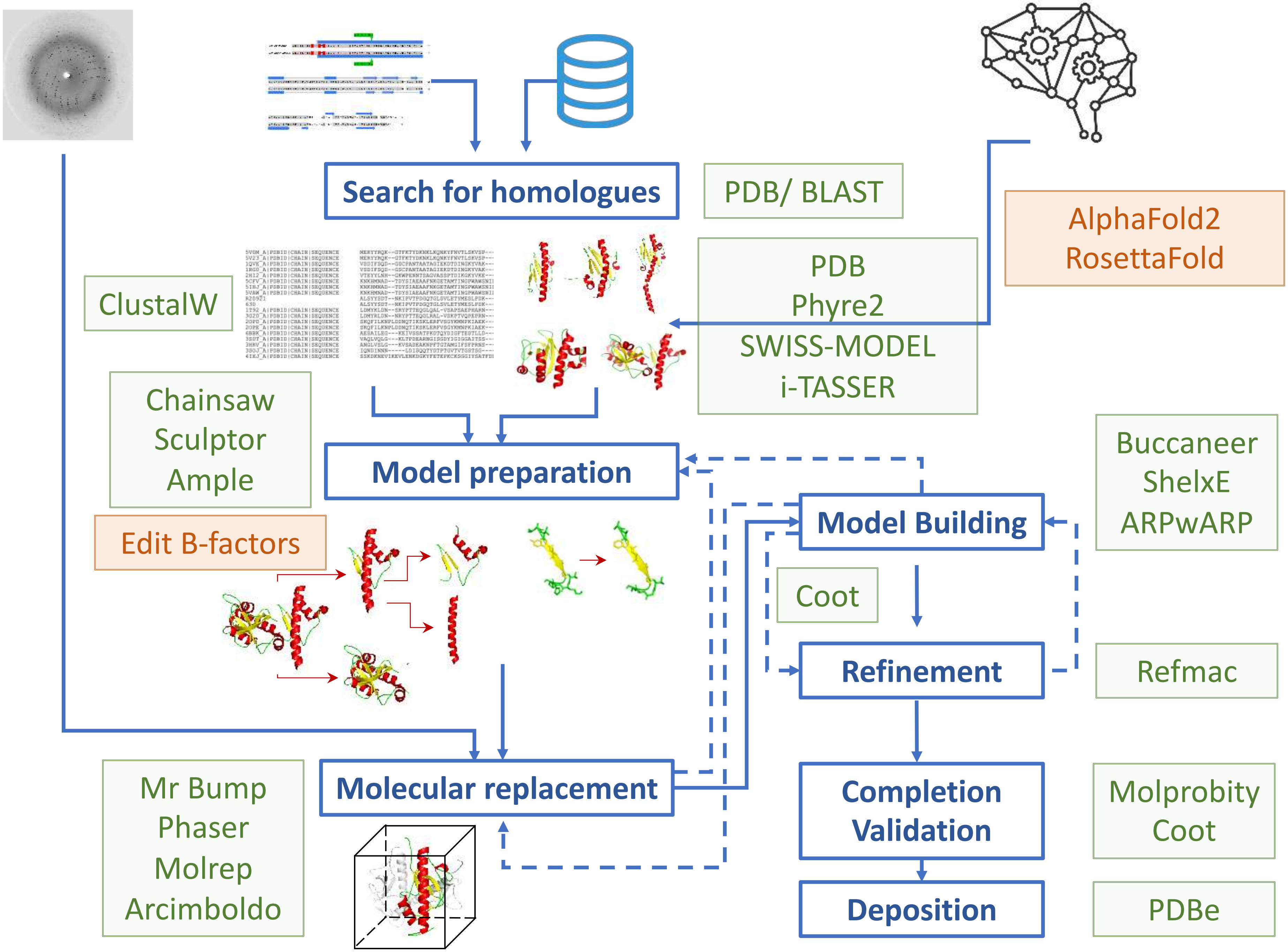
MR case studies

Paula S. Salgado
DLS/CCP4 Workshop 2021


Search for homologues
Model preparation
Model Building
Refinement
Molecular replacement
Completion
Validation
Deposition



Search for homologues

PDB/ BLAST

AlphaFold2
RosettaFold

ClustalW

PDB
Phyre2
SWISS-MODEL
i-TASSER

Chainsaw
Sculptor
Ample

Model preparation

Edit B-factors

Model Building

Buccaneer
ShelxE
ARPwARP

Coot

Refinement

Refmac

Molecular replacement

Mr Bump
Phaser
Molrep
Arcimboldo

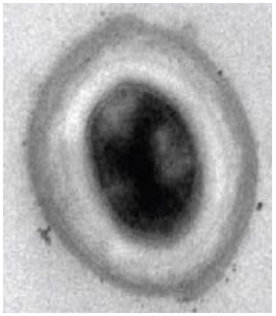
**Completion
Validation**

Molprobity
Coot

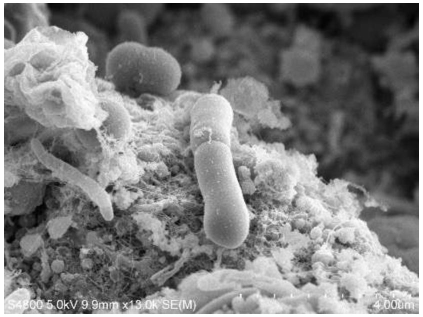
Deposition

PDBe

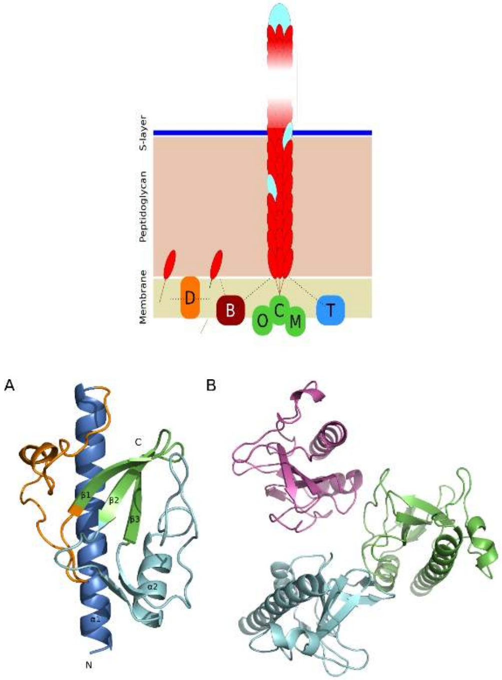
Case studies from *C. difficile* proteins



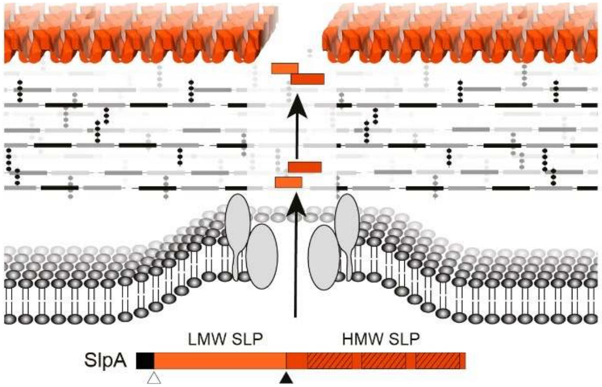
+ Antibiotics



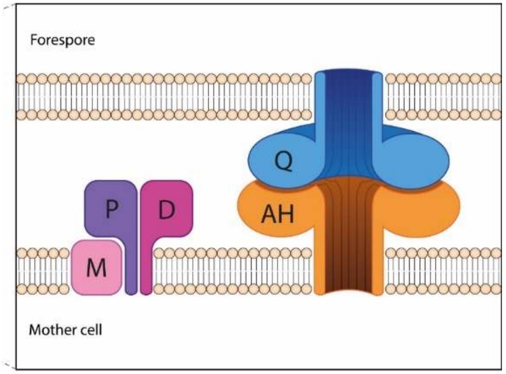
Inflammatory Complications



Type IV pili

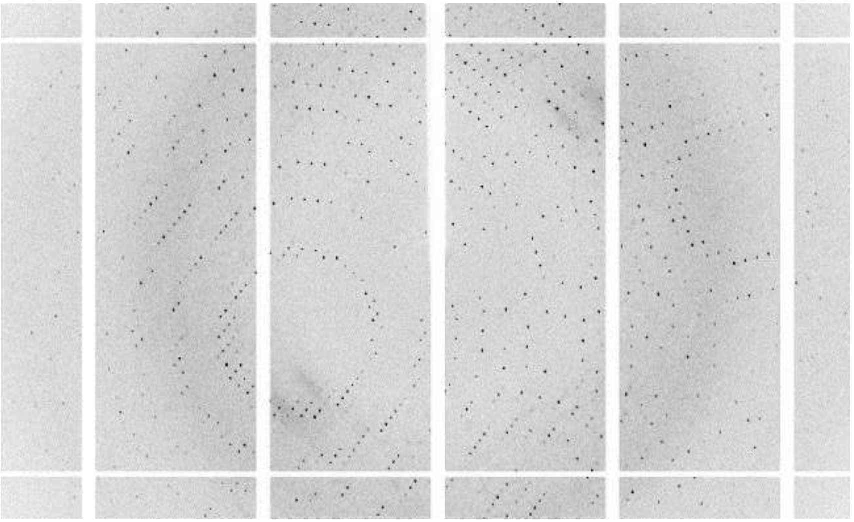


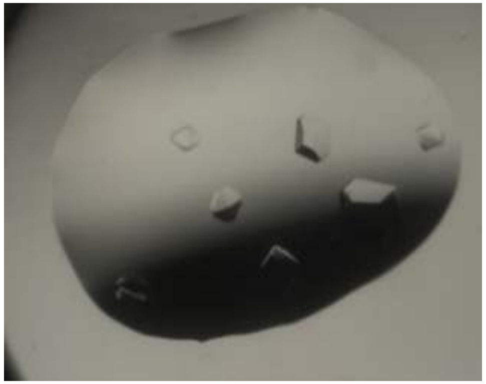
S-layer



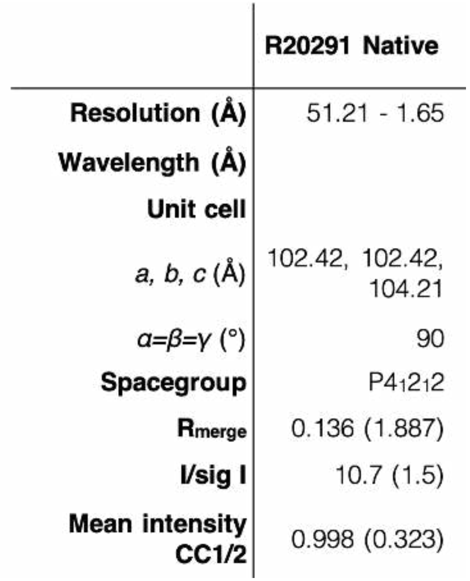
Sporulation/ engulfment

Example 1: CdPilA1

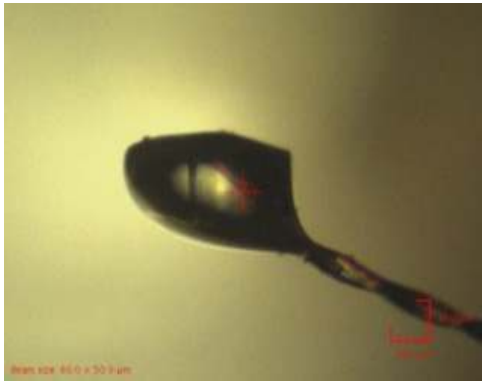


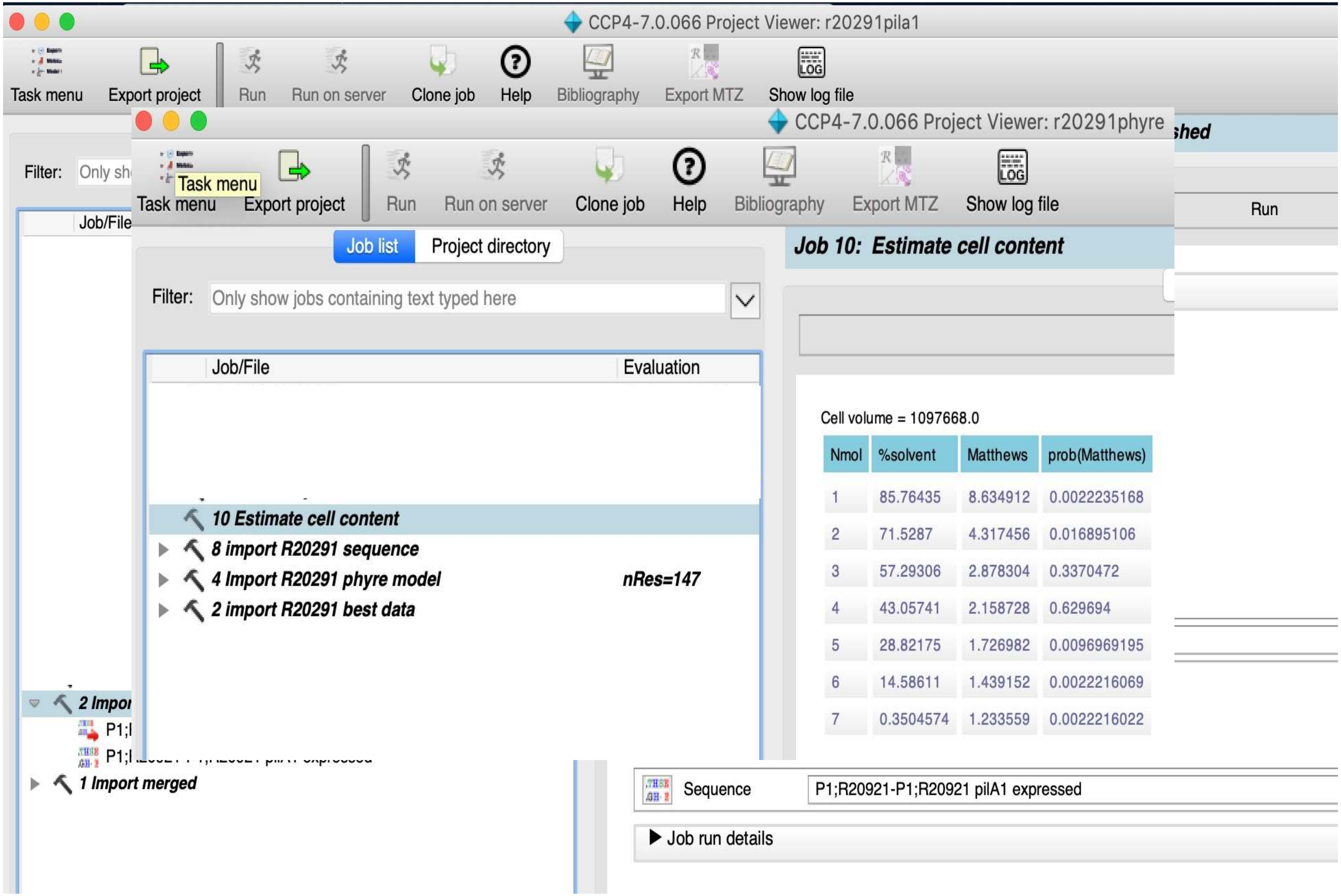


A. Crawshaw, C. Davies



	R20291 Native
Resolution (Å)	51.21 - 1.65
Wavelength (Å)	
Unit cell	
a, b, c (Å)	102.42, 102.42, 104.21
$\alpha=\beta=\gamma$ (°)	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)





10 Estimate cell content

8 import R20291 sequence

4 Import R20291 phyre model — nRes=147

2 import R20291 best data

2 Impor

P1;l

P1;l

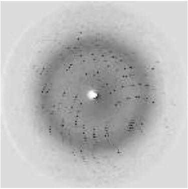
1 Import merged

Job 10: Estimate cell content

Cell volume = 1097668.0

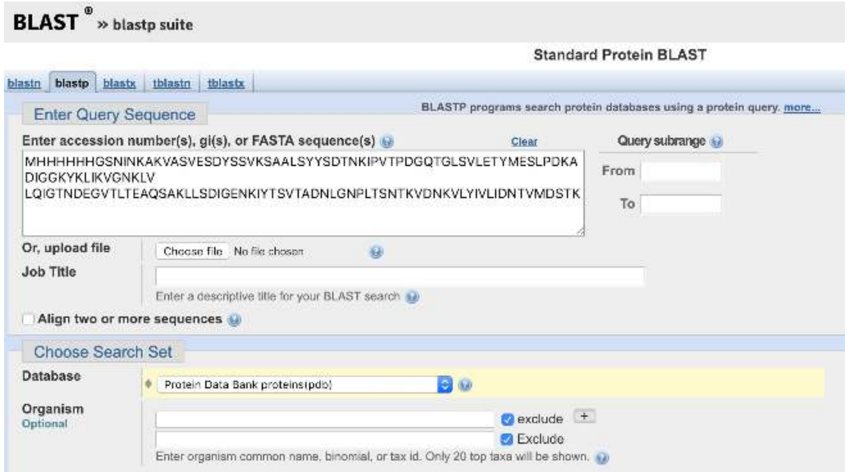
Nmol	%solvent	Matthews	prob(Matthews)
1	85.76435	8.634912	0.0022235168
2	71.5287	4.317456	0.016895106
3	57.29306	2.878304	0.3370472
4	43.05741	2.158728	0.629694
5	28.82175	1.726982	0.0096969195
6	14.58611	1.439152	0.0022216069
7	0.3504574	1.233559	0.0022216022

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) Clear

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

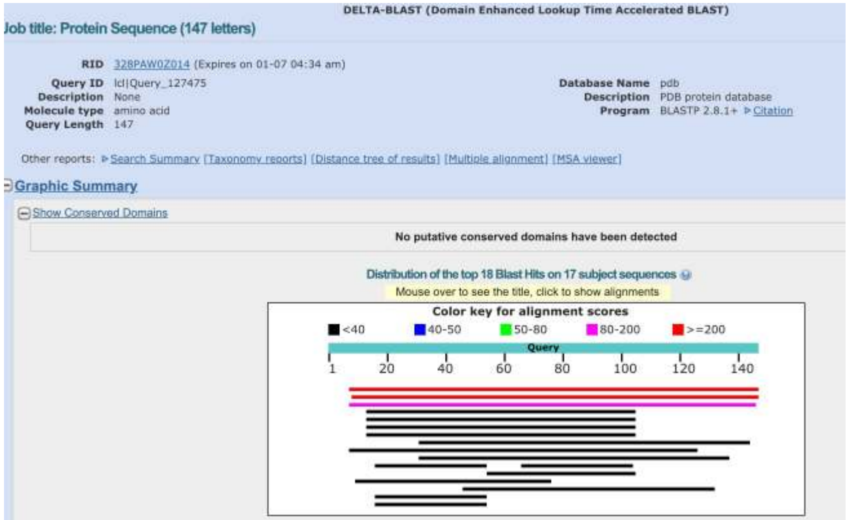
Database: Protein Data Bank proteins(pdb)

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]

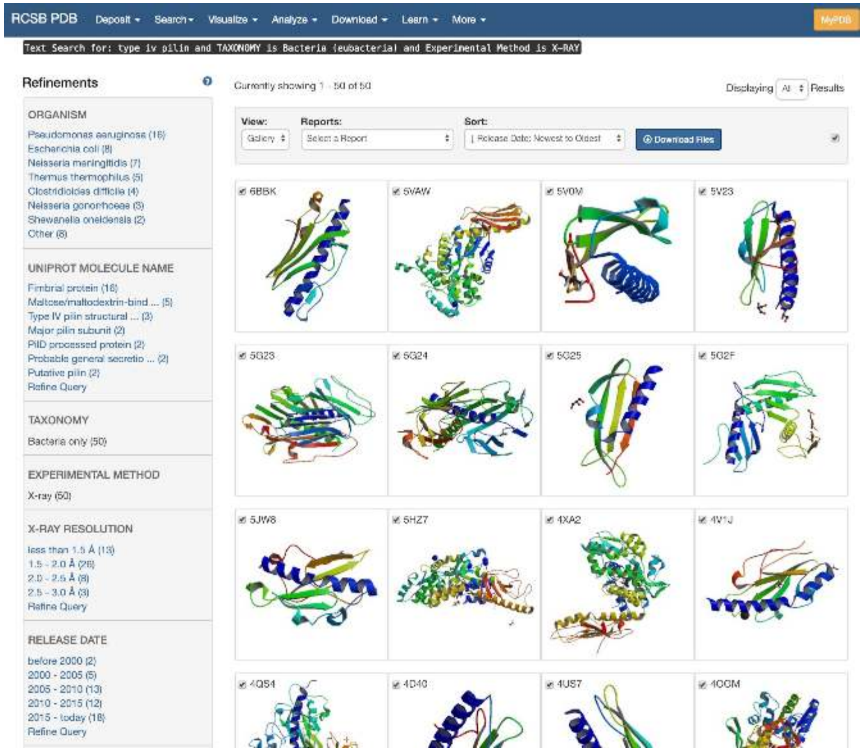
Graphic Summary

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



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

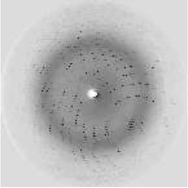
Currently showing 1 - 50 of 50

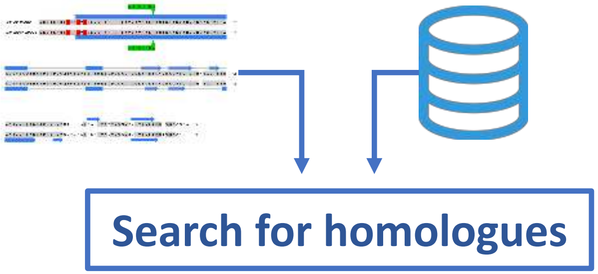
Displaying All Results

View: Gallery

Reports: Select a Report

Sort: Release Date: Newest to Oldest

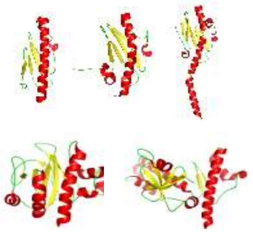




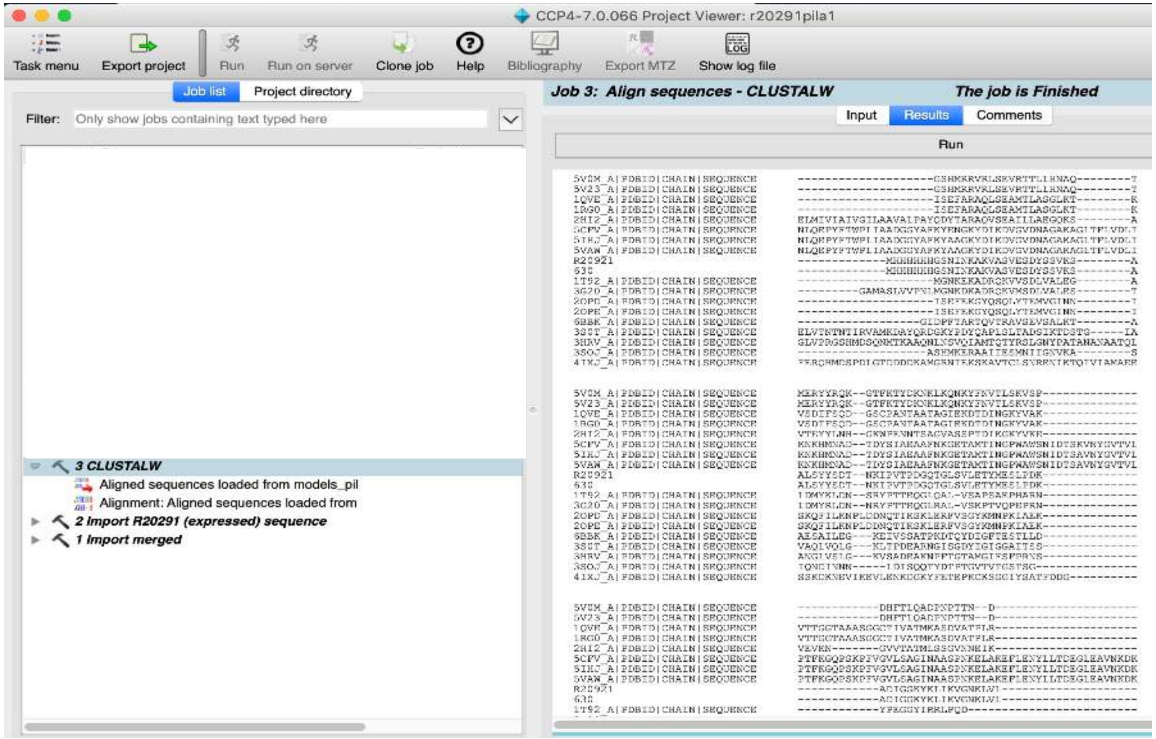
Search for homologues

PDB/ BLAST

ClustalW



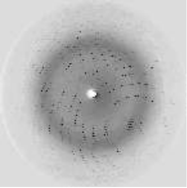
PDB

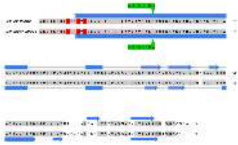


Job 3: Align sequences - CLUSTALW

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

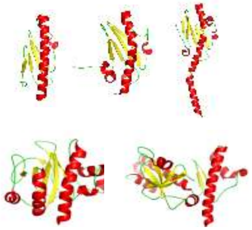




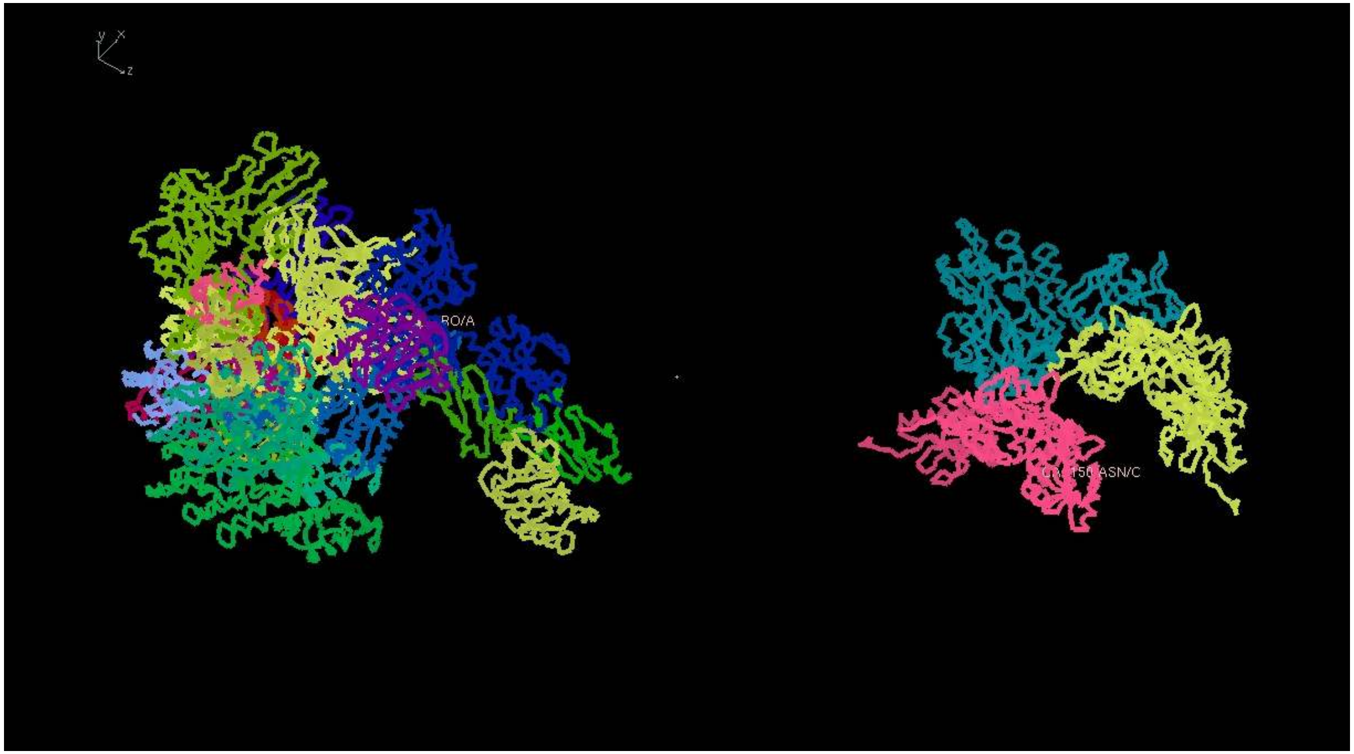
Search for homologues

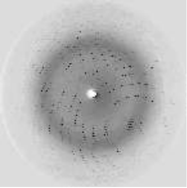
ClustalW

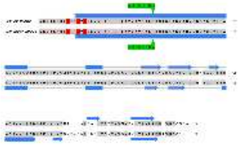
```
5VOM_A|PDBID|CHAIN|SEQUENCE      MERYYRQK--GTFKTYDKNKLKQNKYFNVTLSKVSP---
5V23_A|PDBID|CHAIN|SEQUENCE      MERYYRQK--GTFKTYDKNKLKQNKYFNVTLSKVSP---
1QVE_A|PDBID|CHAIN|SEQUENCE      VSDIFSQD--GSCPANTAATAGIEKDTDINGKYVAK---
1RGD_A|PDBID|CHAIN|SEQUENCE      VSDIFSQD--GSCPANTAATAGIEKDTDINGKYVAK---
2H12_A|PDBID|CHAIN|SEQUENCE      VTEYYLNH--GKWPENNTSAGVASSPTDIKGKYVKE---
5CFV_A|PDBID|CHAIN|SEQUENCE      KNKHMNAD--TDYSIAEAAFNKGETAMTINGPWAWSNI
5INJ_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---
3SUT_A|PDBID|CHAIN|SEQUENCE      VAQLVQLG---KLTPDEARNGISGDYIGIGGAITSS---
3HNV_A|PDBID|CHAIN|SEQUENCE      ANGLVSLG---KVSADEAKNPFTGTAMGIFSFPRNS---
3SOJ_A|PDBID|CHAIN|SEQUENCE      IQNDINNN-----LDISQQTYDTPTGVTVTGSTSG---
4IKJ_A|PDBID|CHAIN|SEQUENCE      SSKDKNEVIKEVLENKDGKYFETEPKCKSGGIYSATFDI
```



PDB

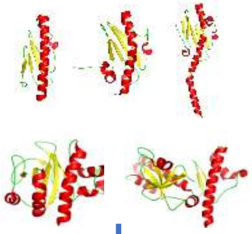






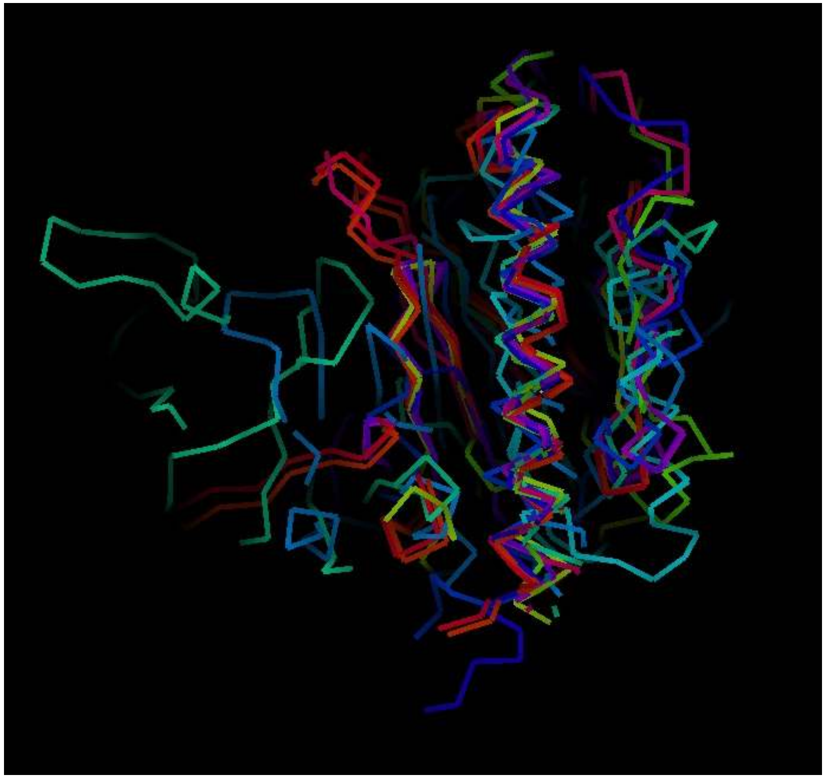
Search for homologues

ClustalW



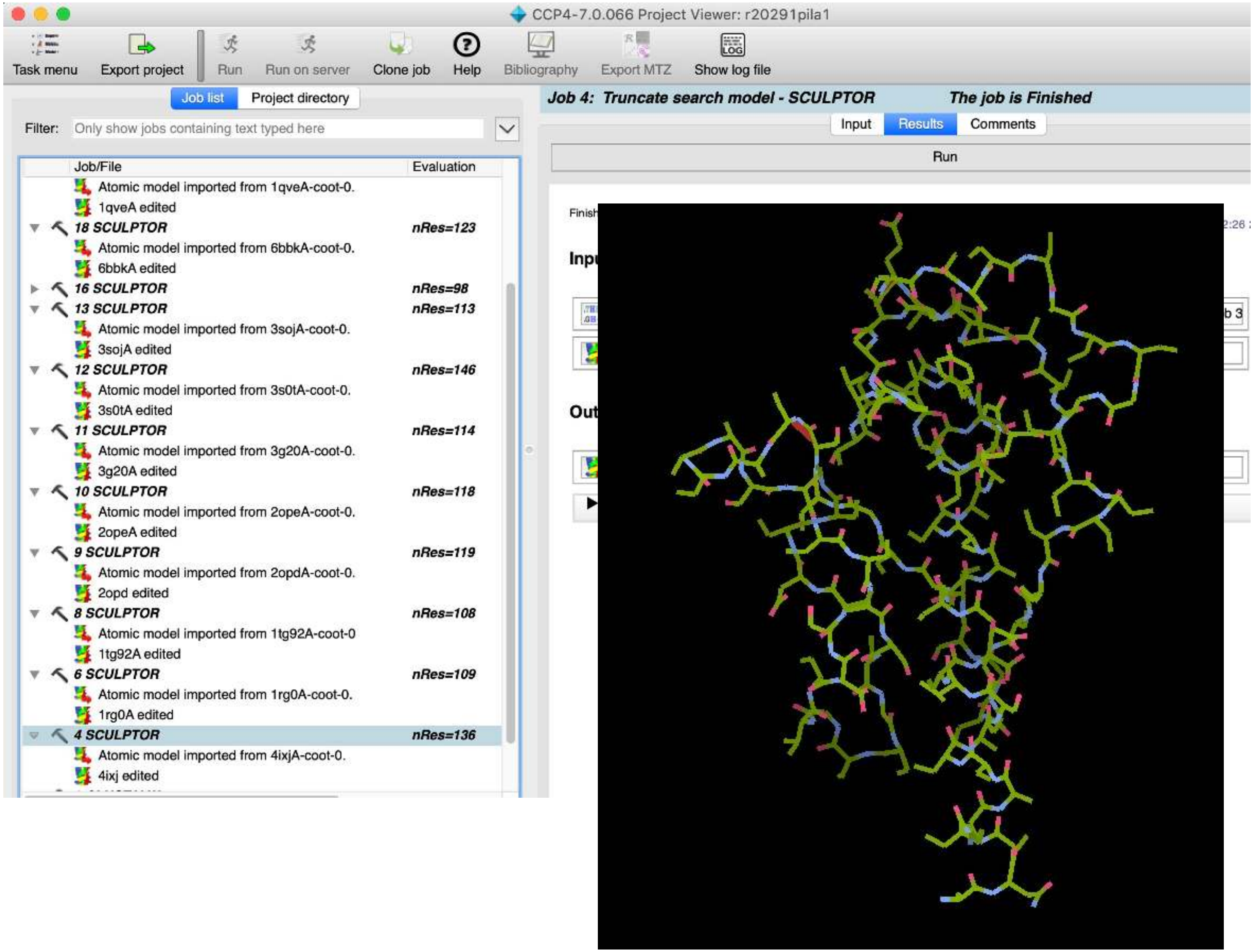
PDB

Model preparation



Sculptor

Model preparation

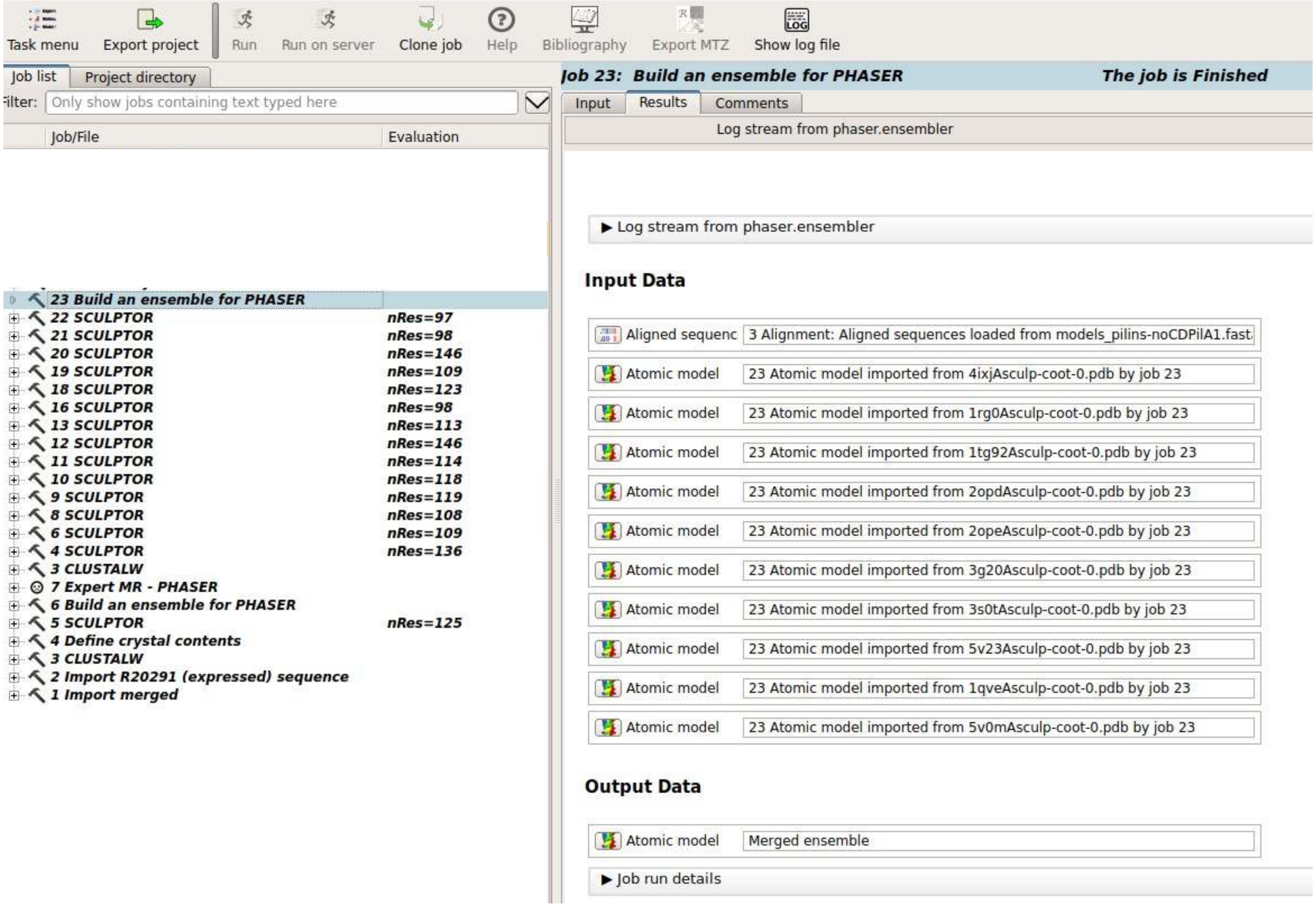


Filter: Only show jobs containing text typed here

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

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	

Model preparation



Filter: Only show jobs containing text typed here

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	

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

Log stream from phaser.ensembler

▶ Log stream from phaser.ensembler

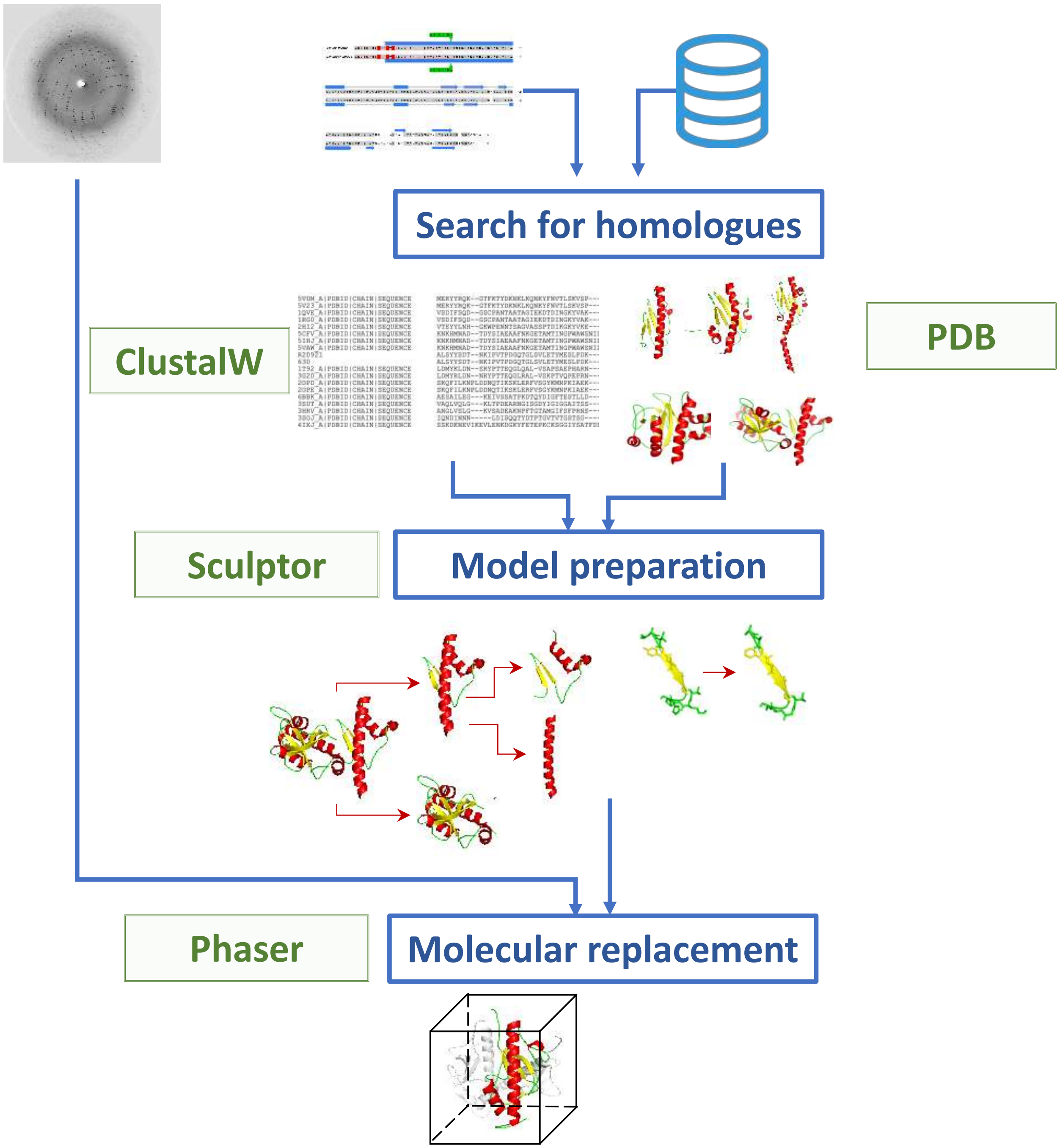
Input Data

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

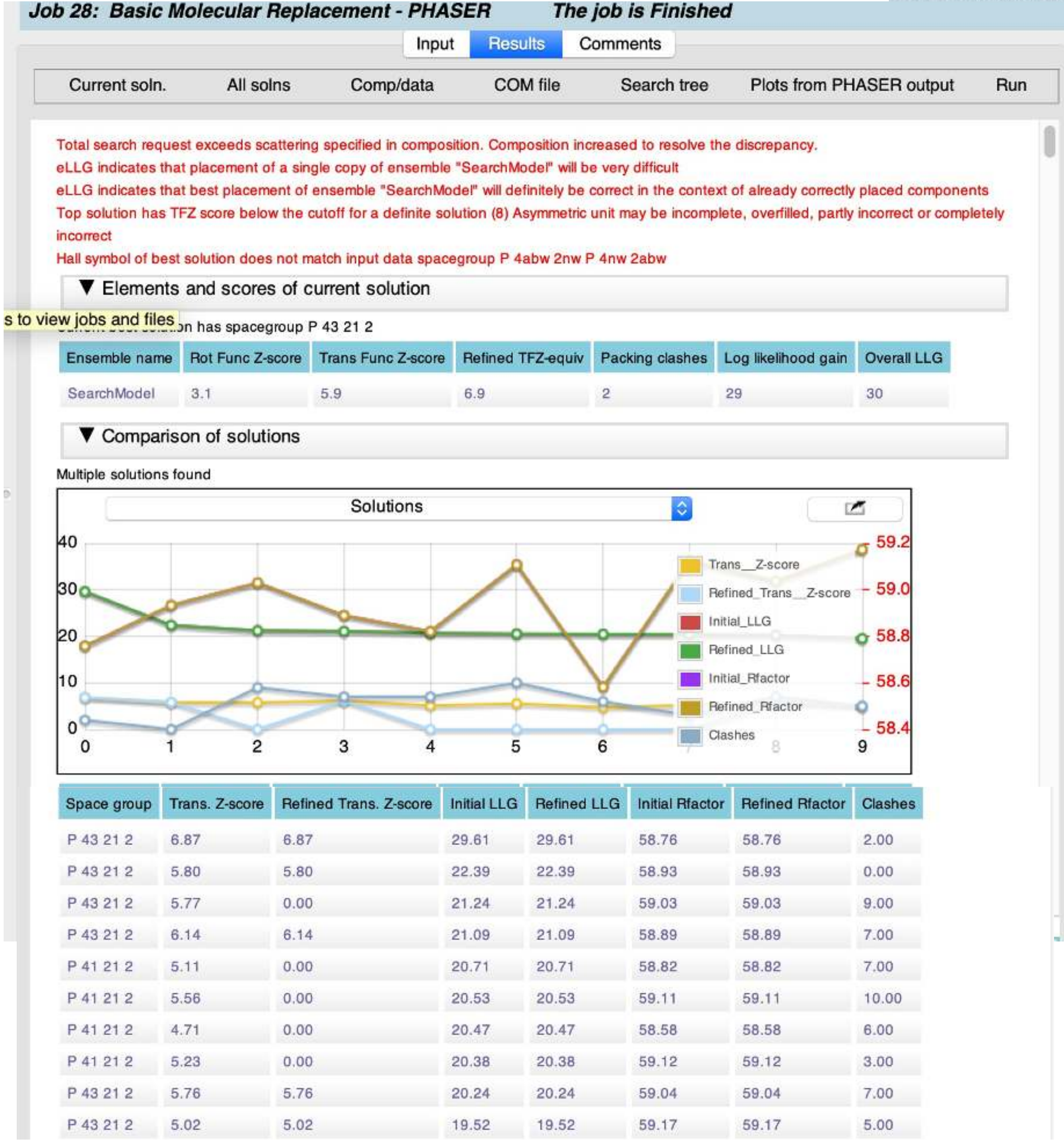
Sculptor

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

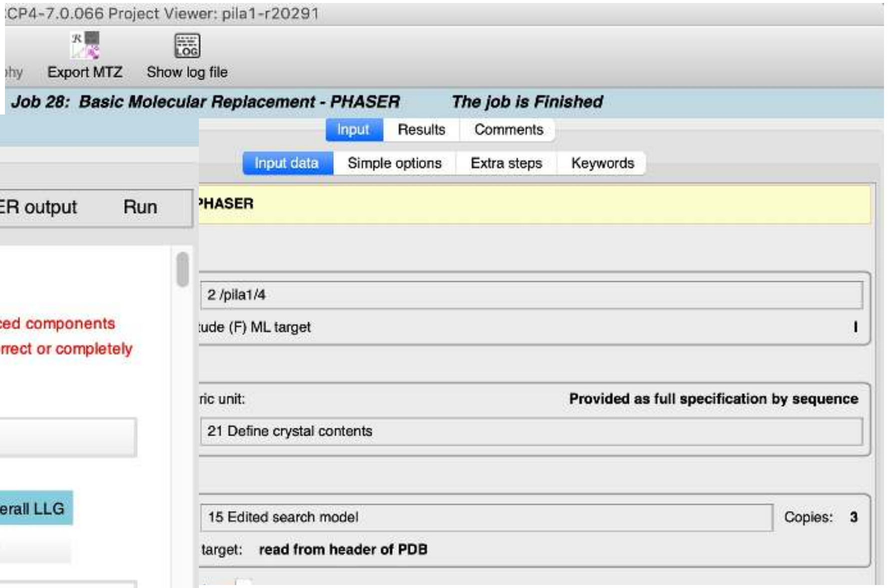
on has spacegroup P 43 21 2

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*

HASER

2 /pila1/4

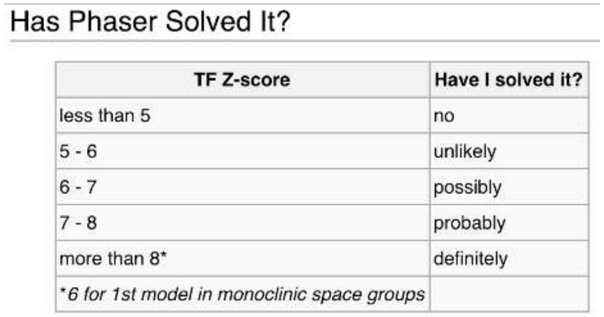
tude (F) ML target

ric unit: **Provided as full specification by sequence**

21 Define crystal contents

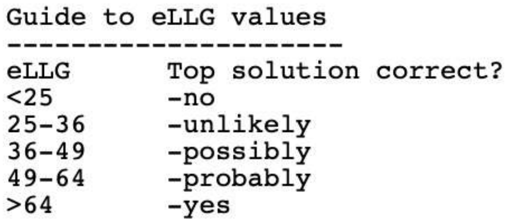
15 Edited search model Copies: **3**

target: **read from header of PDB**

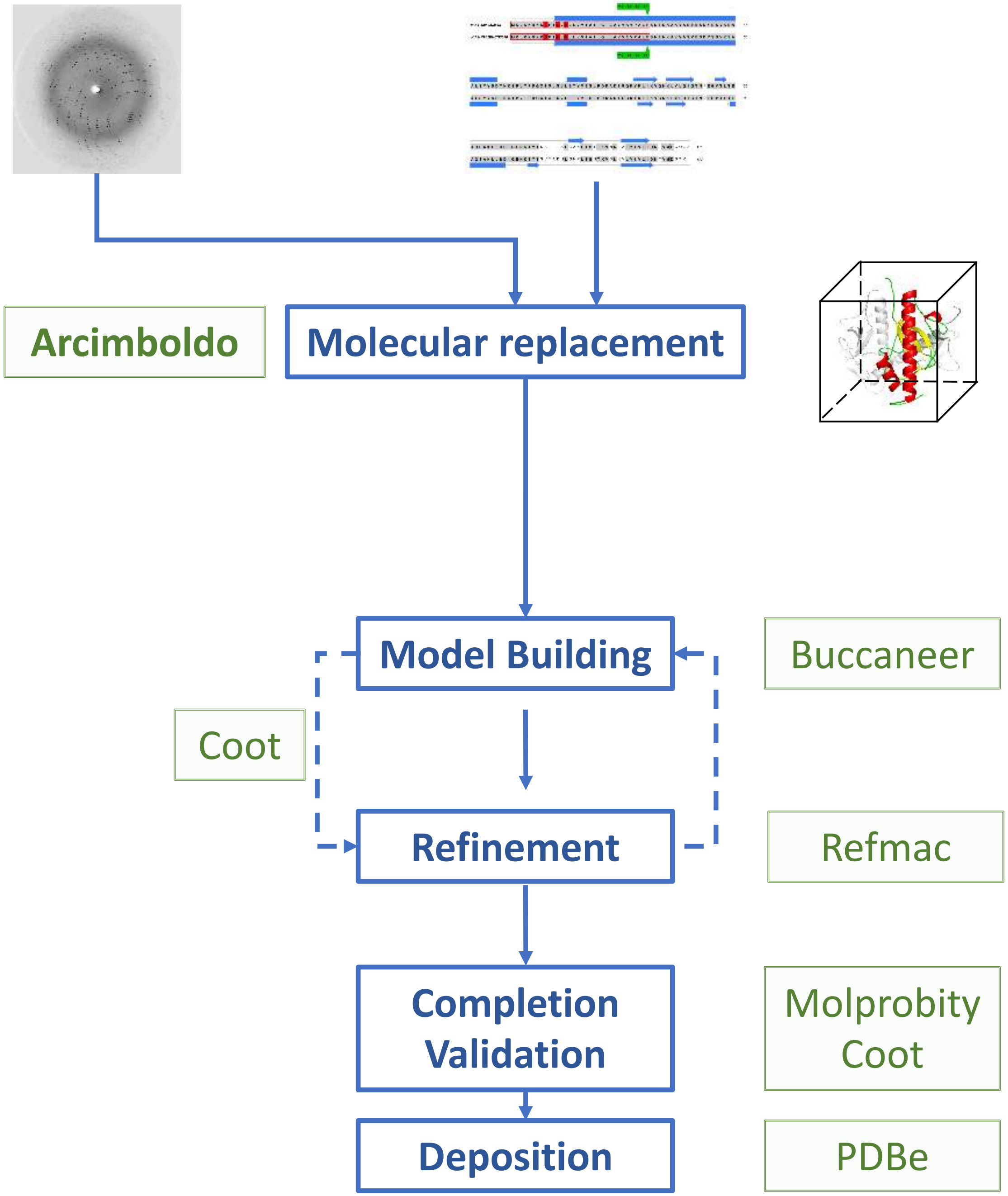


Has Phaser Solved It?

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

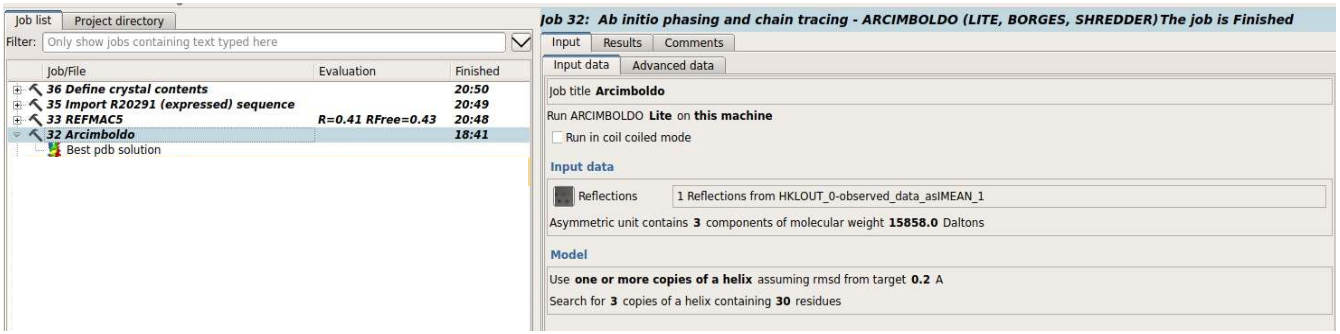
Molprobity
Coot

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

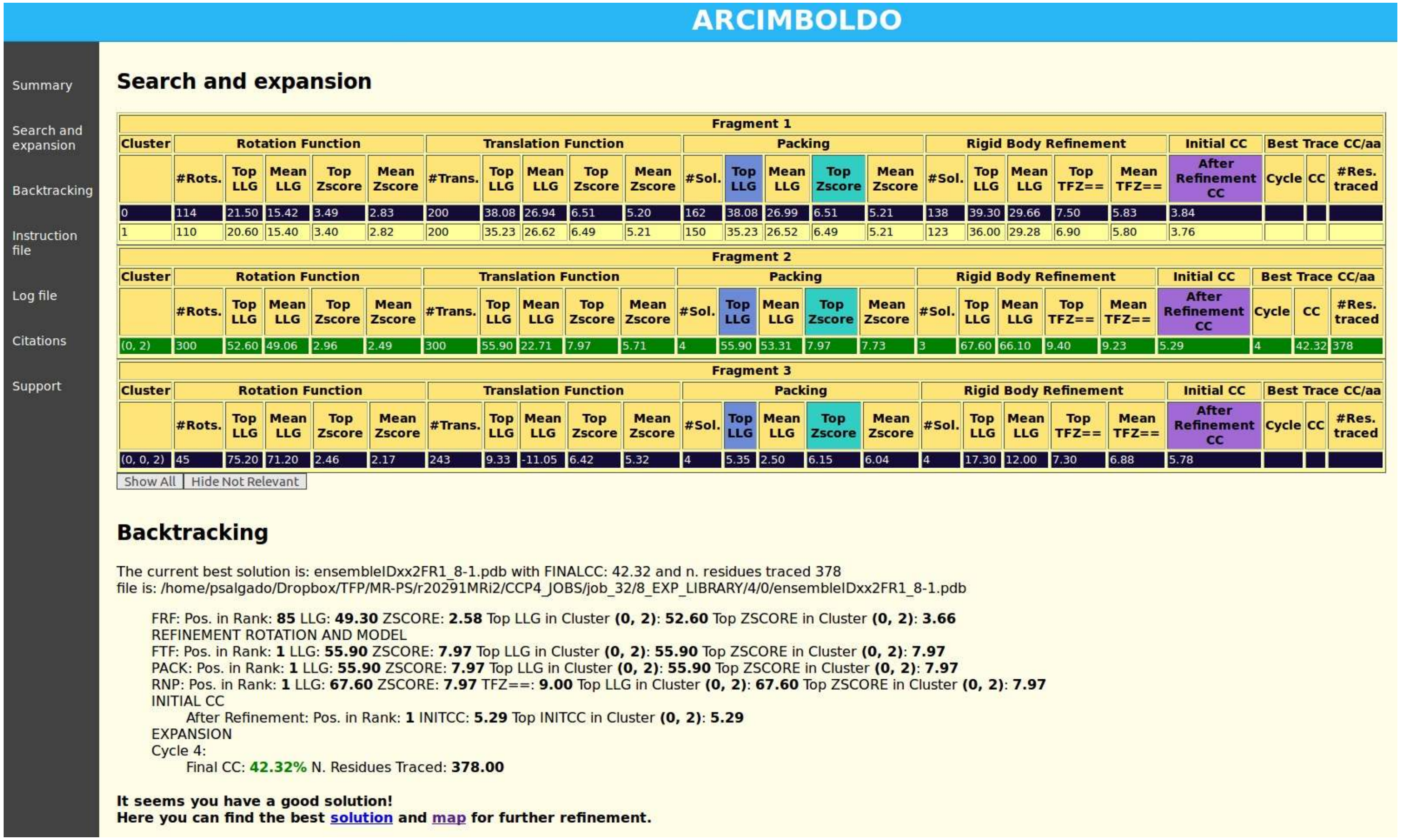
Reflections: 1 Reflections from HKLOUT_0-observed_data_asIMEAN_1

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

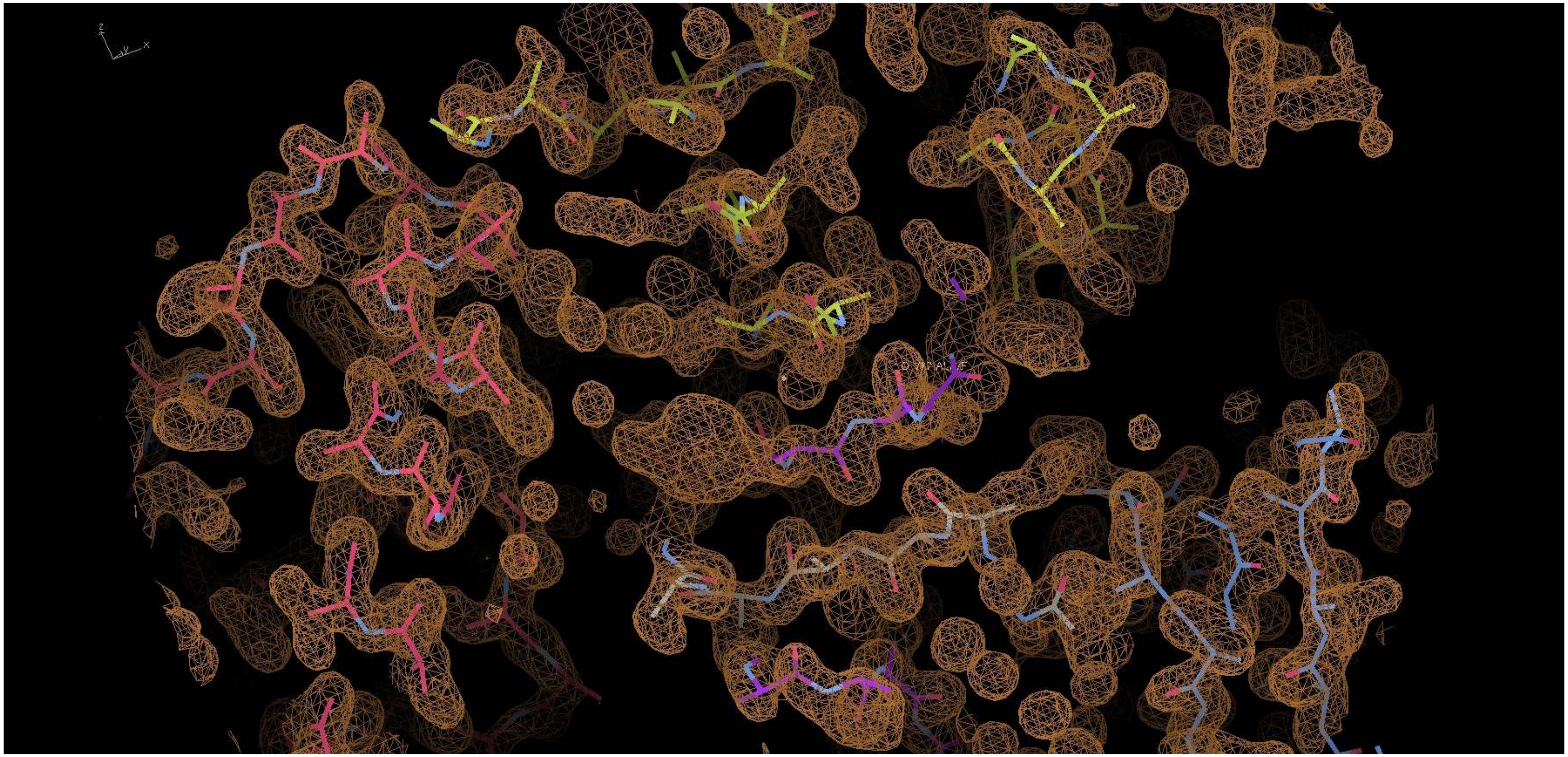
The current best solution is: ensembleIDxx2FR1_8-1.pdb with FINALCC: 42.32 and n. residues traced 378
file is: /home/psalgado/Dropbox/TFP/MR-PS/r20291MRi2/CCP4_JOBS/job_32/8_EXP_LIBRARY/4/0/ensembleIDxx2FR1_8-1.pdb

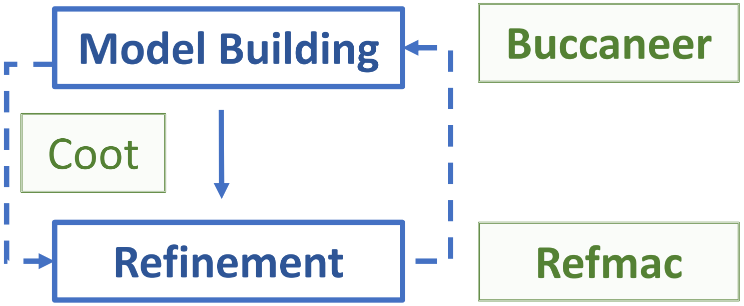
FRF: Pos. in Rank: **85** LLG: **49.30** ZSCORE: **2.58** Top LLG in Cluster **(0, 2)**: **52.60** Top ZSCORE in Cluster **(0, 2)**: **3.66**
REFINEMENT ROTATION AND MODEL
FTF: Pos. in Rank: **1** LLG: **55.90** ZSCORE: **7.97** Top LLG in Cluster **(0, 2)**: **55.90** Top ZSCORE in Cluster **(0, 2)**: **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





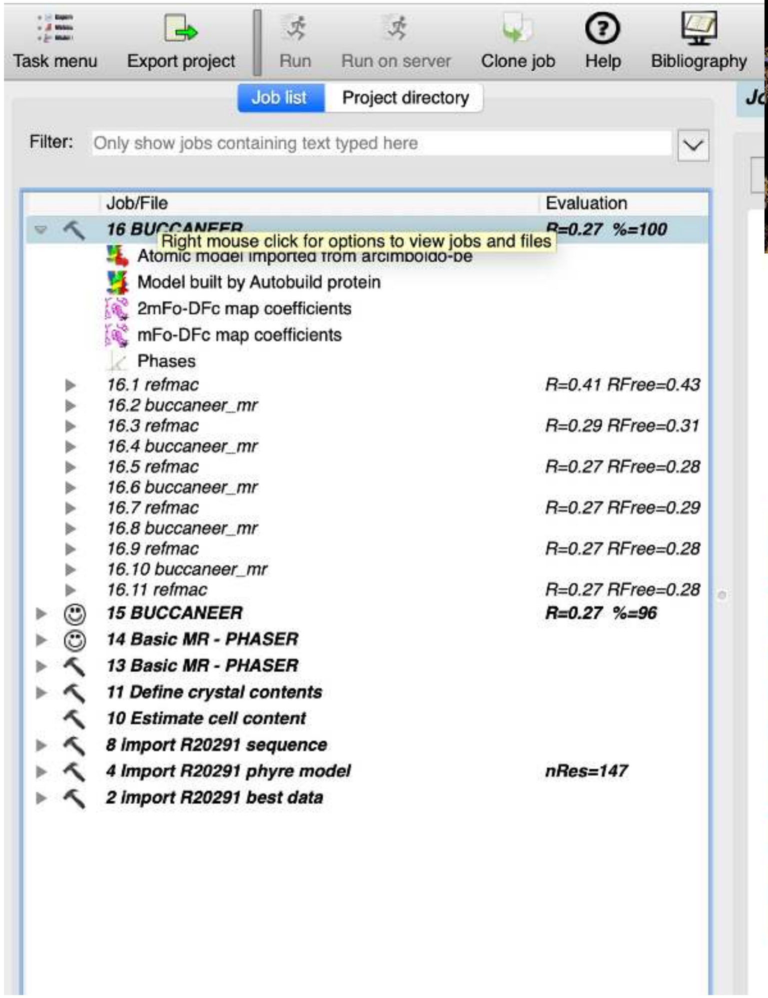
Model Building

Buccaneer

Coot

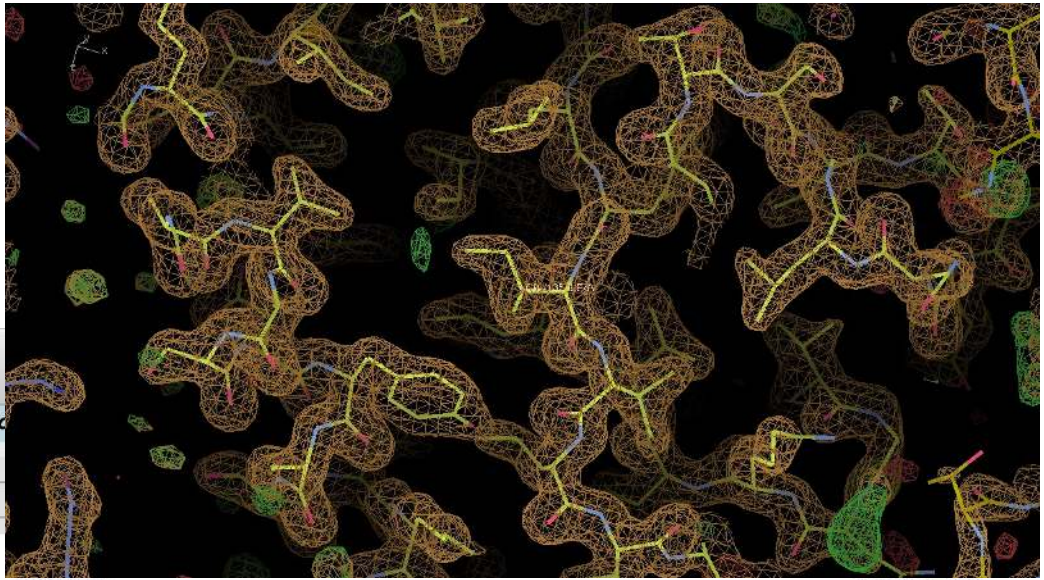
Refinement

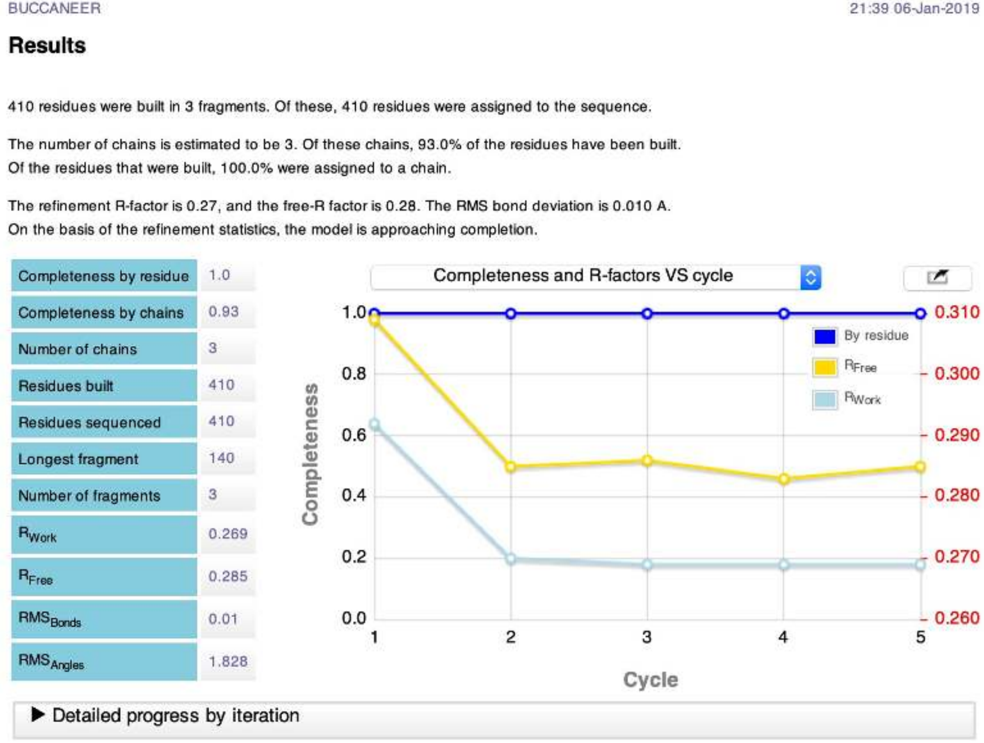
Refmac



Right mouse click for options to view jobs and files

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	





BUCCANEER 21:39 06-Jan-2019

Results

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

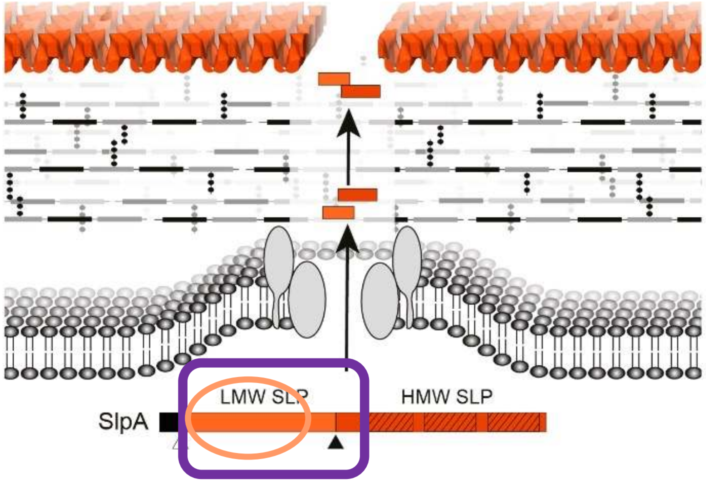
The number of chains is estimated to be 3. Of these chains, 93.0% of the residues have been built. Of the residues that were built, 100.0% were assigned to a chain.

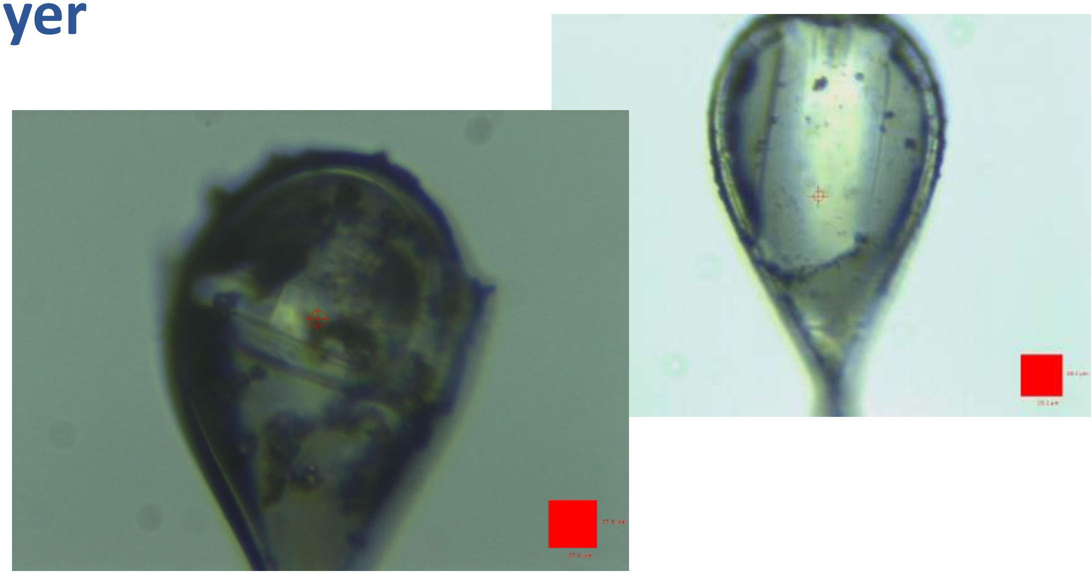
The refinement R-factor is 0.27, and the free-R factor is 0.28. The RMS bond deviation is 0.010 A. On the basis of the refinement statistics, the model is approaching completion.

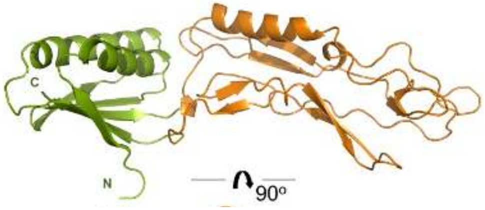
Completeness by residue	1.0
Completeness by chains	0.93
Number of chains	3
Residues built	410
Residues sequenced	410
Longest fragment	140
Number of fragments	3
R_{Work}	0.269
R_{Free}	0.285
RMS_{Bonds}	0.01
RMS_{Angles}	1.828

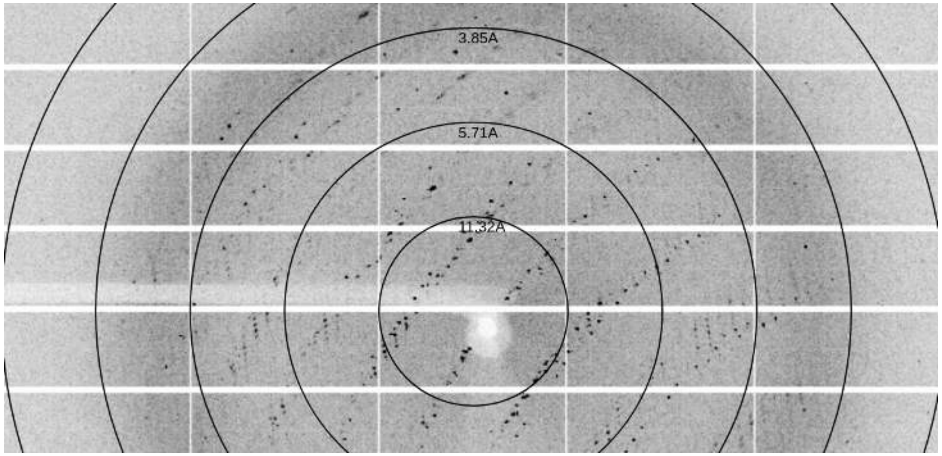
▶ Detailed progress by iteration

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

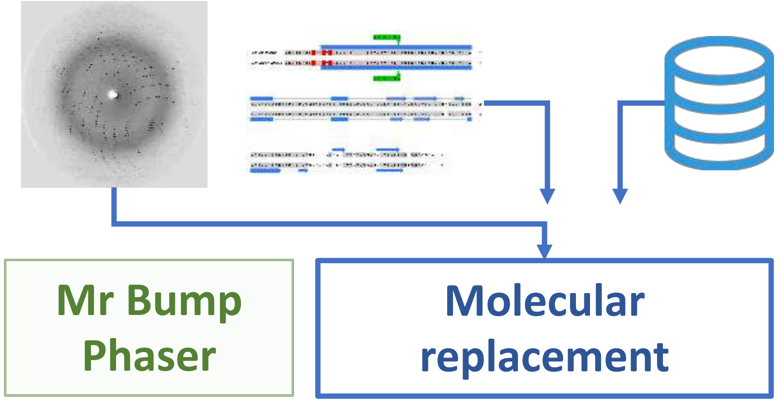






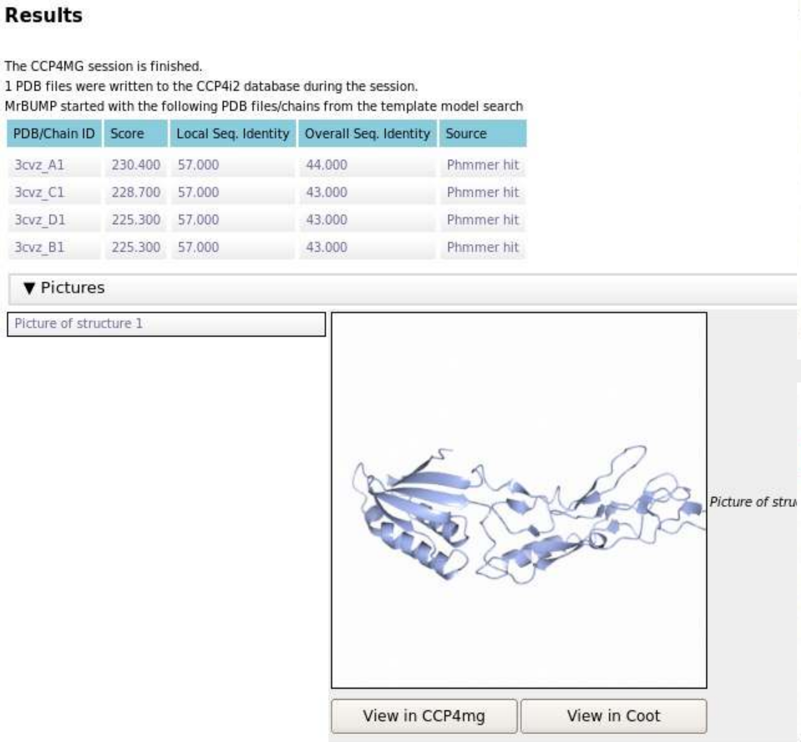


	Overall	High res.
Space group	C2	
Unit cell		
a, b, c (Å)	173.25, 29.58, 144.63	
α=γ, β (°)	90.0, 94.2	
Resolution (Å)	144.24 – 2.57	2.57-2.69
I/σI	10.2	2.2
Half-set correlation CC(1/2)	0.993	0.838
Completeness %	99	98



Mr Bump Phaser

Molecular replacement

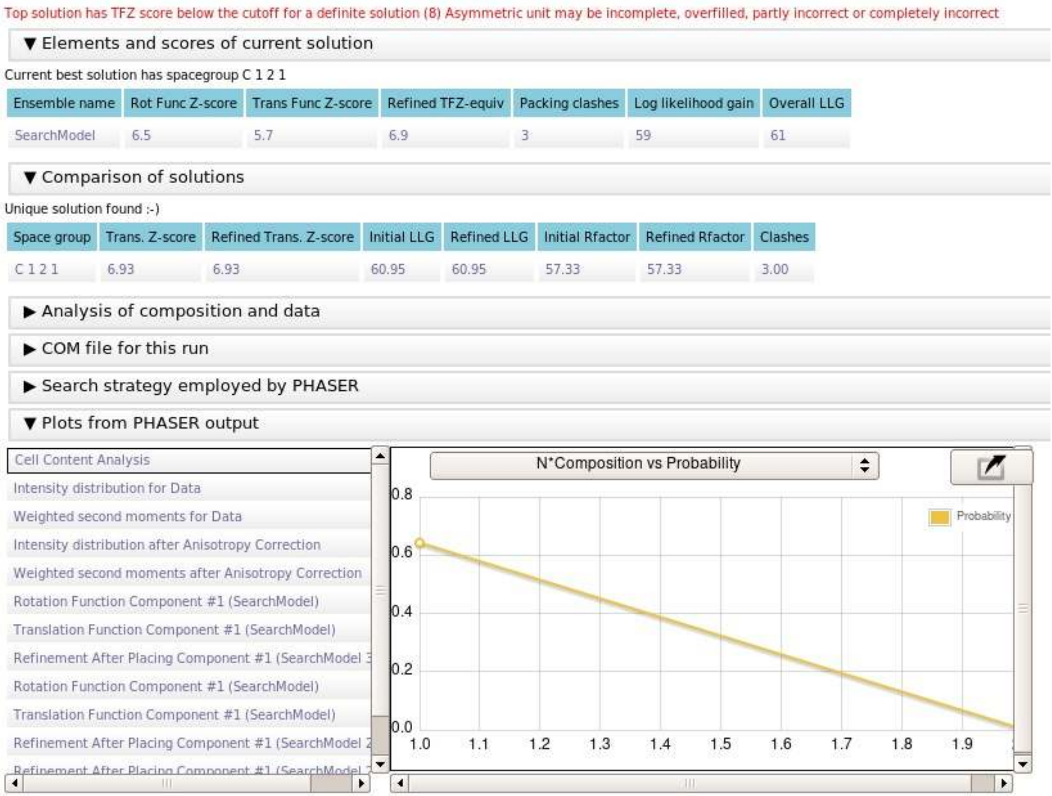


Results

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

▼ Pictures



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 :-)

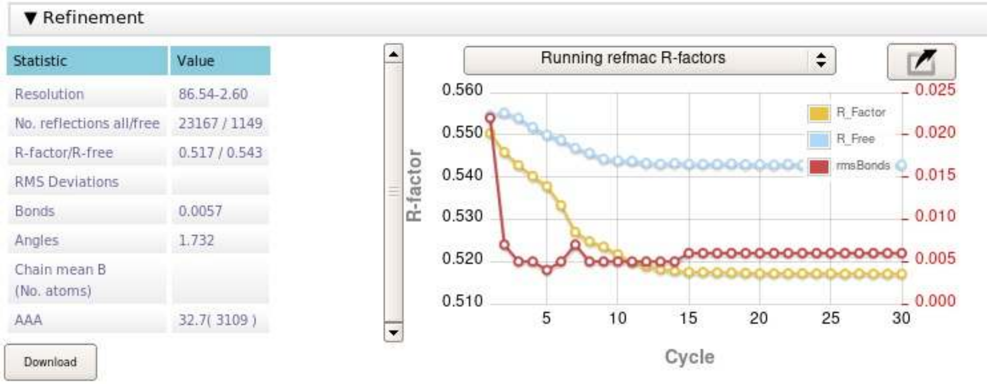
Space group	Trans. Z-score	Refined Trans. Z-score	Initial LLG	Refined LLG	Initial Rfactor	Refined Rfactor	Clashes
C 1 2 1	6.93	6.93	60.95	60.95	57.33	57.33	3.00

▶ Analysis of composition and data

▶ COM file for this run

▶ Search strategy employed by PHASER

▼ Plots from PHASER output

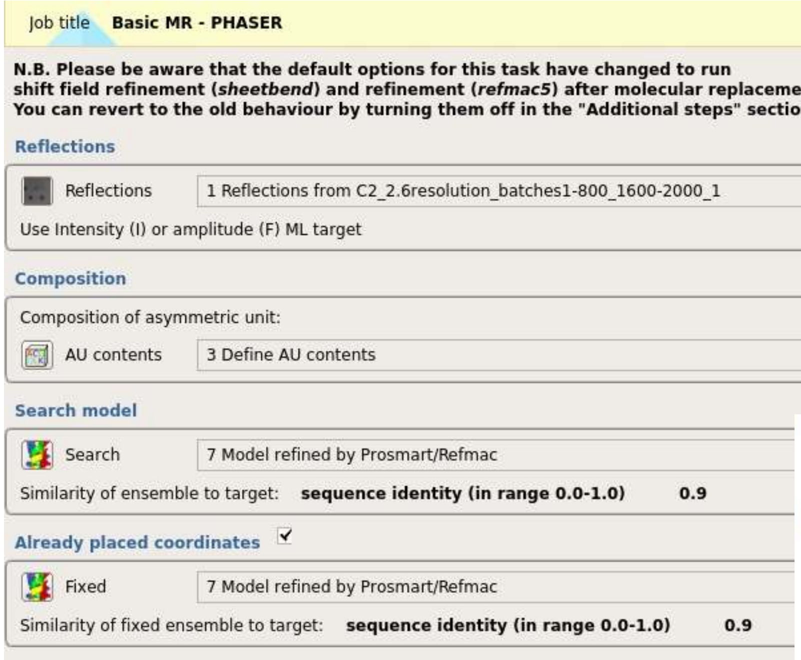


▼ Refinement

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

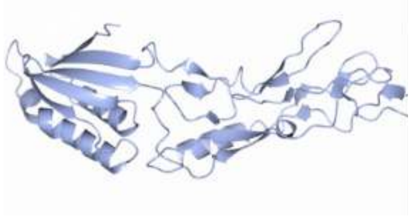
Search: 7 Model refined by Prosmart/Refmac

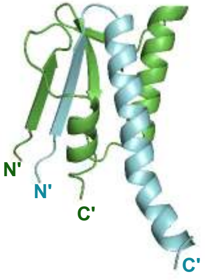
Similarity of ensemble to target: **sequence identity (in range 0.0-1.0)** **0.9**

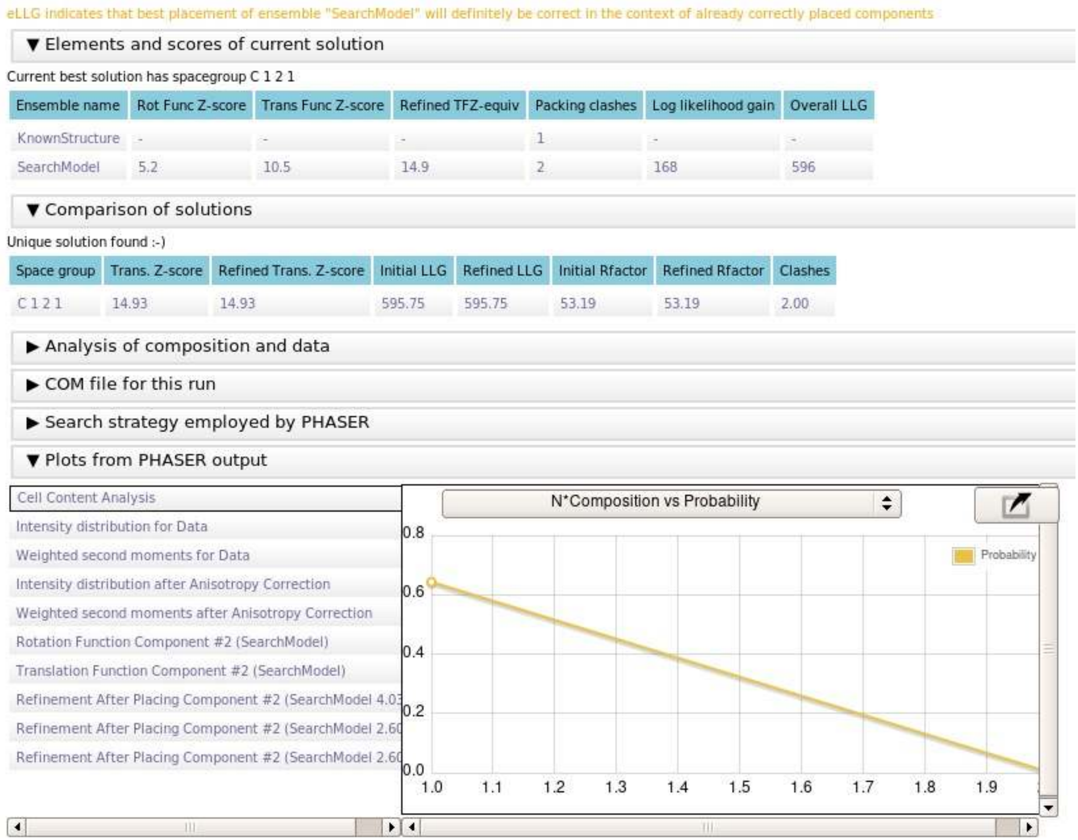
Already placed coordinates ☑

Fixed: 7 Model refined by Prosmart/Refmac

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







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

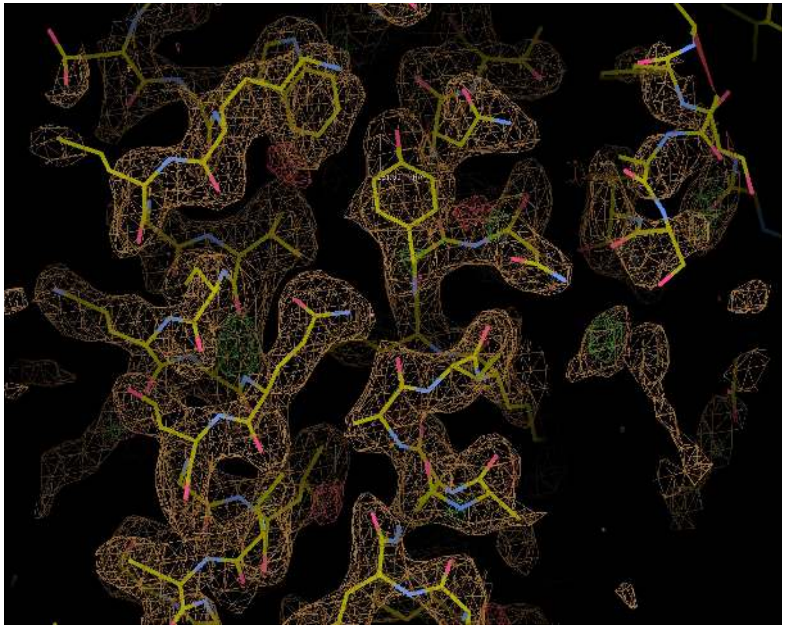
▶ Analysis of composition and data

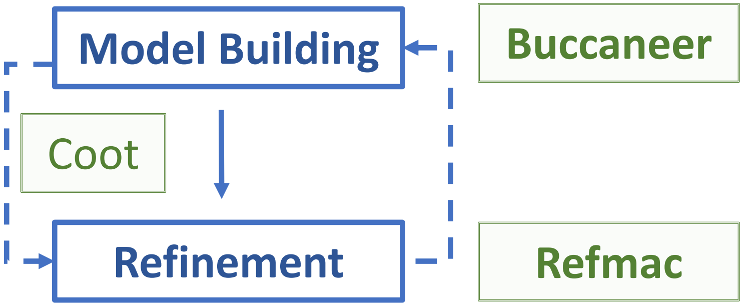
▶ COM file for this run

▶ Search strategy employed by PHASER

▼ Plots from PHASER output

Cell Content Analysis
Intensity distribution for Data
Weighted second moments for Data
Intensity distribution after Anisotropy Correction
Weighted second moments after Anisotropy Correction
Rotation Function Component #2 (SearchModel)
Translation Function Component #2 (SearchModel)
Refinement After Placing Component #2 (SearchModel 4.03
Refinement After Placing Component #2 (SearchModel 2.60
Refinement After Placing Component #2 (SearchModel 2.60





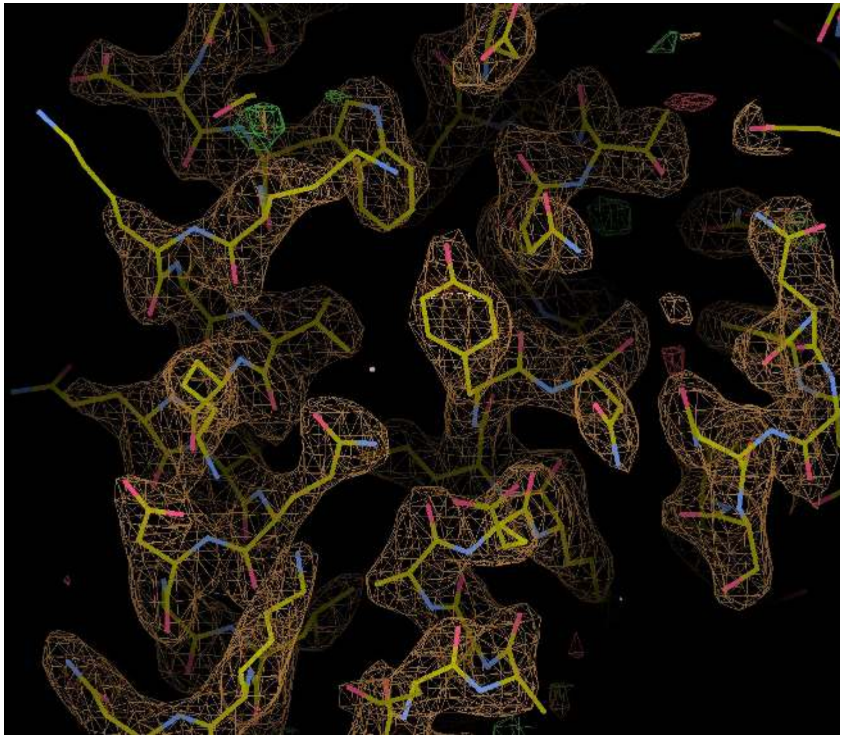
Model Building

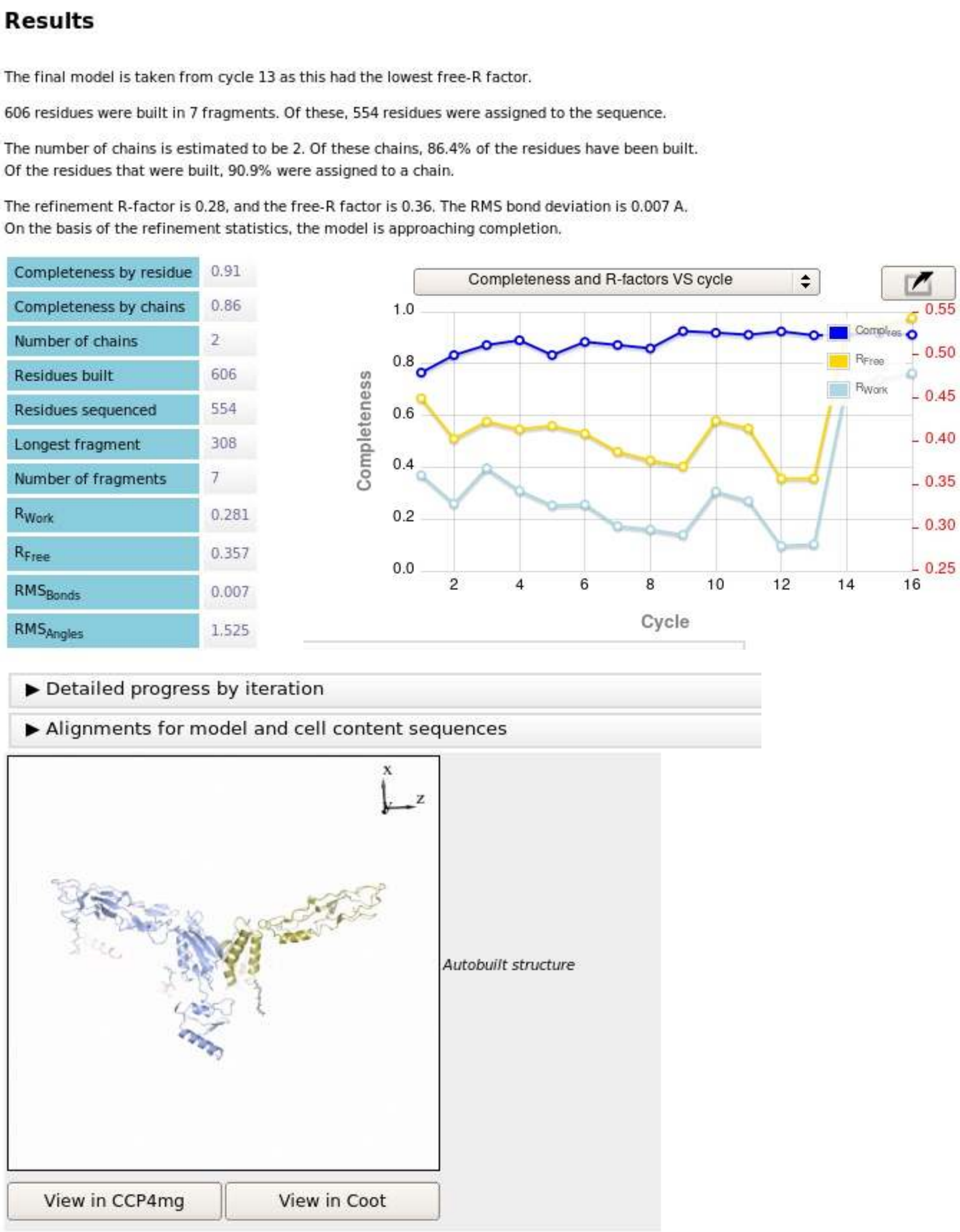
Coot

Refinement

Buccaneer

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.

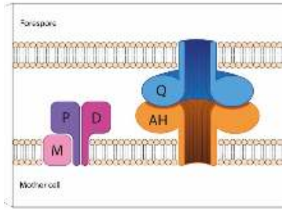
The number of chains is estimated to be 2. Of these chains, 86.4% of the residues have been built. Of the residues that were built, 90.9% were assigned to a chain.

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.

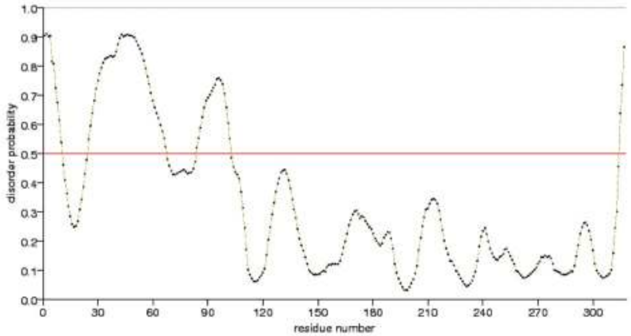
Completeness by residue	0.91
Completeness by chains	0.86
Number of chains	2
Residues built	606
Residues sequenced	554
Longest fragment	308
Number of fragments	7
R_{Work}	0.281
R_{Free}	0.357
RMS_{Bonds}	0.007
RMS_{Angles}	1.525

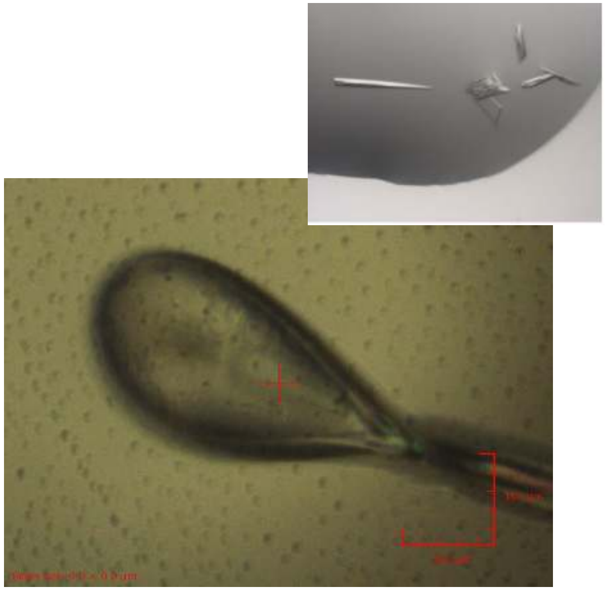
Autobuilt structure

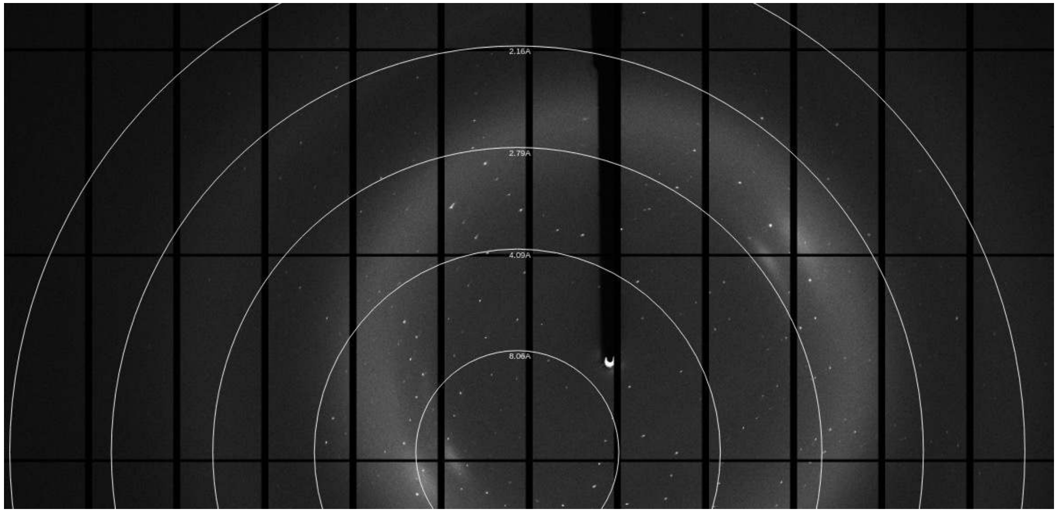
Case study 3: *C. difficile* PG hydrolase



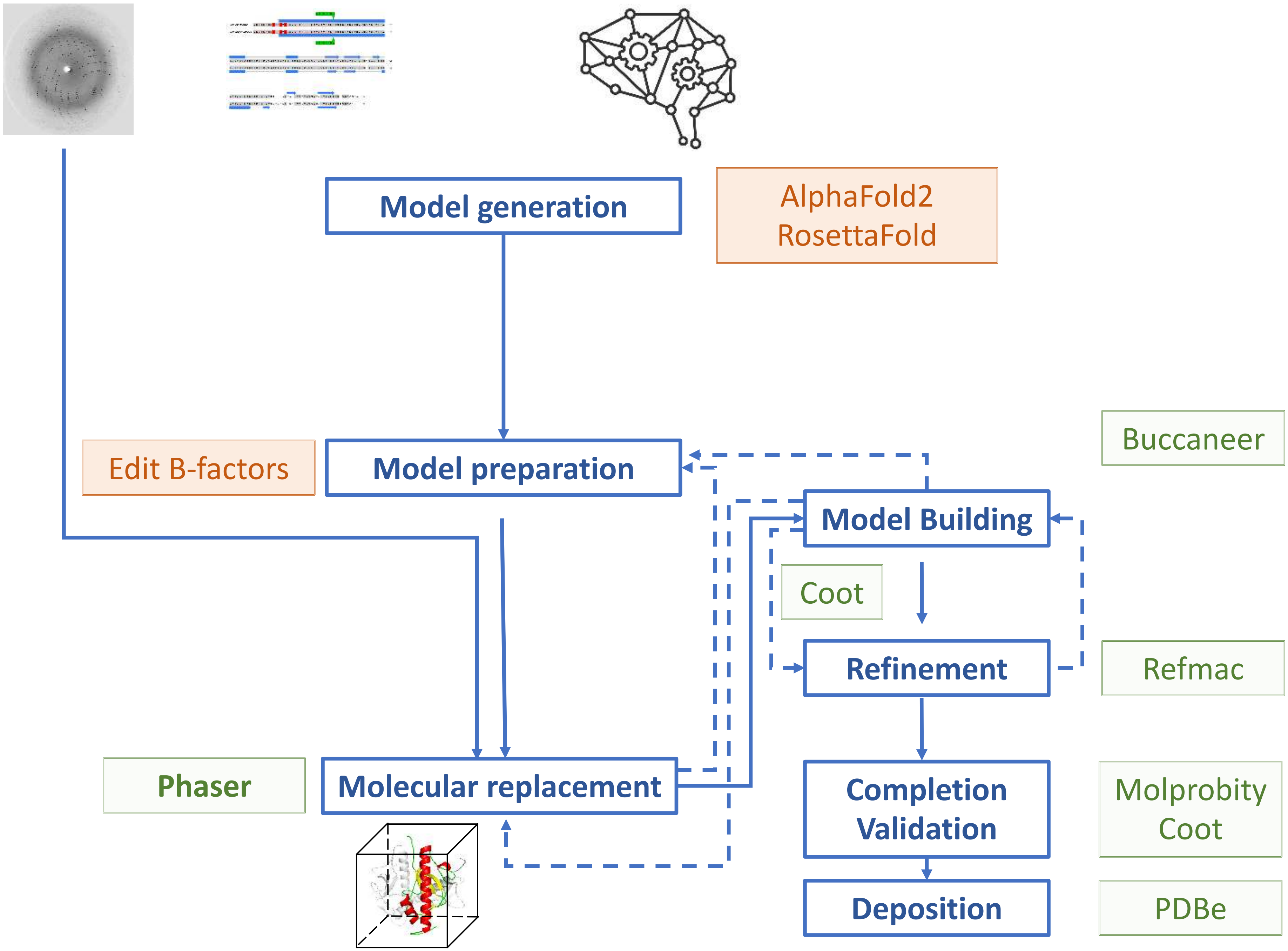
```
1   GAMGNQDDFL KFLVNSSYPE AKVEGNDTEN KKNNKNKETS KENKEESKEE  50
51  NTKSKDASKV DNKKESEKEY IKLYVGKENV PDIESKNSDT TETNTTSSSD 100
101 YKDDLRVTKE NPRILIYHTH GCETYSNSPD GNYHSRDKKN SVMEVGSALT 150
151 SALDSKGWGV VHTTKYHDYP SYNNSYASSL KTIQSILPKY NSVDIAIDLH 200
201 RDARDLTNPA TKEKDHLKYT TMINGERVSK FFFVVGGKNT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







	Overall	High res.
Space group	P2 2_1 2_1	
Unit cell		
a, b, c (Å)	47.62, 51,88, 81.80	
α=β=γ (°)	90.0	
Resolution (Å)	43.81 - 1.56	1.59 - 1.56
Rmerge	0.148	1.644
I/σI	7.1	1.2
Half-set correlation CC(1/2)	0.996	0.548
Completeness %	100	99.9



Model generation

AlphaFold2
RosettaFold

Edit B-factors

Model preparation

Buccaneer

Model Building

Coot

Refinement

Refmac

Phaser

Molecular replacement

Completion
Validation

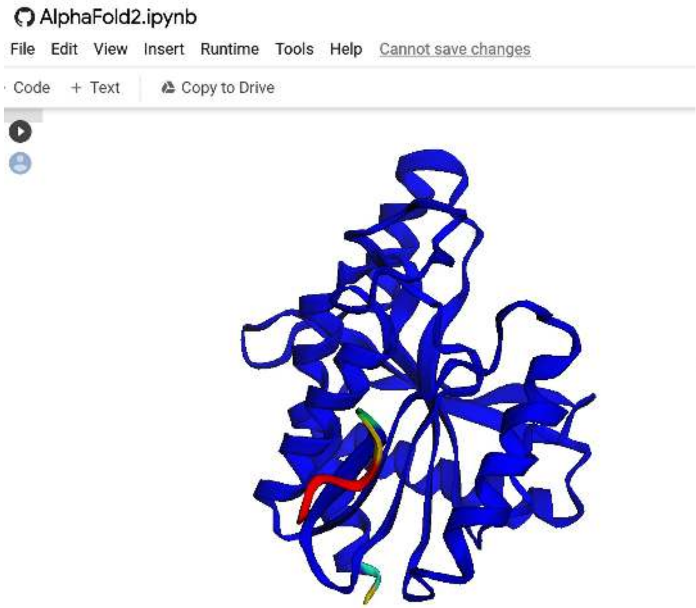
Molprobity
Coot

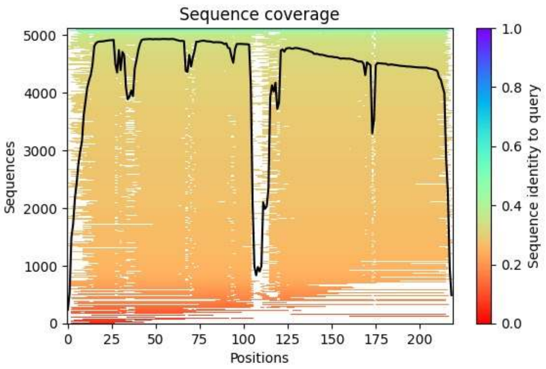
Deposition

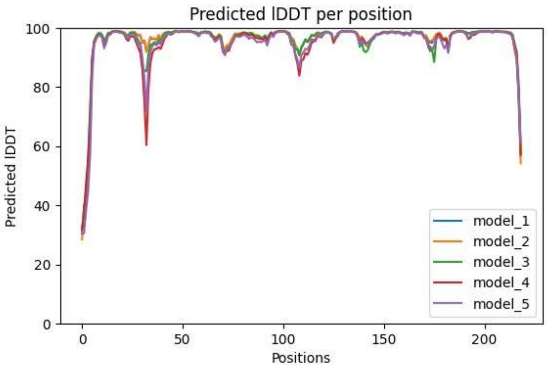
PDBe

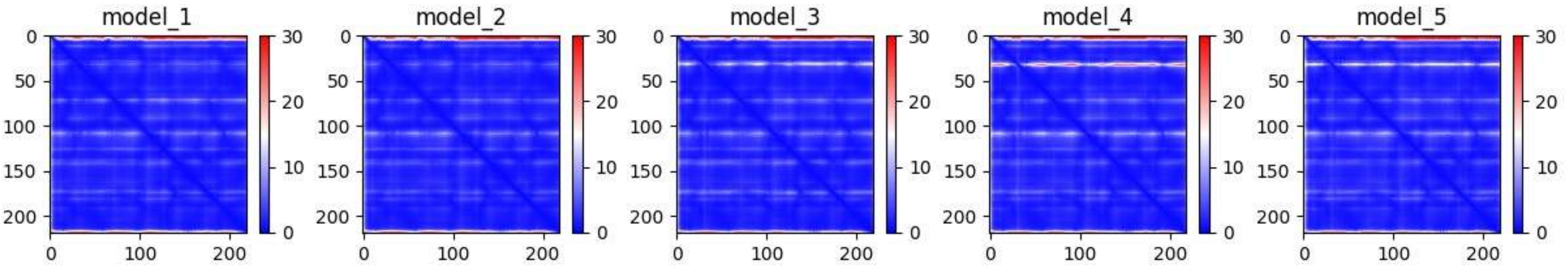
Model generation

AlphaFold2





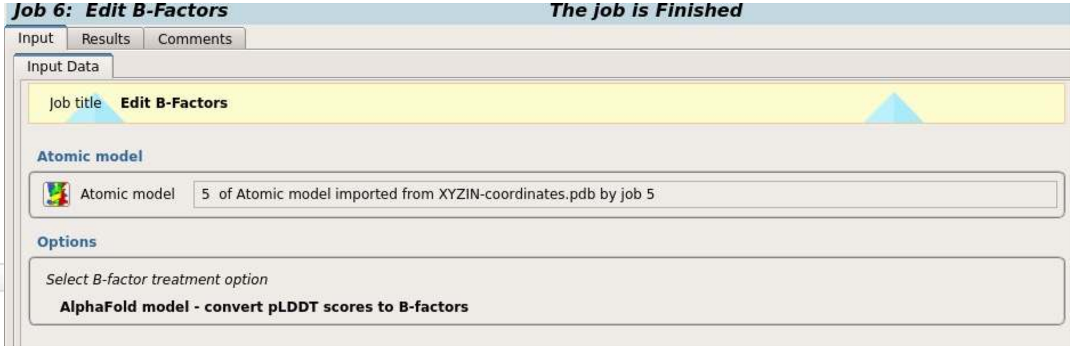




Model preparation

Molecular replacement

Edit B-factors


Job 6: Edit B-Factors **The job is Finished**

Job title **Edit B-Factors**

Atomic model

Atomic model: 5 of Atomic model imported from XYZIN-coordinates.pdb by job 5

Options

Select B-factor treatment option

AlphaFold model - convert pLDDT scores to B-factors

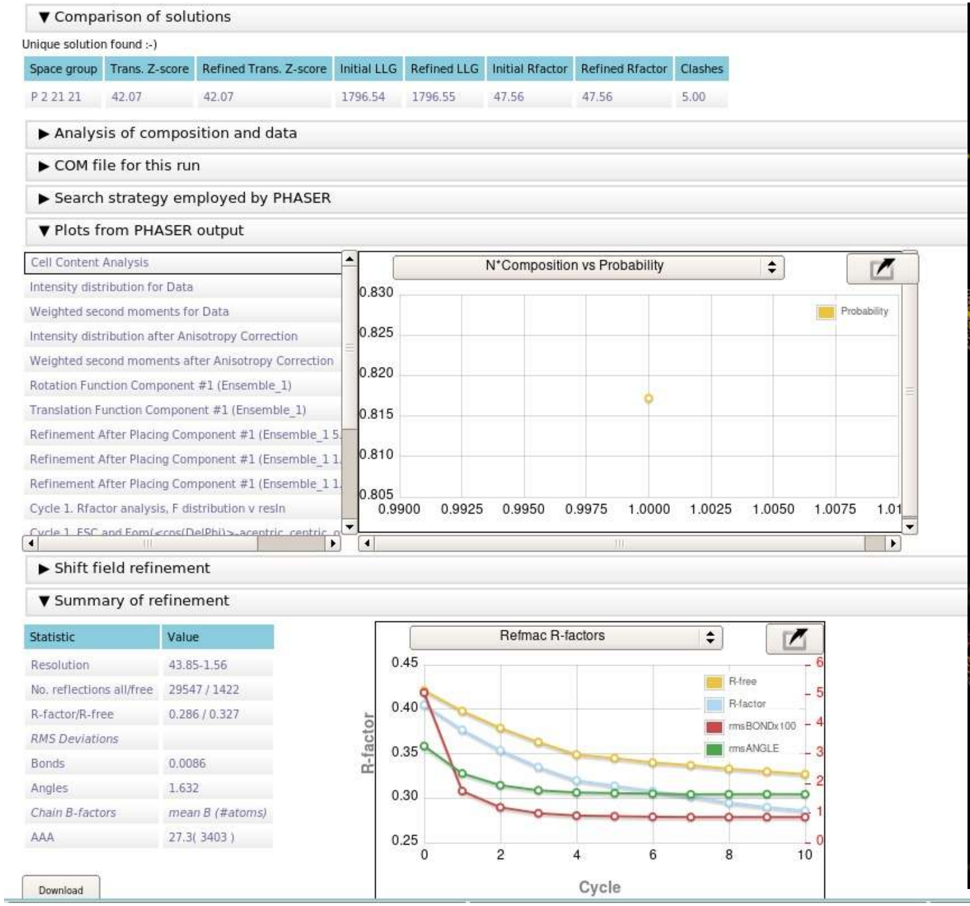
Phaser

Expert MR - PHASER AF2 model

▼ Elements and scores of current solution

Current best solution has spacegroup P 2 21 21

Ensemble name	Rot Func Z-score	Trans Func Z-score	Refined TFZ-equiv	Packing clashes	Log likelihood gain	Overall LLG
Ensemble_1	5.1	8.7	42.1	5	131	1797


▼ Comparison of solutions

Unique solution found :-)

Space group	Trans. Z-score	Refined Trans. Z-score	Initial LLG	Refined LLG	Initial Rfactor	Refined Rfactor	Clashes
P 2 21 21	42.07	42.07	1796.54	1796.55	47.56	47.56	5.00

▶ Analysis of composition and data

▶ COM file for this run

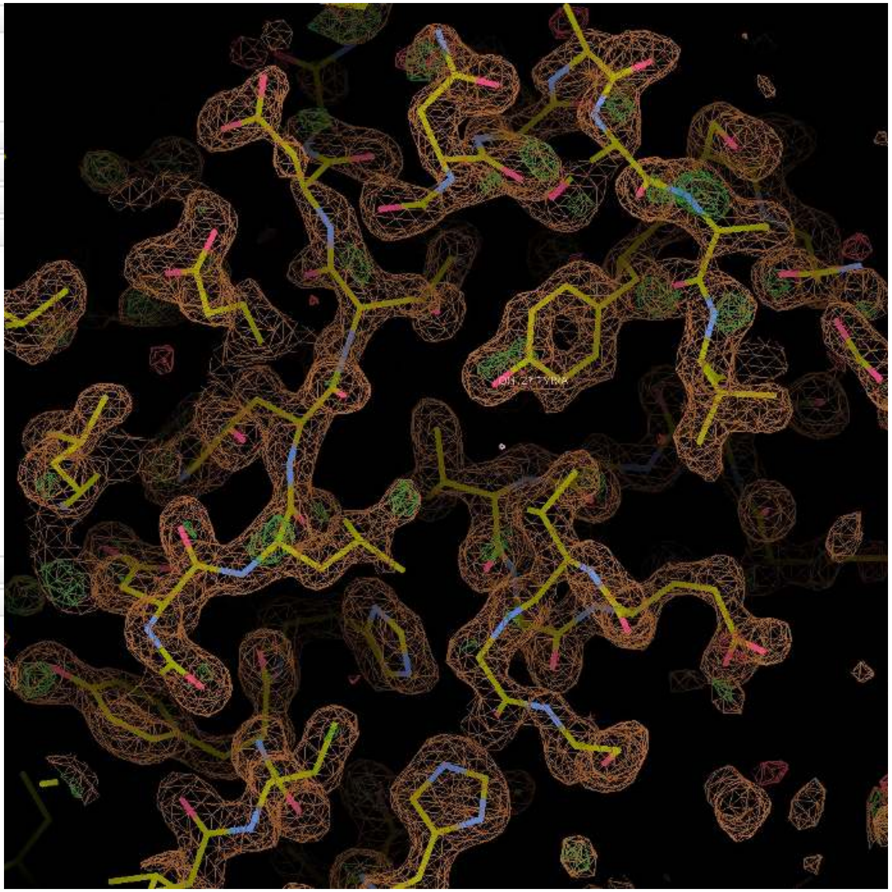
▶ Search strategy employed by PHASER

▼ Plots from PHASER output

▶ Shift field refinement

▼ Summary of refinement

Statistic	Value
Resolution	43.85-1.56
No. reflections all/free	29547 / 1422
R-factor/R-free	0.286 / 0.327
RMS Deviations	
Bonds	0.0086
Angles	1.632
Chain B-factors	*mean B (#atoms)*
AAA	27.3(3403)

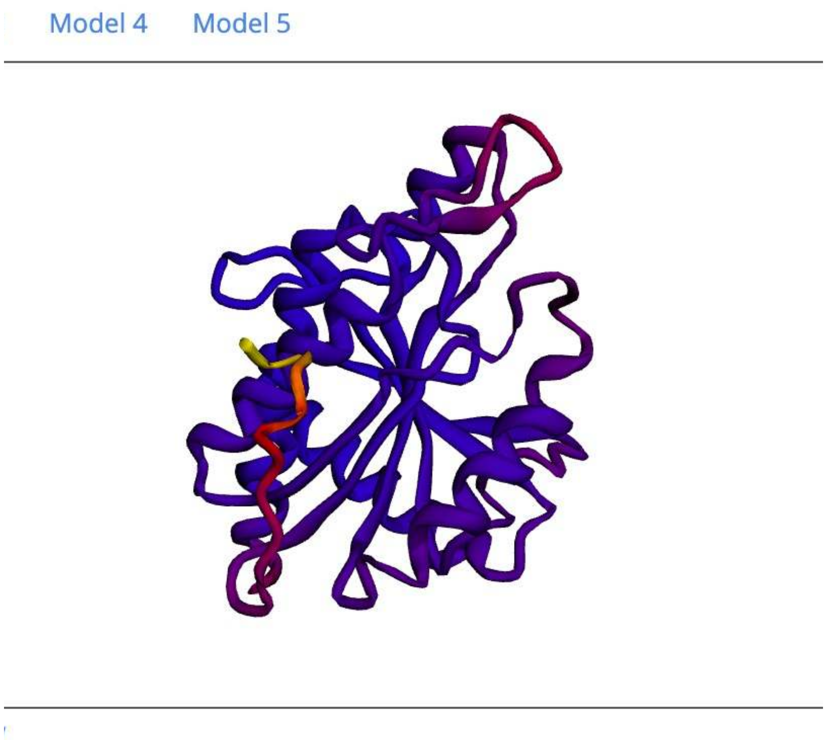


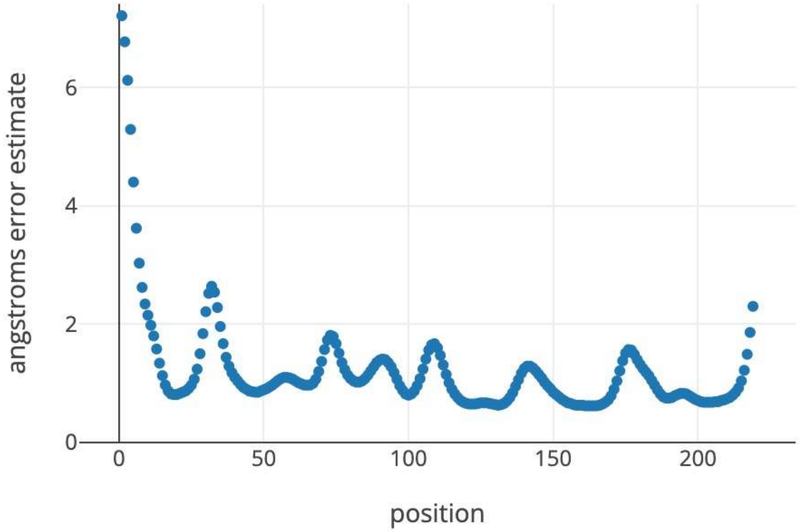
Model generation

RosettaFold

Model preparation

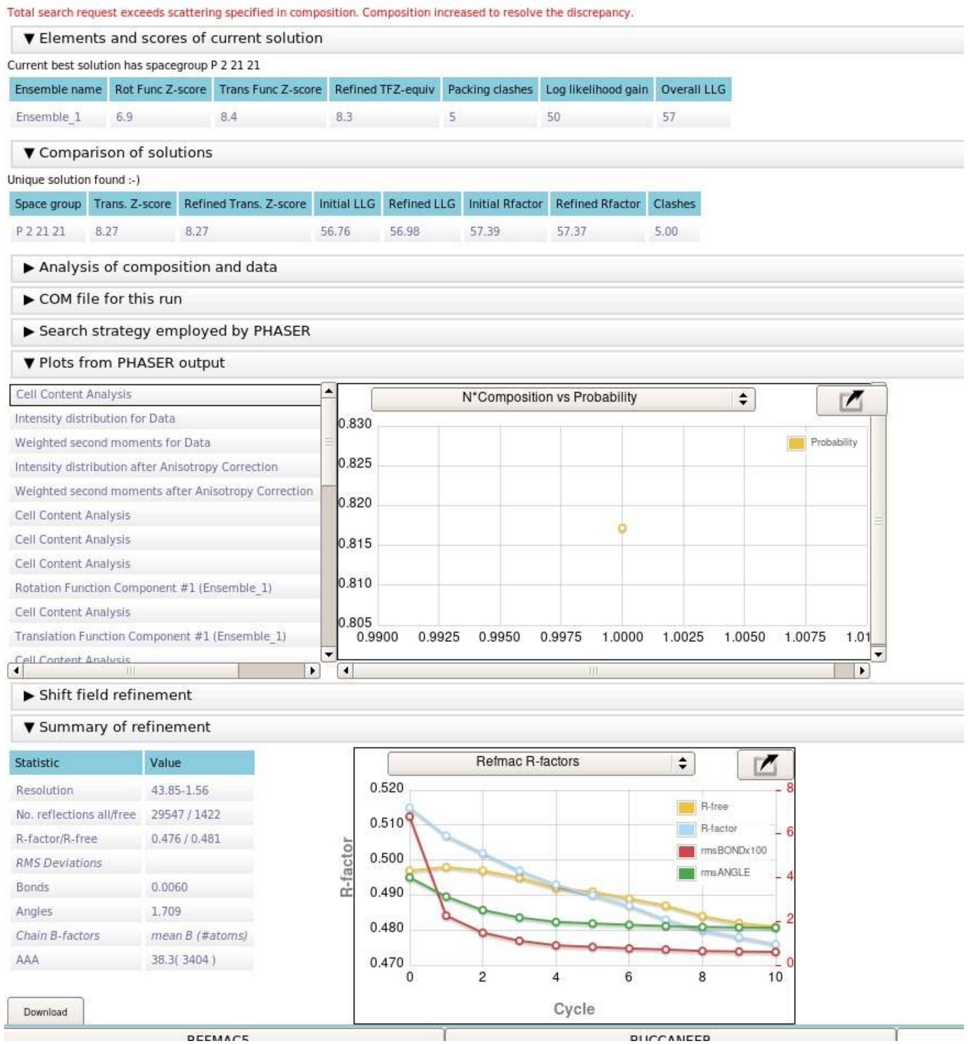
Edit B-factors





Molecular replacement

Phaser



Total search request exceeds scattering specified in composition. Composition increased to resolve the discrepancy.

▼ Elements and scores of current solution

Current best solution has spacegroup P 2 21 21

Ensemble name	Rot Func Z-score	Trans Func Z-score	Refined TFZ-equiv	Packing clashes	Log likelihood gain	Overall LLG
Ensemble_1	6.9	8.4	8.3	5	50	57

▼ Comparison of solutions

Unique solution found :-)

Space group	Trans. Z-score	Refined Trans. Z-score	Initial LLG	Refined LLG	Initial Rfactor	Refined Rfactor	Clashes
P 2 21 21	8.27	8.27	56.76	56.98	57.39	57.37	5.00

▶ Analysis of composition and data

▶ COM file for this run

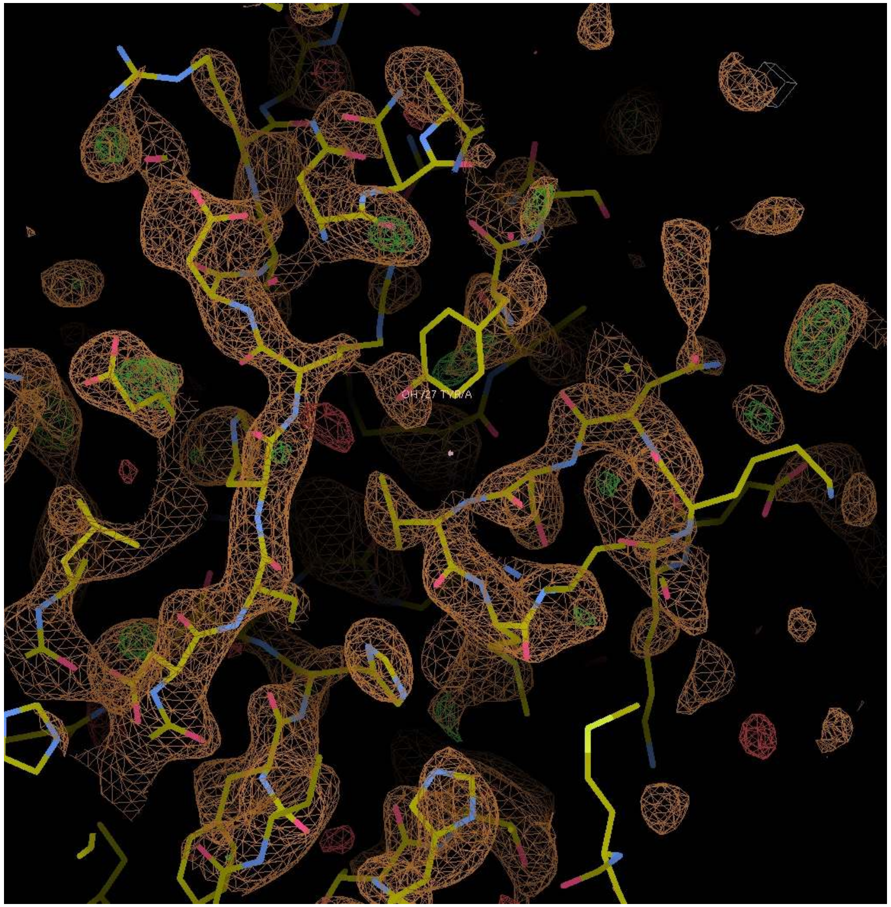
▶ Search strategy employed by PHASER

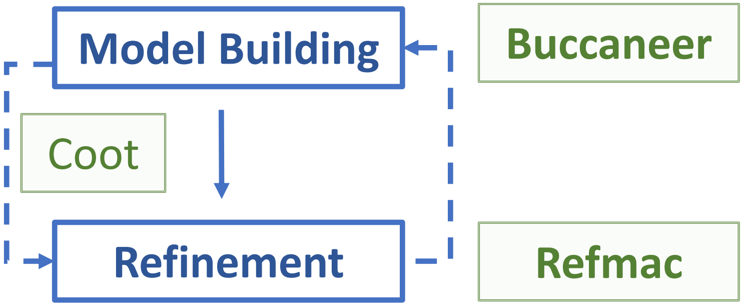
▼ Plots from PHASER output

▶ Shift field refinement

▼ Summary of refinement

Statistic	Value
Resolution	43.85-1.56
No. reflections all/free	29547 / 1422
R-factor/R-free	0.476 / 0.481
RMS Deviations	
Bonds	0.0060
Angles	1.709
Chain B-factors	*mean B (#atoms)*
AAA	38.3(3404)





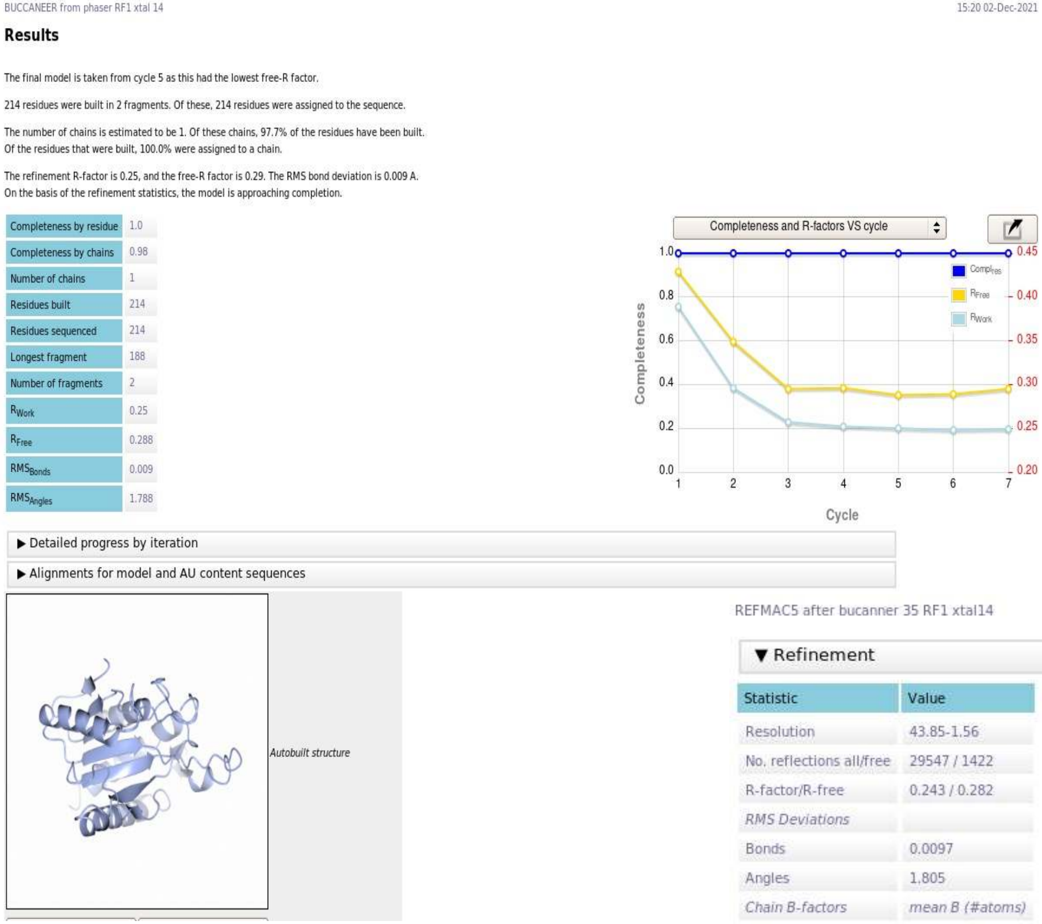
Model Building

Coot

Refinement

Buccaneer

Refmac



BUCCANEER from phaser RF1 xtal 14

15:20 02-Dec-2021

Results

The final model is taken from cycle 5 as this had the lowest free-R factor.

214 residues were built in 2 fragments. Of these, 214 residues were assigned to the sequence.

The number of chains is estimated to be 1. Of these chains, 97.7% of the residues have been built. Of the residues that were built, 100.0% were assigned to a chain.

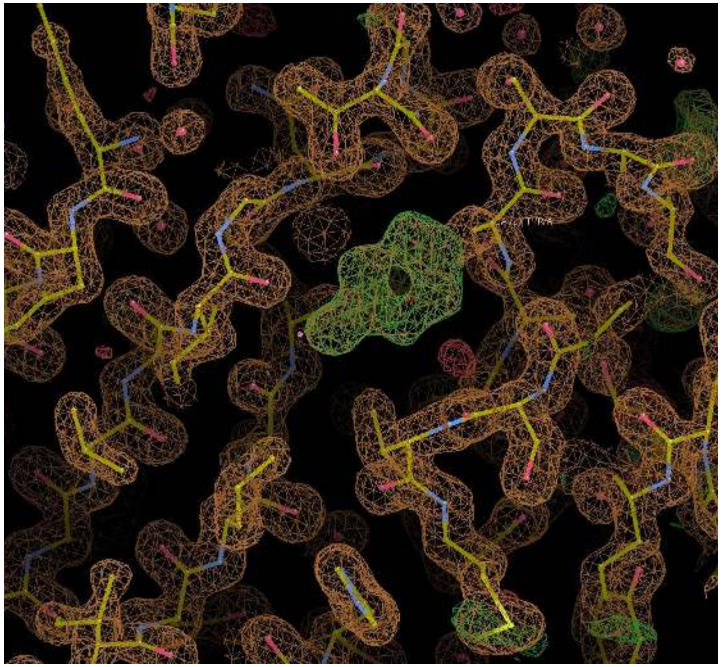
The refinement R-factor is 0.25, and the free-R factor is 0.29. The RMS bond deviation is 0.009 A. On the basis of the refinement statistics, the model is approaching completion.

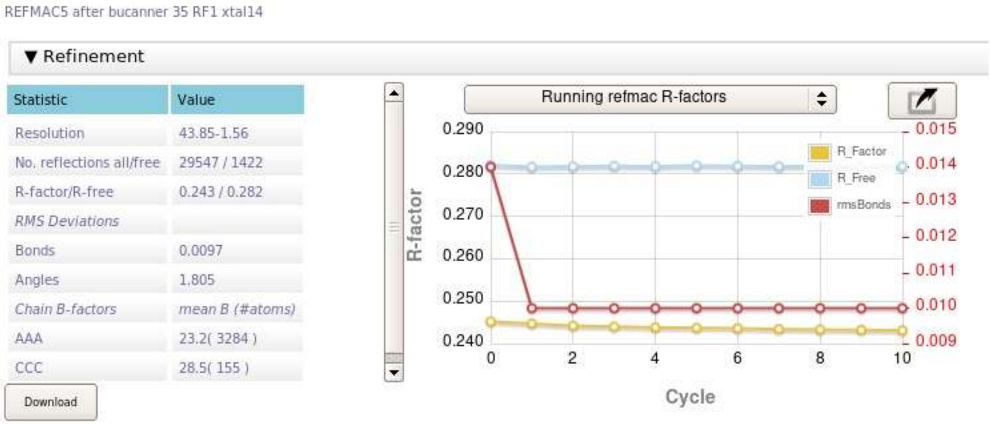
Completeness by residue	1.0
Completeness by chains	0.98
Number of chains	1
Residues built	214
Residues sequenced	214
Longest fragment	188
Number of fragments	2
R_{Work}	0.25
R_{Free}	0.288
RMS_{Bonds}	0.009
RMS_{Angles}	1.788

▶ Detailed progress by iteration

▶ Alignments for model and AU content sequences

Autobuilt structure

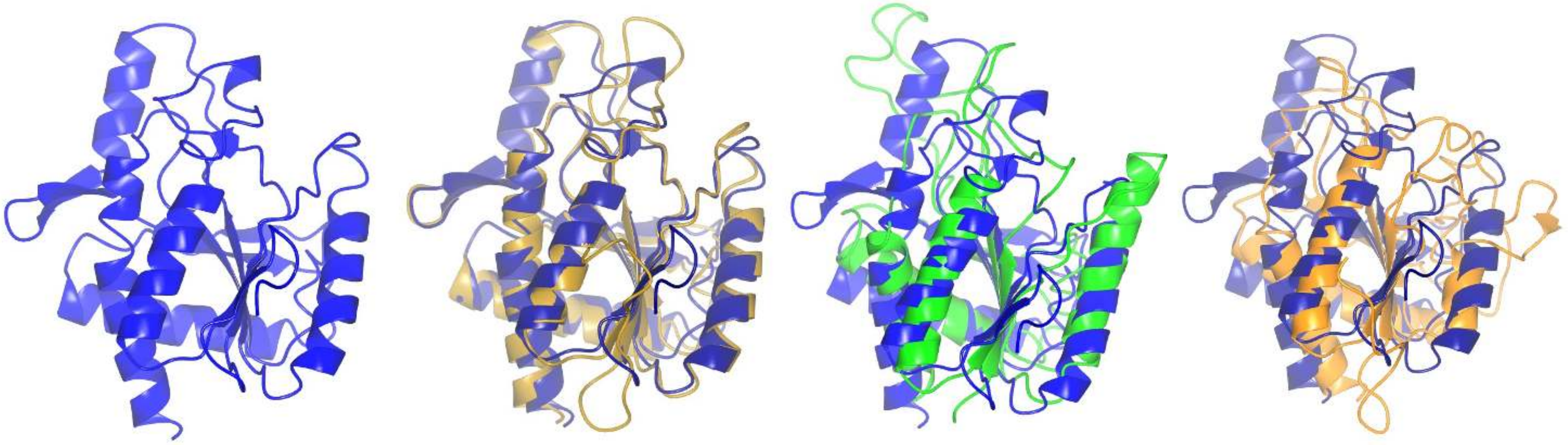




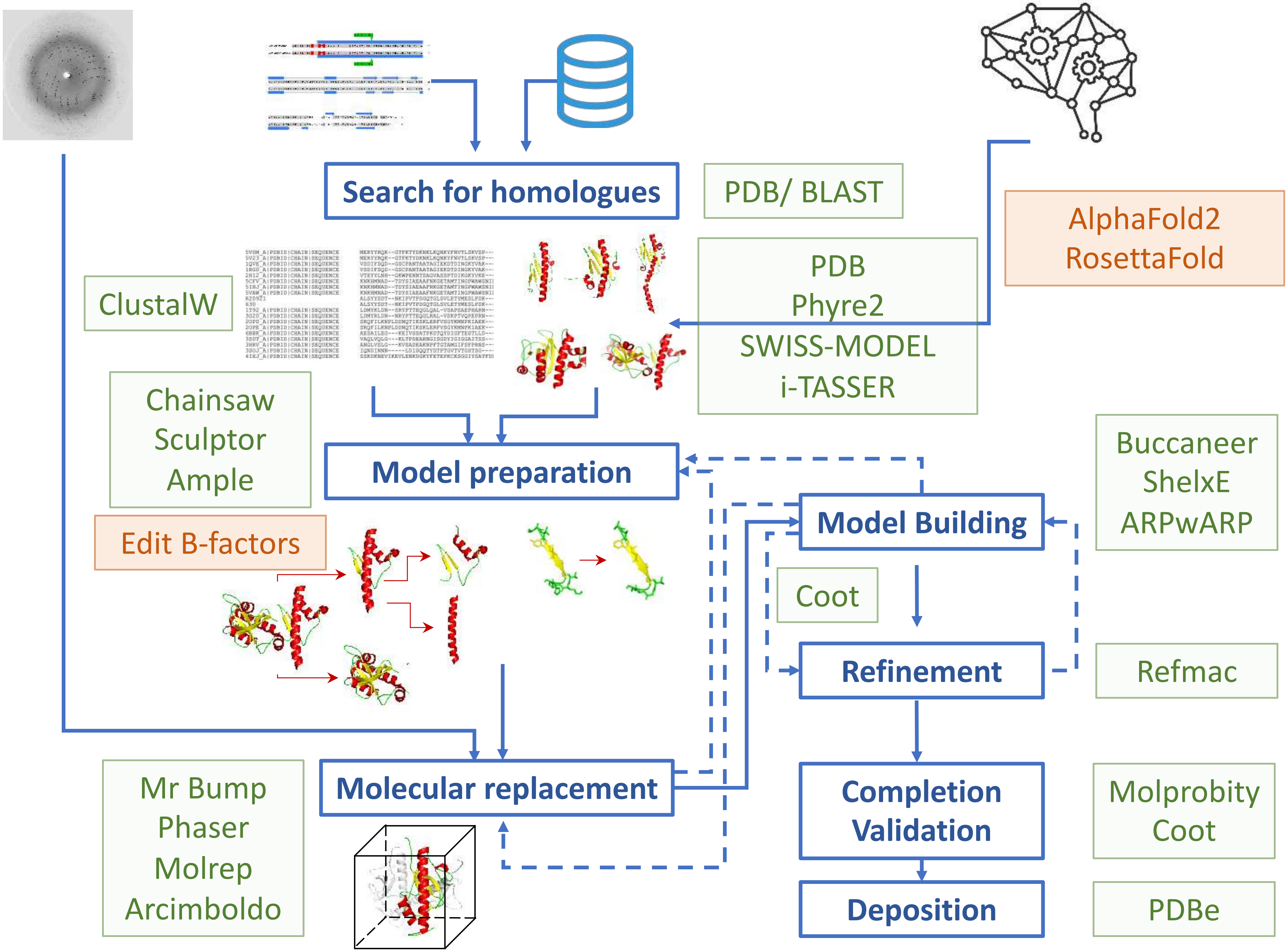
REFMAC5 after bucanner 35 RF1 xtal14

▼ Refinement

Statistic	Value
Resolution	43.85-1.56
No. reflections all/free	29547 / 1422
R-factor/R-free	0.243 / 0.282
RMS Deviations	
Bonds	0.0097
Angles	1.805
Chain B-factors	*mean B (#atoms)*
AAA	23.2(3284)
CCC	28.5(155)



AlphaFold2	RosettaFold	SWISS-MODEL	Phyre2
Phaser: TFZ – 42.1 R = 29% Rfree = 32%	Phaser: TFZ – 8.2 R = 48% Rfree = 48%	Phaser: TFZ – 3.6 No solution	Phaser: TFZ – 3.7 No solution
Buccaneer: 215 aa R = 26% Rfree = 31%	Buccaneer: 214 aa R = 25% Rfree = 29%	-	-
Refmac: R = 25% Rfree = 30%	Refmac: R = 24% Rfree = 28%	-	-



Search for homologues

PDB/ BLAST

AlphaFold2
RosettaFold

ClustalW

PDB
Phyre2
SWISS-MODEL
i-TASSER

Chainsaw
Sculptor
Ample

Model preparation

Edit B-factors

Model Building

Buccaneer
ShelxE
ARPwARP

Coot

Refinement

Refmac

Molecular replacement

Mr Bump
Phaser
Molrep
Arcimboldo

Completion Validation

Molprobity
Coot

Deposition

PDBe

Newcastle Structural Biology Laboratory

Newcastle University

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

Arnaud Baslé

The University Of Sheffield.

Rob Fagan
Joe Kirk
Per Bullough
Oishik Banerji
Jason Wilson

University of Glasgow

Gill Douce
Filipa Vaz

Imperial College London

Neil Fairweather
Ed Couchman

BBSRC Doctoral Training Partnerships

MRC Medical Research Council

wellcome