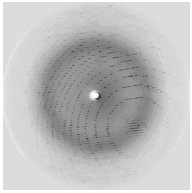
MR case studies

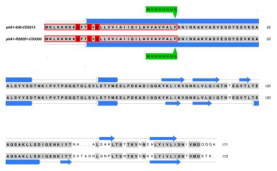
Paula S. Salgado

DLS/CCP4 Workshop 2022

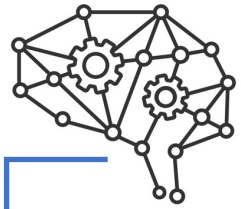
SALGADO LAB

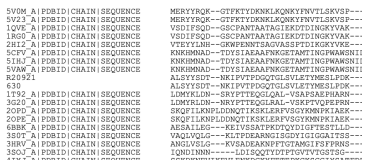
@paulasalgado@mastodon.social @pssalgado salgadolab.org



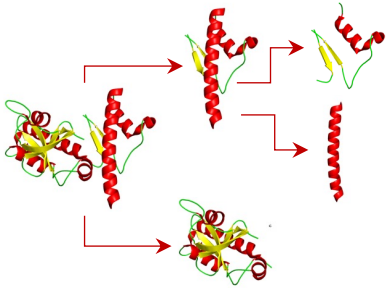


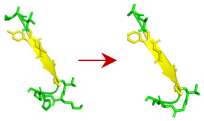
Search for homologues





Model preparation

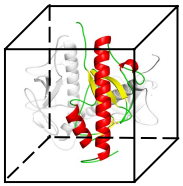




Model Building

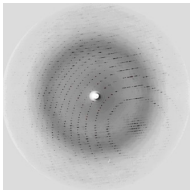
Refinement

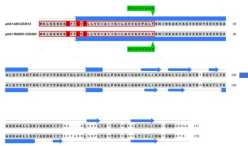
Molecular replacement



Completion
Validation

Deposition



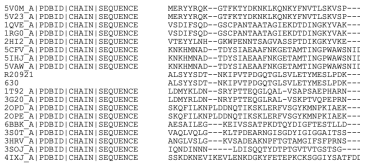


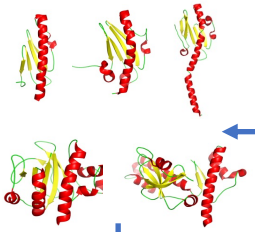
Search for homologues

PDB/ BLAST

AlphaFold2
RosettaFold

ClustalW



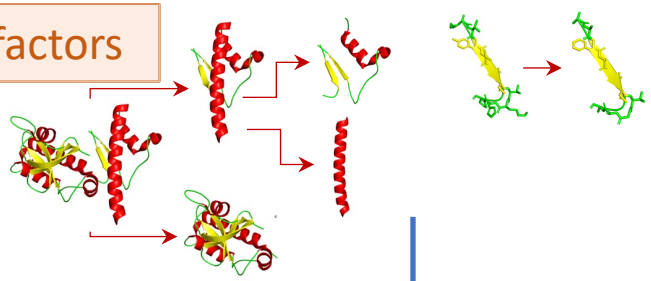


PDB
Phyre2
SWISS-MODEL
i-TASSER

Chainsaw
Sculptor
Ample

Model preparation

Edit B-factors



Model Building

Buccaneer
ShelxE
ARPwARP

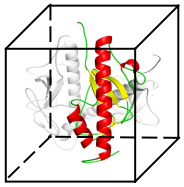
Coot

Refinement

Refmac

Mr Bump
Phaser
Molrep
Arcimboldo

Molecular replacement



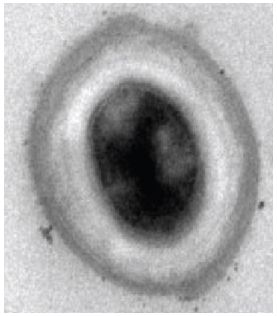
Completion Validation

Molprobity
Coot

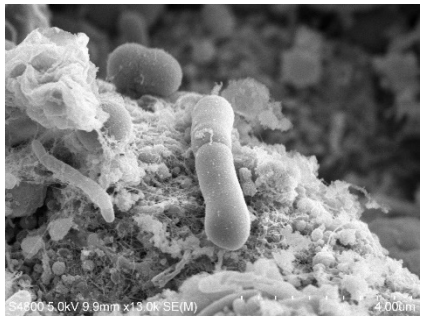
Deposition

PDBe

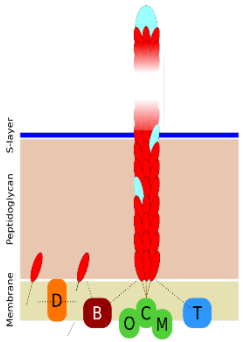
Case studies from *C. difficile* proteins

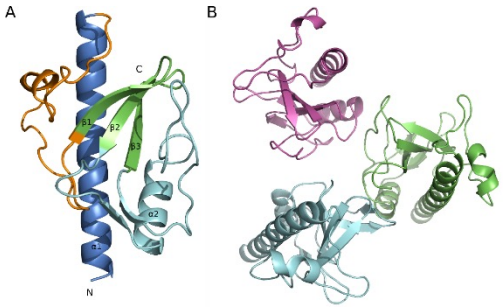


+
Antibiotics

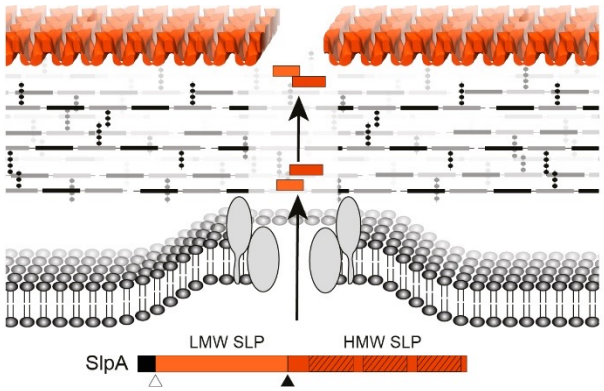


Inflammatory
Complications

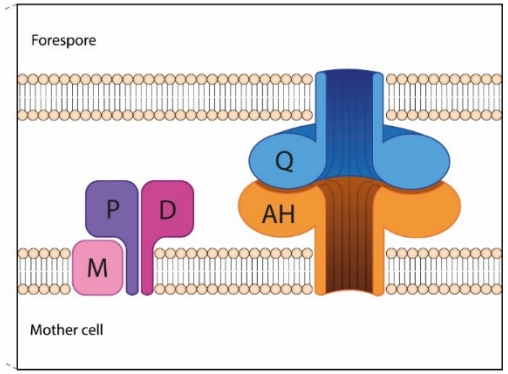




Type IV pili

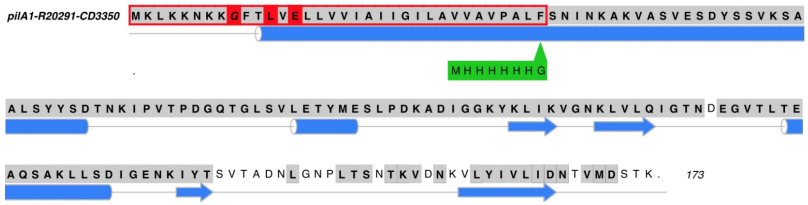


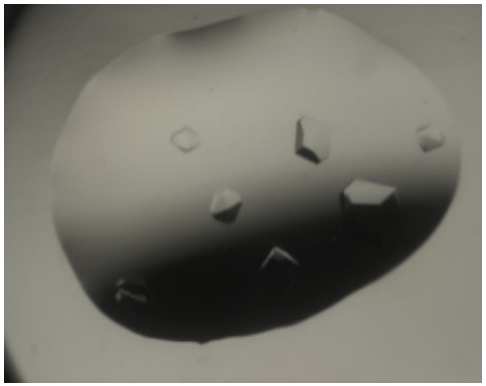
S-layer



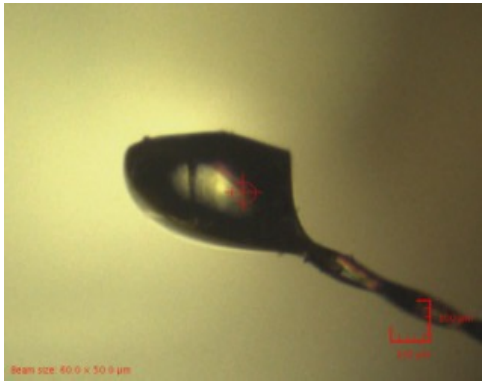
Sporulation/
engulfment

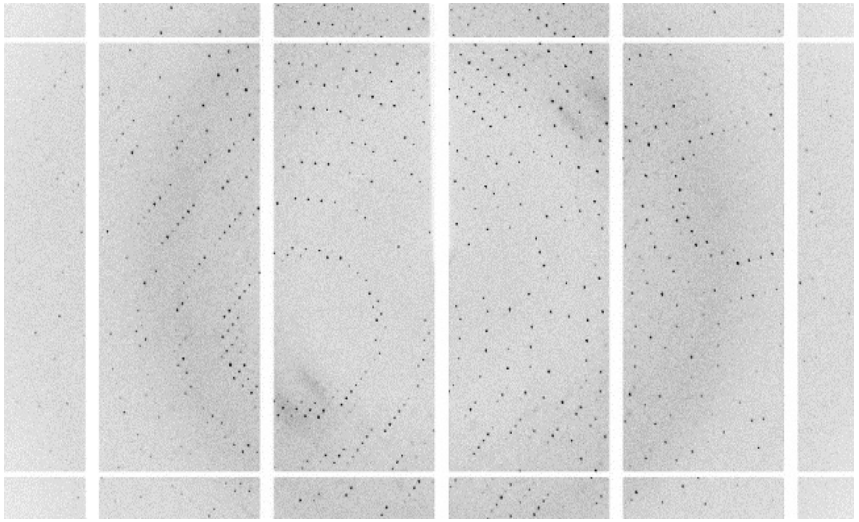
Example 1: CdPilA1



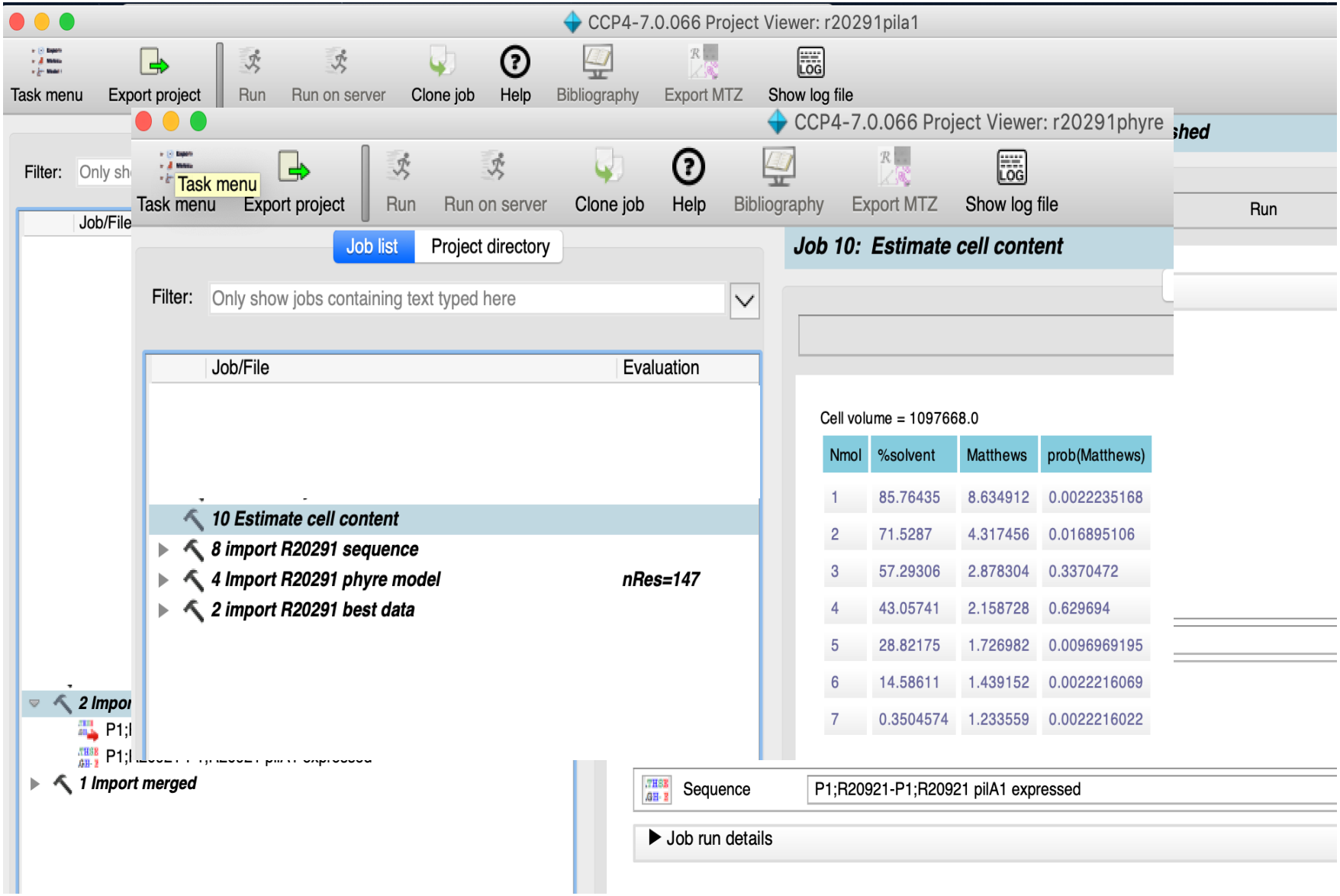


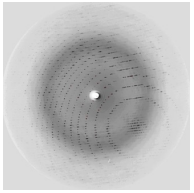
A. Crawshaw, C. Davies





	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)


CCP4-7.0.066 Project Viewer: r20291pila1
Task menu
Export project
Run
Run on server
Clone job
Help
Bibliography
Export MTZ
Show log file
CCP4-7.0.066 Project Viewer: r20291phyre
Filter:
Job list
Project directory
Filter: Only show jobs containing text typed here
Job/File
Evaluation
10 Estimate cell content
8 import R20291 sequence
4 Import R20291 phyre model
nRes=147
2 import R20291 best data
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
Job run details



Search for homologues

PDB/ BLAST

BLAST® » blastp suite

Standard Protein BLAST

blastn | blastp | blastx | tblastn | tblastx

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 328PAW0ZQ14 (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

<40 40-50 50-80 80-200 >=200

Query

1 20 40 60 80 100 120 140

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB

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

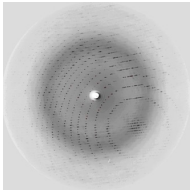
View: Gallery Reports: Select a Report Sort: Release Date: Newest to Oldest Download Files

6BBK 5VAW 5V0M 5V23

5G23 5G24 5G25 5G2F

5JW8 5HZ7 4XA2 4V1J

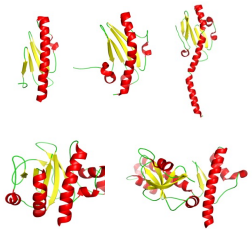
4QS4 4D40 4US7 4OGM



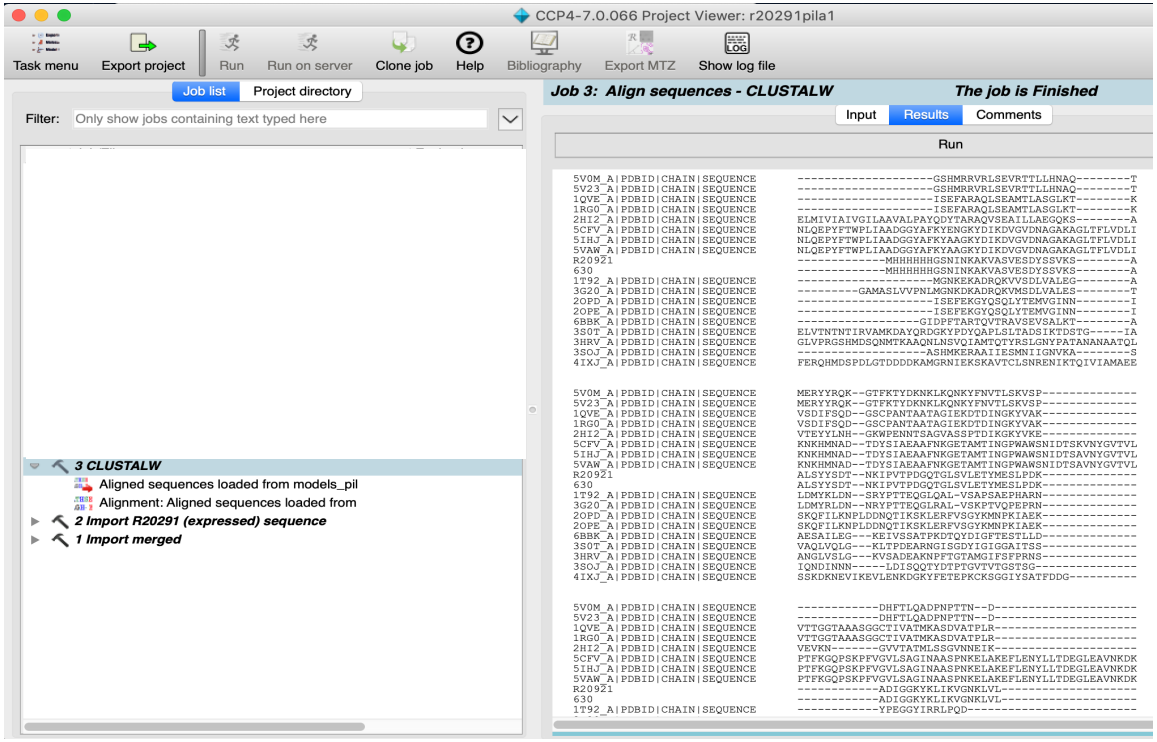
Search for homologues

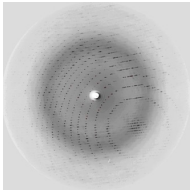
PDB/ BLAST

ClustalW



PDB

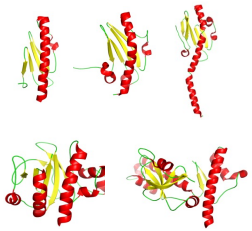




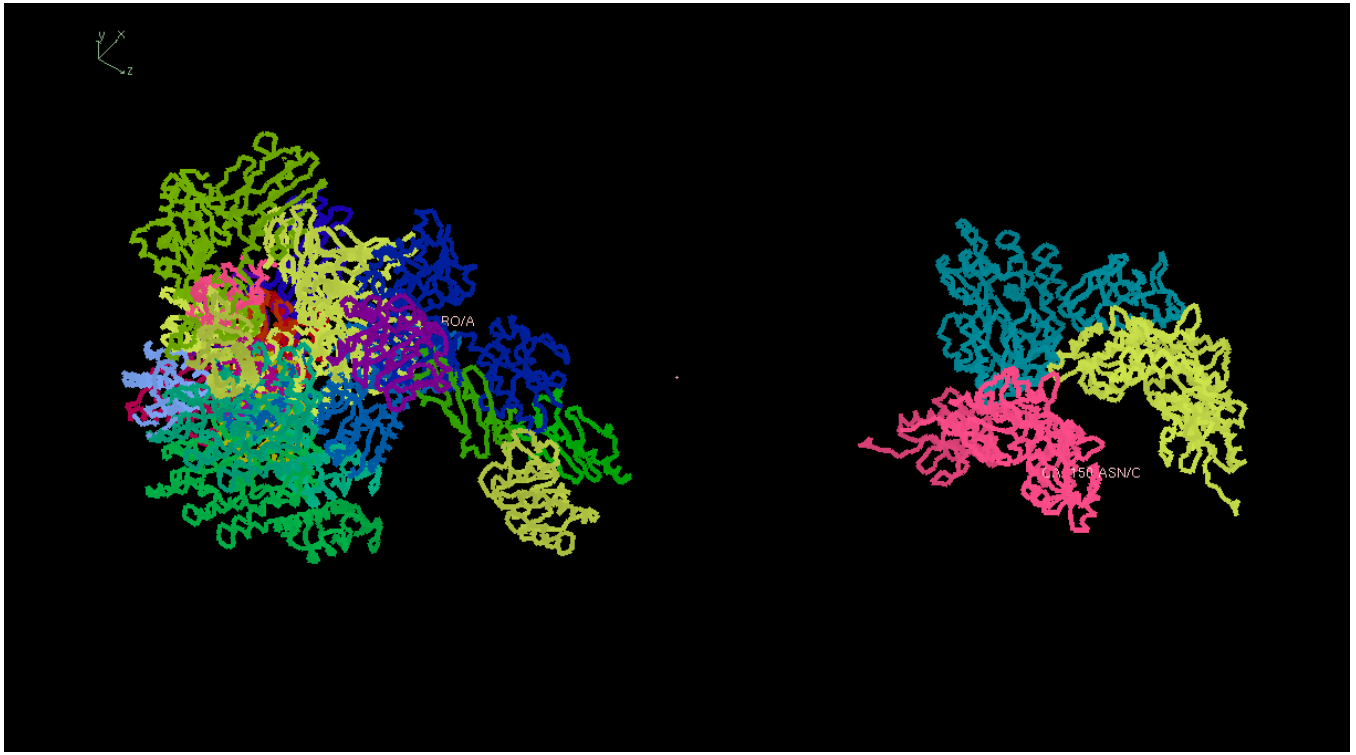
Search for homologues

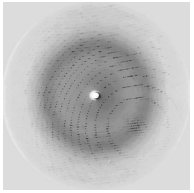
ClustalW

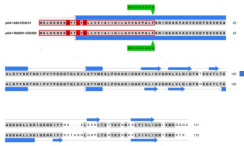
```
5V0M_A|PDBID|CHAIN|SEQUENCE      MERYYRQK--GTPKTYDKNKLKQNKYFNVTLSKVSP---
5VZ3_A|PDBID|CHAIN|SEQUENCE      MERYYRQK--GTPKTYDKNKLKQNKYFNVTLSKVSP---
1QVE_A|PDBID|CHAIN|SEQUENCE      VSDIFSQD--GSCPANTAATAGIEKDTDINGKYVAK---
1RG0_A|PDBID|CHAIN|SEQUENCE      VSDIFSQD--GSCPANTAATAGIEKDTDINGKYVAK---
2H12_A|PDBID|CHAIN|SEQUENCE      VTEYYLNH--GKWFENNTEAGVASSFTDIKGKYVKE---
5CFV_A|PDBID|CHAIN|SEQUENCE      KNKHMNAD--TDYSIAEAAPNKGETAMTINGPWAWSNI
5IHJ_A|PDBID|CHAIN|SEQUENCE      KNKHMNAD--TDYSIAEAAPNKGETAMTINGPWAWSNI
5VAW_A|PDBID|CHAIN|SEQUENCE      KNKHMNAD--TDYSIAEAAPNKGETAMTINGPWAWSNI
R20921                           ALSYYSDT--NKIPVTPDQGTGLSVLETYMESLPDK---
630                              ALSYYSDT--NKIPVTPDQGTGLSVLETYMESLPDK---
1T92_A|PDBID|CHAIN|SEQUENCE      LDMYKLDN--SRYPTTEQGLQAL-VSAPSAEPHARN---
3G20_A|PDBID|CHAIN|SEQUENCE      LDMYKLDN--NRYPTTEQGLRAL-VSKPTVQPEPRN---
2OPD_A|PDBID|CHAIN|SEQUENCE      SKQFILKNPLDDNQTIKSKLERFVSGYKMNPKIAEK---
2OPE_A|PDBID|CHAIN|SEQUENCE      SKQFILKNPLDDNQTIKSKLERFVSGYKMNPKIAEK---
6BBK_A|PDBID|CHAIN|SEQUENCE      AESAILEG---KEIVSSATPKDTQVDIGFTESTLLG---
3S0T_A|PDBID|CHAIN|SEQUENCE      VAQLVQLG----KLTPDEARNGISGDYIGIGGAITSS---
3HRV_A|PDBID|CHAIN|SEQUENCE      ANGLVSLG----KVSADEAKNPTGTAMGIFSFPRNS---
3SOJ_A|PDBID|CHAIN|SEQUENCE      IQNDINNN-----LDISQQTYDTPTGVTVTGSTSG----
4IXJ_A|PDBID|CHAIN|SEQUENCE      SSKDRNEVIKEVLENKDGKYFEIEPKCKSGGIYSATFD
```



PDB

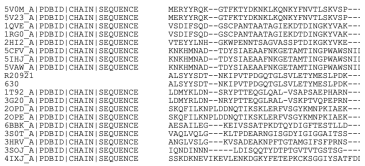


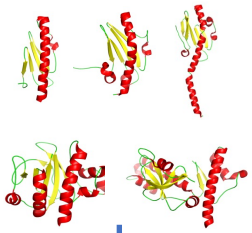




Search for homologues

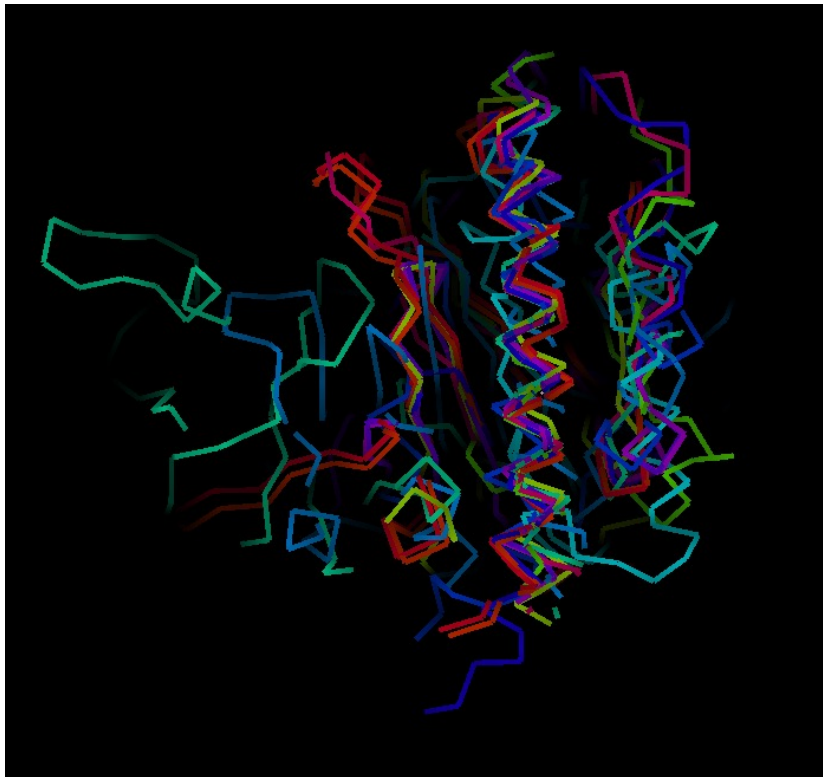
ClustalW





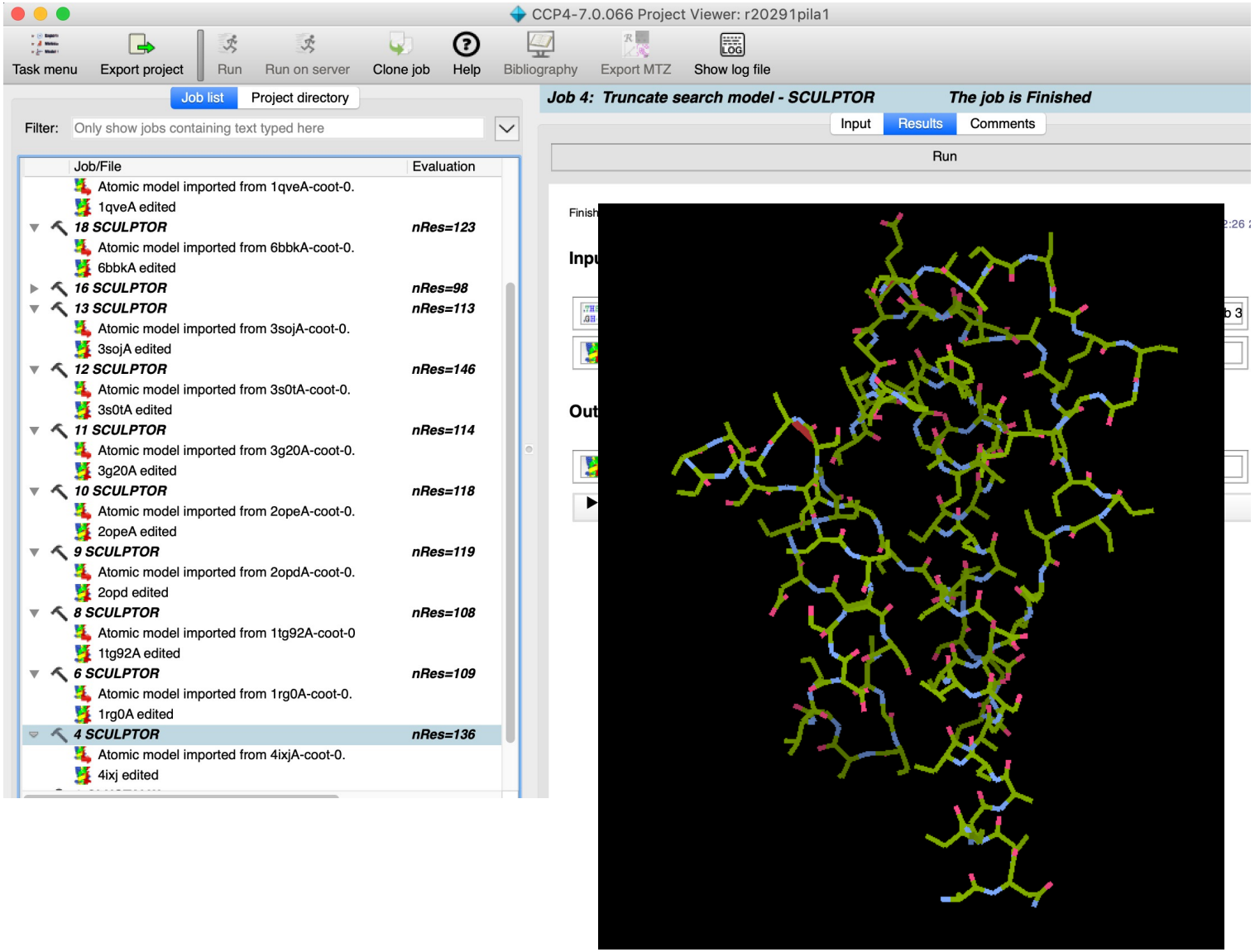
Model preparation

PDB

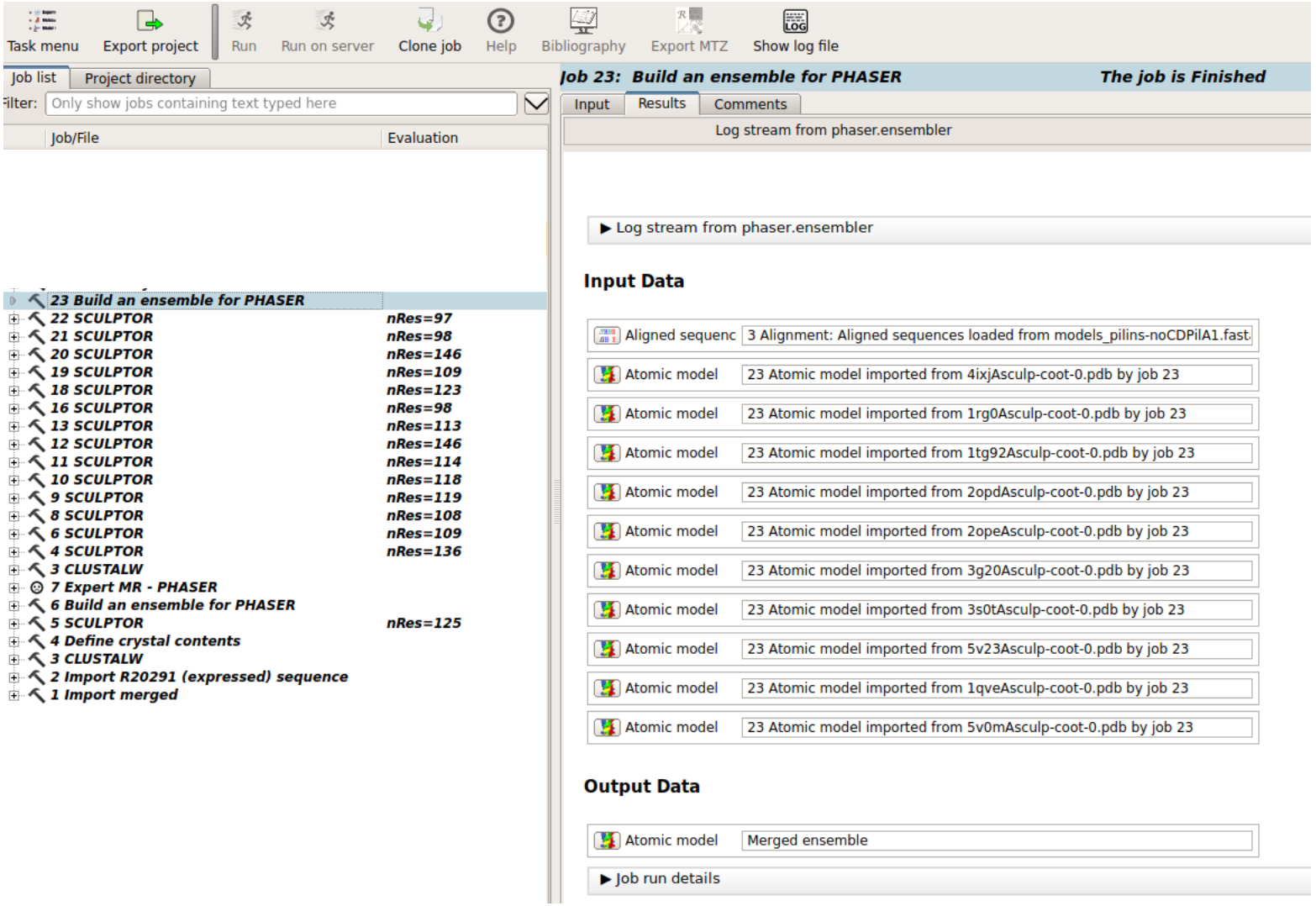


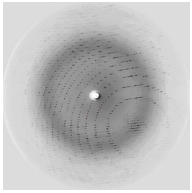
Sculptor

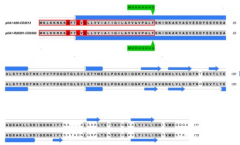
Model preparation



Model preparation

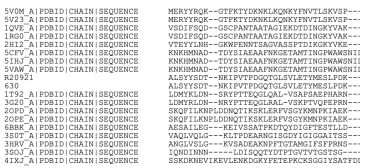






Search for homologues

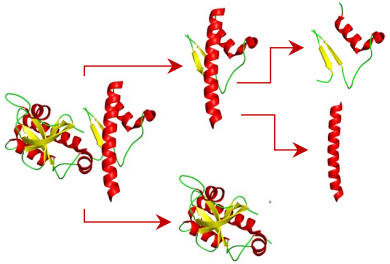
ClustalW

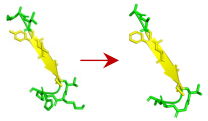


PDB

Sculptor

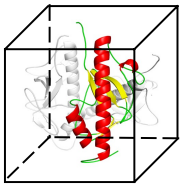
Model preparation





Phaser

Molecular replacement



Phaser

Molecular replacement

CP4-7.0.066 Project Viewer: pila1-r20291

Export MTZ | Show log file

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

Input | Results | Comments

Input data | Simple options | Extra steps | Keywords

PHASER

2 /pila1/4

ude (F) ML target I

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

21 Define crystal contents

15 Edited search model Copies: 3

target: **read from header of PDB**

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

Input | Results | Comments

Current soln. | All solns | Comp/data | COM file | Search tree | Plots from PHASER output | Run

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

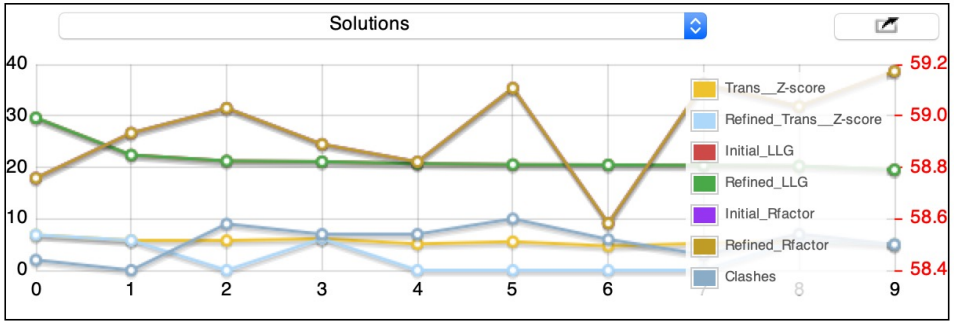
s to view jobs and files

n 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

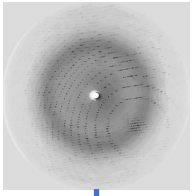


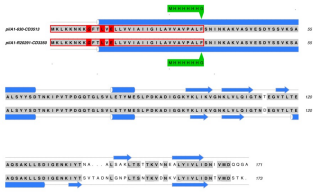
Space group	Trans. Z-score	Refined Trans. Z-score	Initial LLG	Refined LLG	Initial Rfactor	Refined Rfactor	Clashes
P 43 21 2	6.87	6.87	29.61	29.61	58.76	58.76	2.00
P 43 21 2	5.80	5.80	22.39	22.39	58.93	58.93	0.00
P 43 21 2	5.77	0.00	21.24	21.24	59.03	59.03	9.00
P 43 21 2	6.14	6.14	21.09	21.09	58.89	58.89	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

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*	

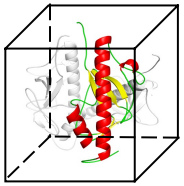
```
Guide to eLLG values
--------------------
eLLG        Top solution correct?
<25         -no
25-36       -unlikely
36-49       -possibly
49-64       -probably
>64         -yes
```





Arcimboldo

Molecular replacement



Model Building

Buccaneer

Coot

Refinement

Refmac

Completion
Validation

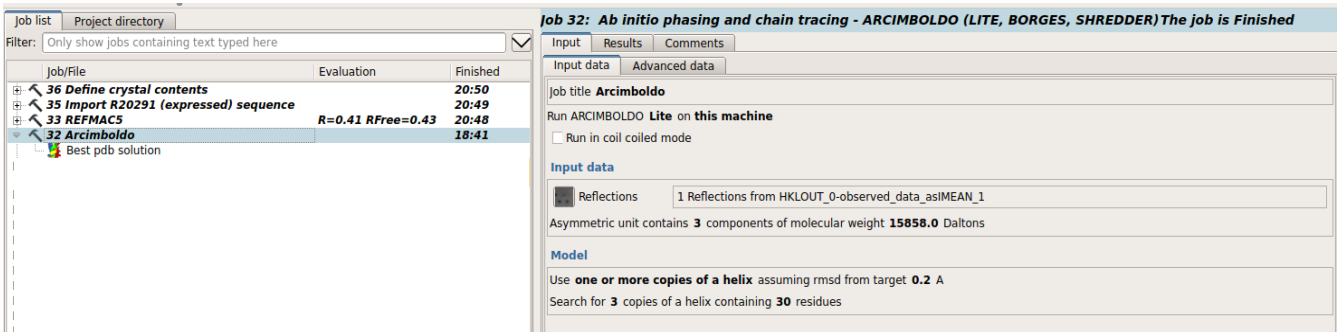
Molprobity
Coot

Deposition

PDBe

Arcimboldo

Molecular replacement



ARCIMBOLDO

Summary

Search and expansion

Backtracking

Instruction file

Log file

Citations

Support

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			

Show All | Hide Not Relevant

Backtracking

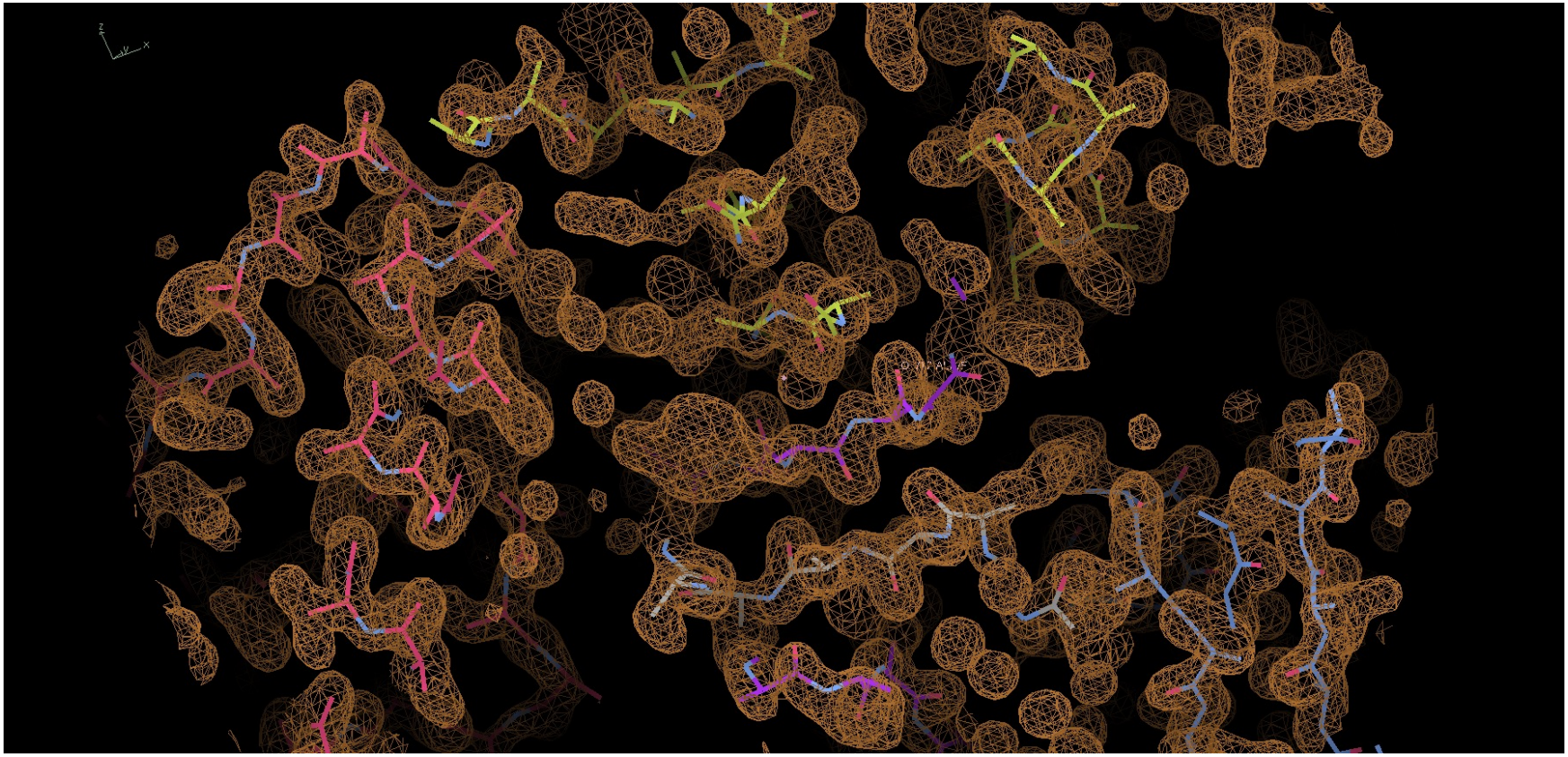
The current best solution is: ensembleIDxx2FR1_8-1.pdb with FINALCC: 42.32 and n. residues traced 378
file is: /home/psalgado/Dropbox/TFP/MR-PS/r20291MRi2/CCP4_JOBS/job_32/8_EXP_LIBRARY/4/0/ensembleIDxx2FR1_8-1.pdb

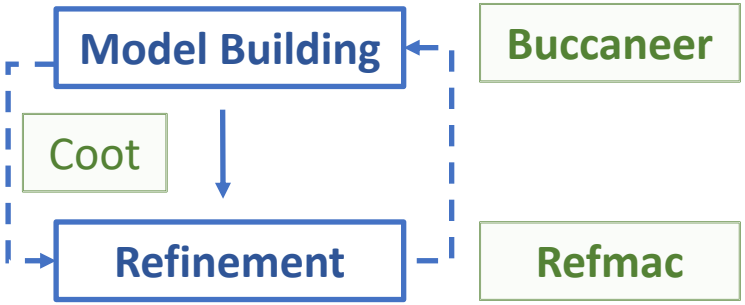
FRF: Pos. in Rank: **85** LLG: **49.30** ZSCORE: **2.58** Top LLG in Cluster **(0, 2)**: **52.60** Top ZSCORE in Cluster **(0, 2)**: **3.66**
REFINEMENT ROTATION AND MODEL
FTF: Pos. in Rank: **1** LLG: **55.90** ZSCORE: **7.97** Top LLG in Cluster **(0, 2)**: **55.90** Top ZSCORE in Cluster **(0, 2)**: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

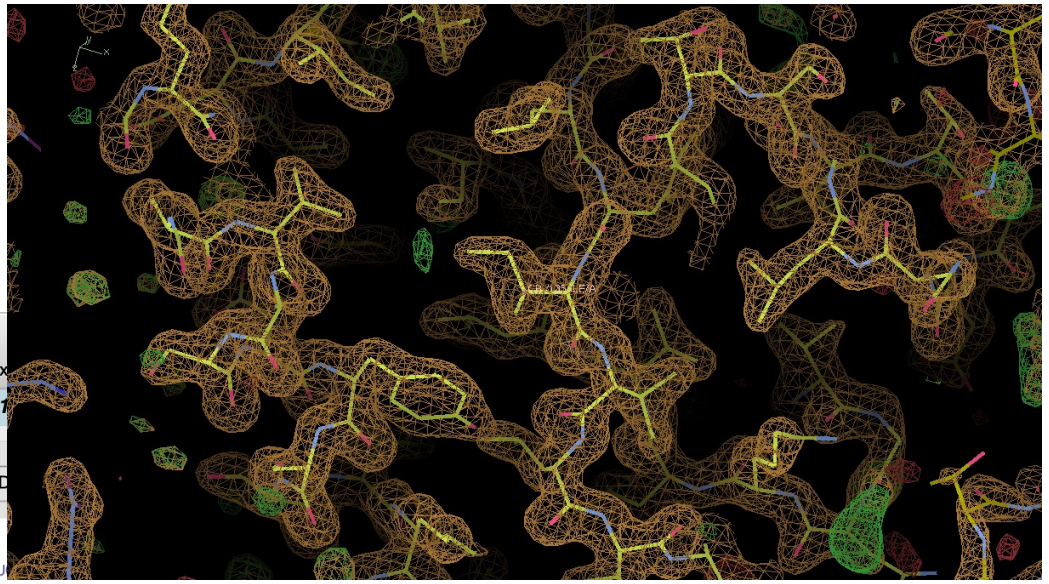
It seems you have a good solution!
Here you can find the best solution and map for further refinement.

Arcimboldo

Molecular replacement







Task menu | Export project | Run | Run on server | Clone job | Help | Bibliography | Ex

Job list | Project directory

Filter: Only show jobs containing text typed here

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	

Right mouse click for options to view jobs and files

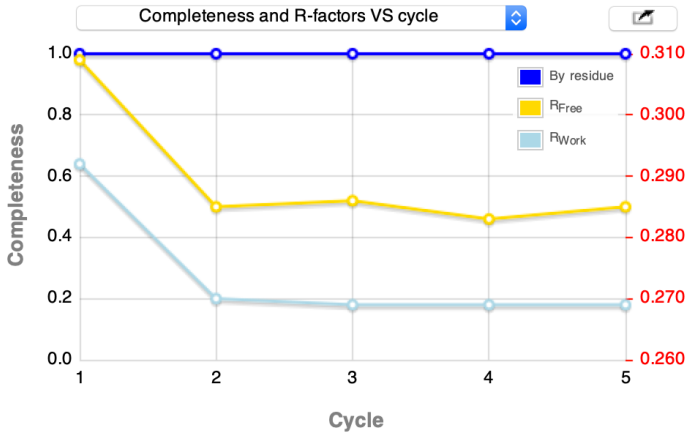
Results

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.

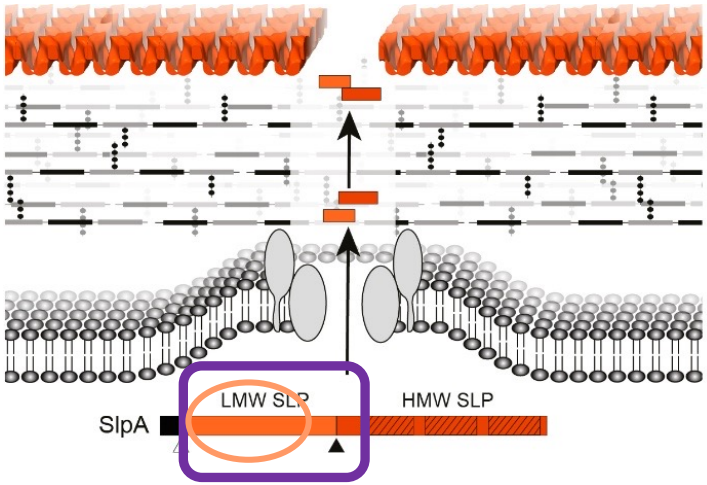
The refinement R-factor is 0.27, and the free-R factor is 0.28. The RMS bond deviation is 0.010 A.
On the basis of the refinement statistics, the model is approaching completion.

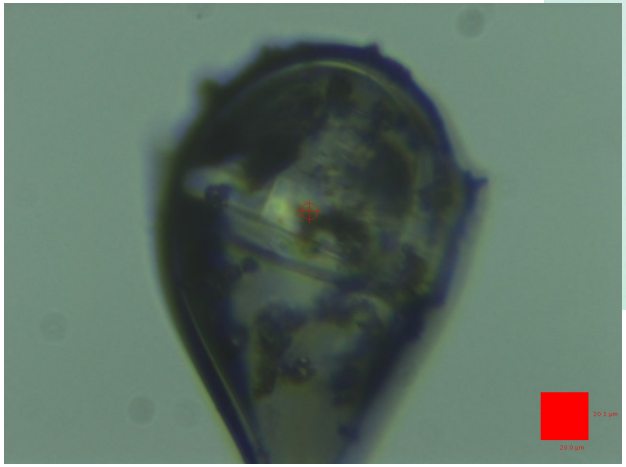
Completeness by residue	1.0
Completeness by chains	0.93
Number of chains	3
Residues built	410
Residues sequenced	410
Longest fragment	140
Number of fragments	3
R_{Work}	0.269
R_{Free}	0.285
RMS_{Bonds}	0.01
RMS_{Angles}	1.828

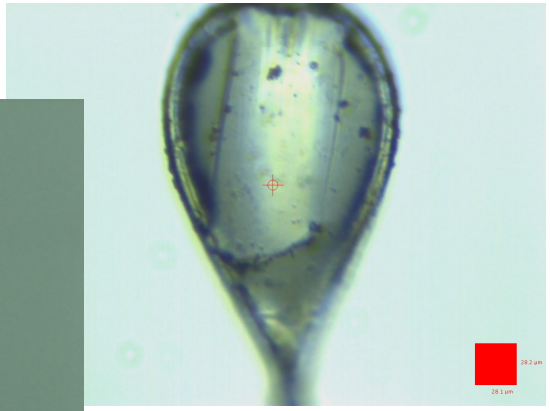


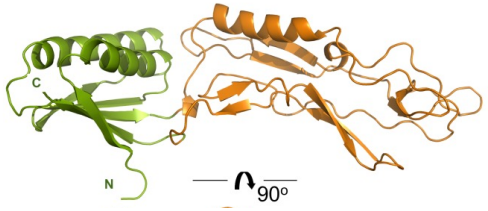
▶ Detailed progress by iteration

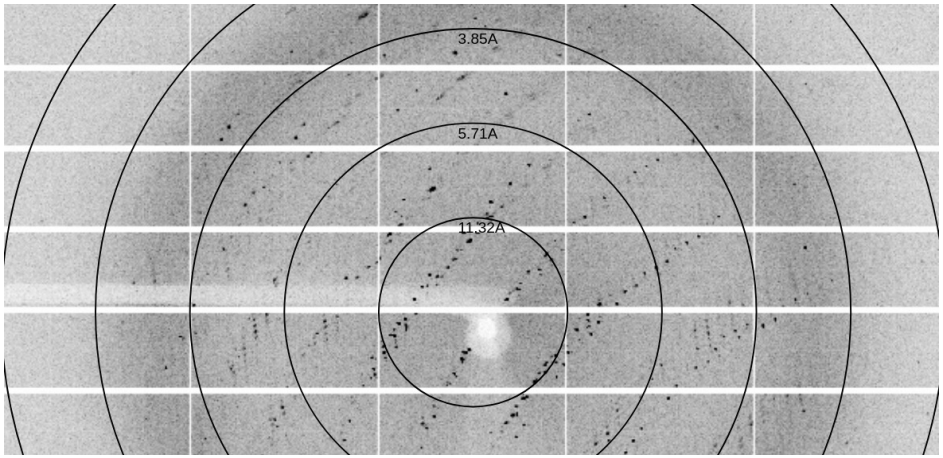
Case study 2: *C. difficile* S-layer



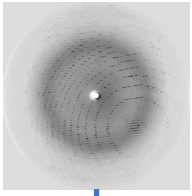


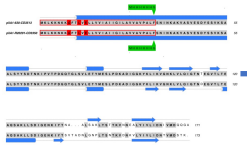






	Overall	High res.
Space group	C2	
Unit cell		
a, b, c (Å)	173.25, 29.58, 144.63	
α=γ, β (°)	90.0, 94.2	
Resolution (Å)	144.24 – 2.57	2.57-2.69
I/σI	10.2	2.2
Half-set correlation CC(1/2)	0.993	0.838
Completeness %	99	98





Mr Bump Phaser

Molecular replacement

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

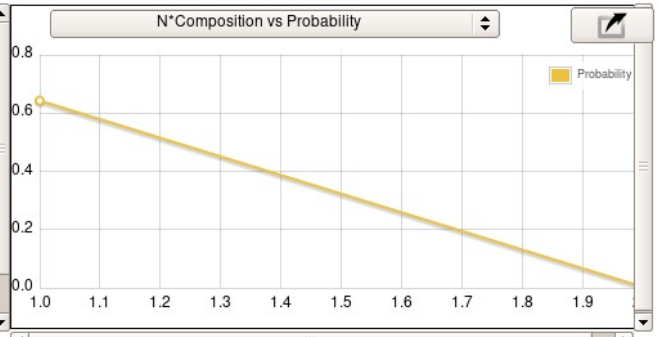
▶ Analysis of composition and data

▶ COM file for this run

▶ Search strategy employed by PHASER

▼ Plots from PHASER output

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 #1 (SearchModel)
Translation Function Component #1 (SearchModel)
Refinement After Placing Component #1 (SearchModel 3
Rotation Function Component #1 (SearchModel)
Translation Function Component #1 (SearchModel)
Refinement After Placing Component #1 (SearchModel 2



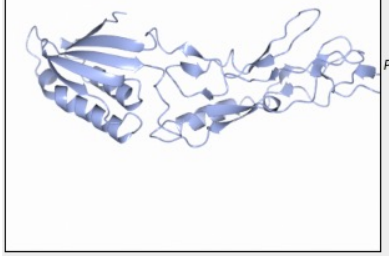
Results

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

Picture of structure 1



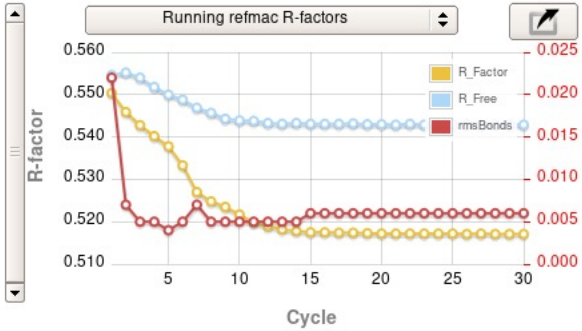
Picture of stru

View in CCP4mg | View in Coot

▼ Refinement

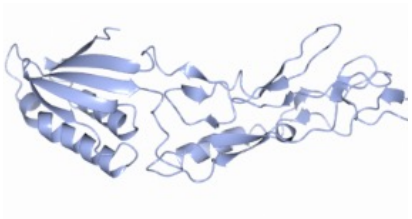
Statistic	Value
Resolution	86.54-2.60
No. reflections all/free	23167 / 1149
R-factor/R-free	0.517 / 0.543
RMS Deviations	
Bonds	0.0057
Angles	1.732
Chain mean B (No. atoms)	
AAA	32.7(3109)

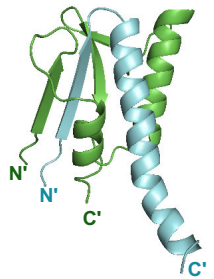
Download



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

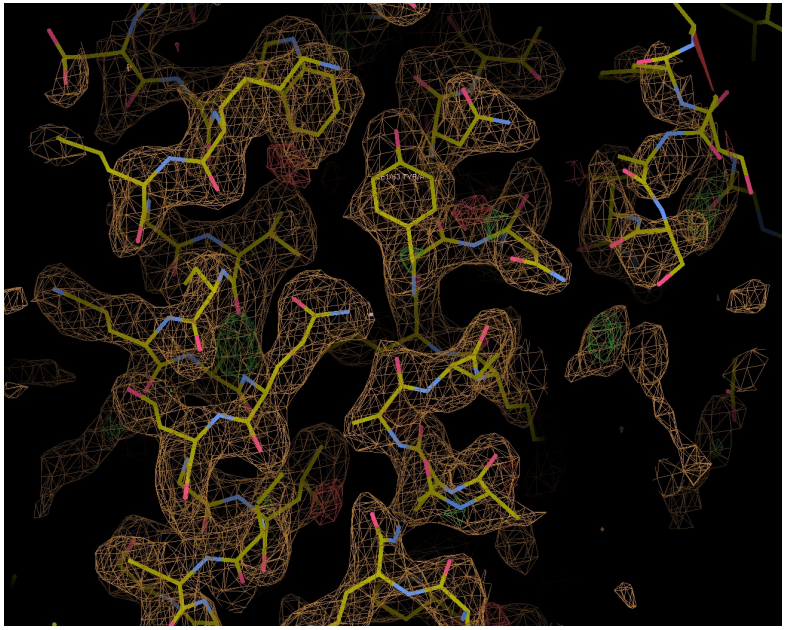
Search: 7 Model refined by Prosmart/Refmac

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 al

▼ 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

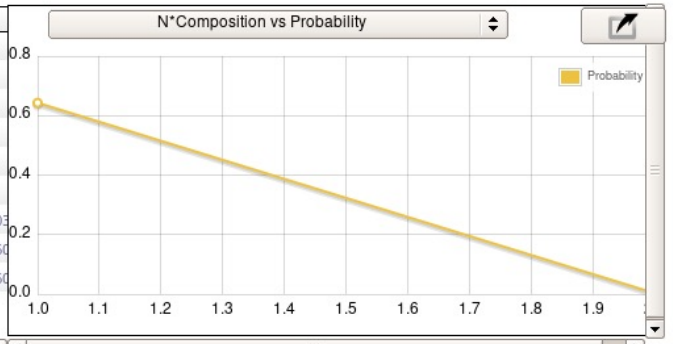
▶ Analysis of composition and data

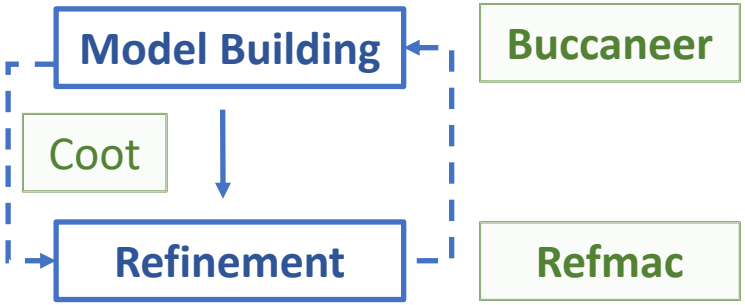
▶ COM file for this run

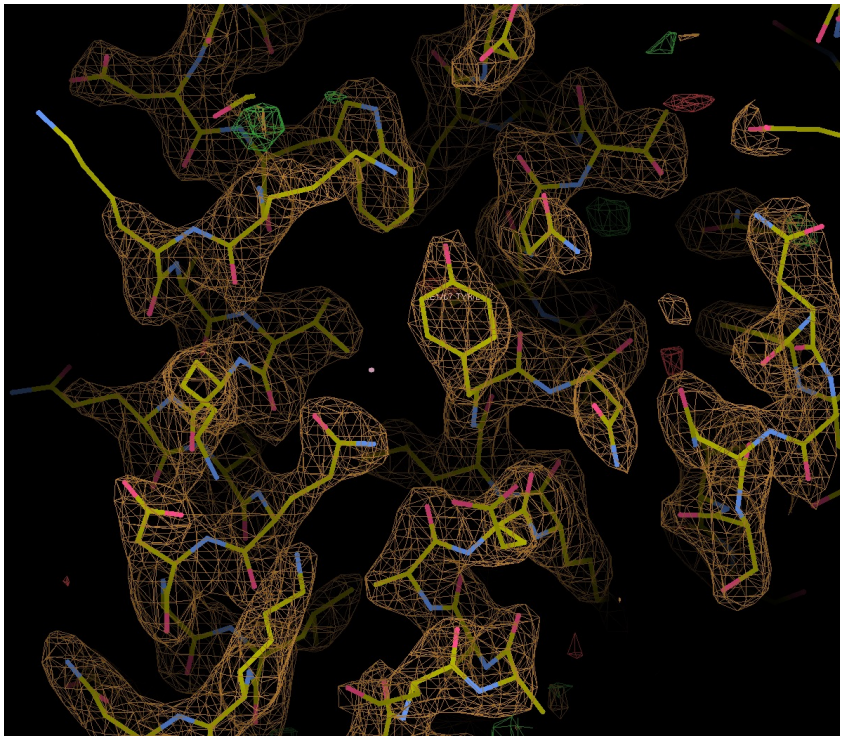
▶ Search strategy employed by PHASER

▼ 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 4.0
- Refinement After Placing Component #2 (SearchModel 2.6
- Refinement After Placing Component #2 (SearchModel 2.6







Results

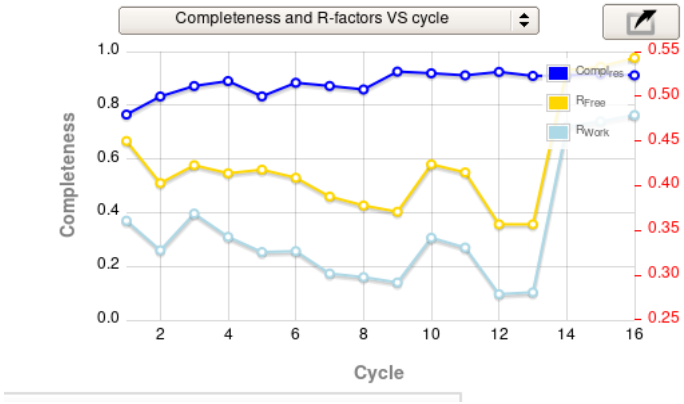
The final model is taken from cycle 13 as this had the lowest free-R factor.

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.

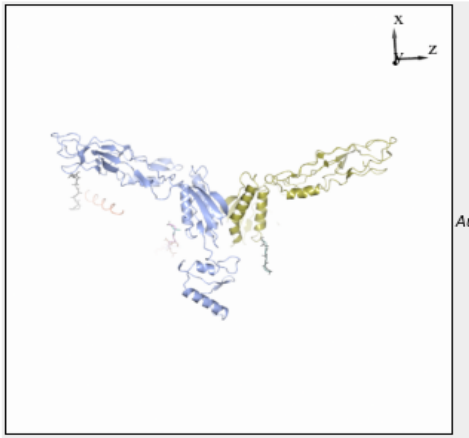
The refinement R-factor is 0.28, and the free-R factor is 0.36. The RMS bond deviation is 0.007 A.
On the basis of the refinement statistics, the model is approaching completion.

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

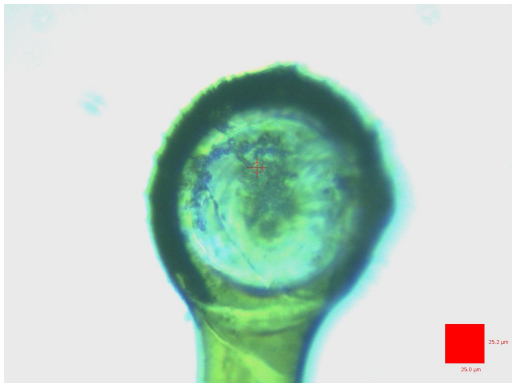
▶ Alignments for model and cell content sequences



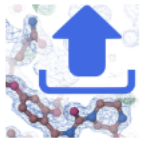
Autobuilt structure

View in CCP4mg | View in Coot

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



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



Import & Replace

Import data from your device or CCP4

job description: import-n-replac

Structure revision R0165.01: asu [DI

Data to import and replace in revision

Import data from local file system

Reflection data Browse

Phases Browse

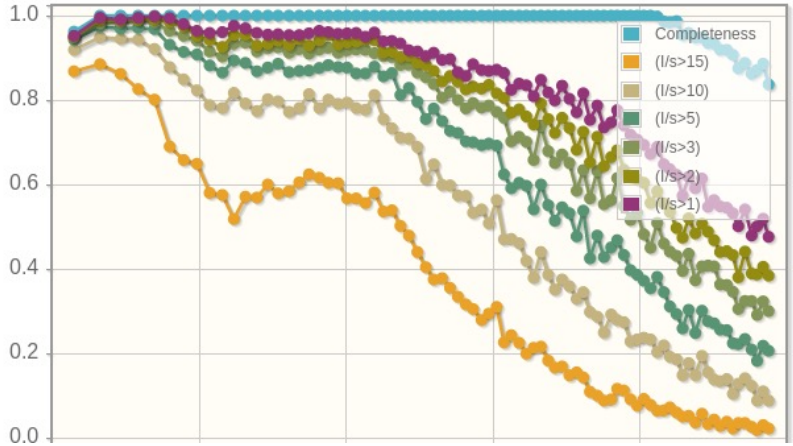
Atomic coordinates Browse

Ligand descriptions Browse

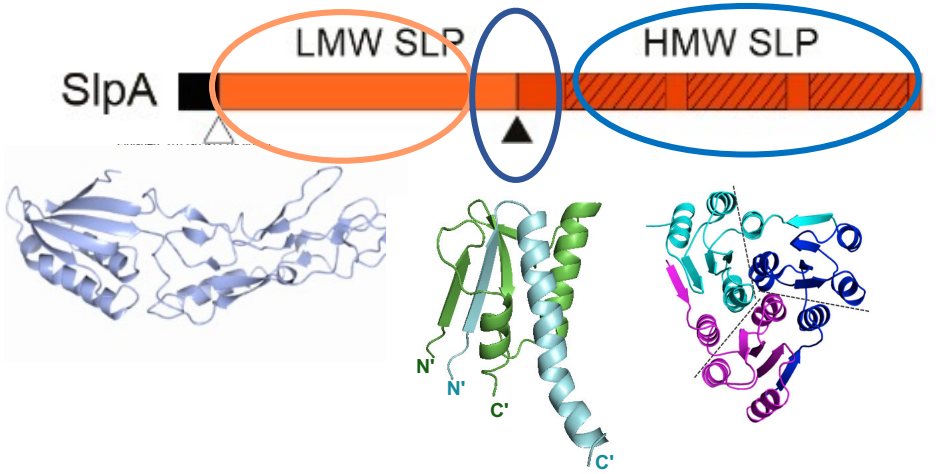
Data analysis (CTruncate)

Graph Data

- ▾ Intensity Completeness analysis
  - Completeness & (I/sigI)>N v resolution
  - Completeness & Rstandard v resolution
- ▸ Wilson plot
- ▸ Intensity anomalous analysis
- ▸ Intensity statistics
- ▸ Cumulative intensity distribution
- ▸ L test for twinning
- ▸ Acentric Moments of I
- ▸ Centric Moments of I
- ▸ Phil plot



Model preparation



[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.

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)

Model modification

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

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/

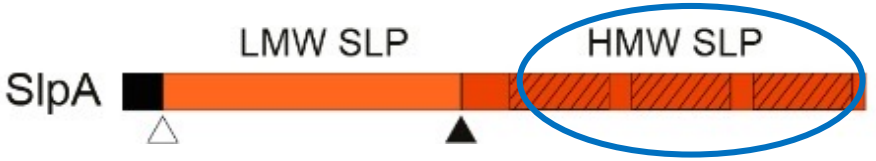
Assigned name : [0169-01] rvB_CWB2s-SM630_B /model/protein/
Estimated seqId : 71.1%

Coordinates UglyMol | ccp4mg | Display

0169-01_rvB_CWB2s-SM630_B.pdb Export

Molecular replacement

Phaser



Molecular Replacement with Phaser

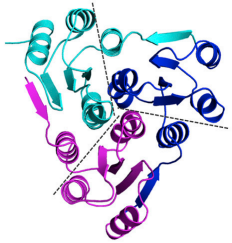
job description: phaser MR

output id: phaser-mr

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

Use current structure: as fixed model

Resolution range (Å): auto *to* auto



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.

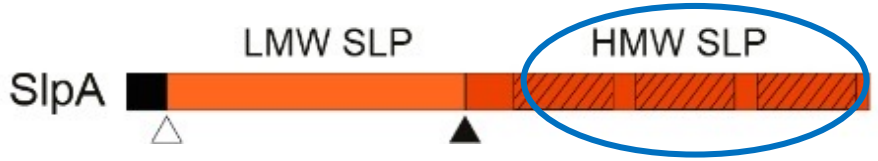
Please consider that phasing scores are lower if, as in this case, not all copies of monomeric units are found. Try to fit the remaining copies in subsequent phasing attempts.

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.

Purge translation peaks	on	Cutoff (%)	75	Max number	40
Purge refinement peaks	on	Cutoff (%)	75	Max number	20

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

[0254] Asymmetric Unit Contents

Suggested ASU contents

	N_{copies}	Structural unit components	Type
1	2	[0164-01] revB1seq /sequence/protein/	PROTEIN
		Total residues/weight:	

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

[0254] Results

Cell volume: 1872135.62 Å^3

Molecule fitting statistics

N_{trial}	Matthews	% solvent	$P_{matthews}$
* 1	3.23	61.99	0.985
2	1.62	23.98	0.015

▾ 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

[0254] Verdict

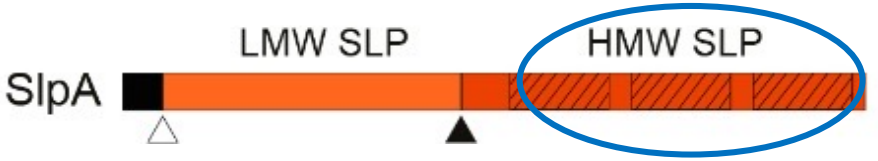
The estimated solvent fraction is within the usua

The suggested composition of ASU corresponds to usual values of solvent fracti

In general, composition of ASU remains a hypothesis until structure is solved. Th
estimations of solvent content may have a negative impact on phasing and dens

Molecular replacement

Phaser



CCP4 v.8.0.000; CCP4 Cloud v.1.7.002

[0255] Molecular

- Phaser (2.8.3)
- Electron Density Cal

[0255] Verdict

Phasing summary

LLG	1037
TFZ	31
R_{free}	0.49(
Found copies	2

Assumed total number o

In general, correctness o

[0255] Output Str

Assigned name: [0255-

Structure and electr

Report | Main Log | Service Log | Errors

Script

MR Result

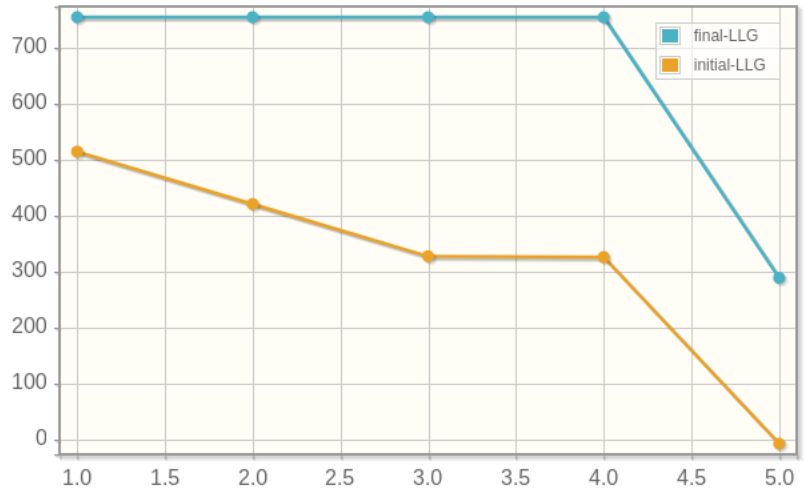
```
** SINGLE solution

** 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
```

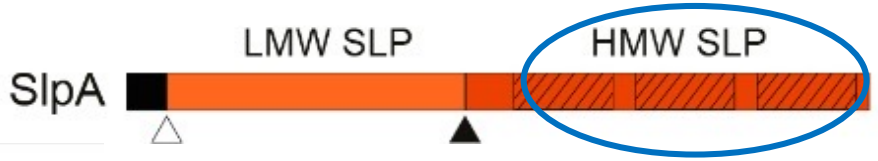
Graph Data

- #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

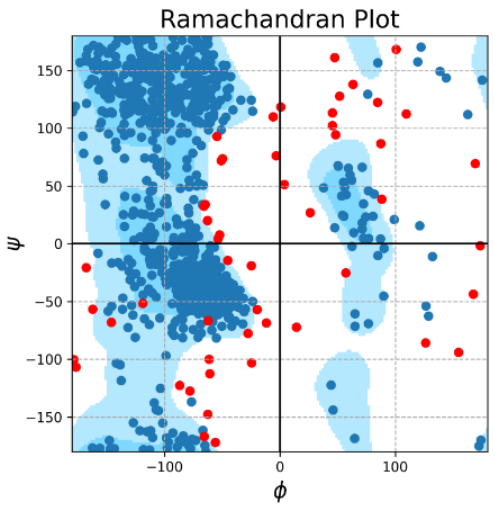
Assigned name: [0257-01] modelcraft /structure/

Structure and electron density UglyMol | ccp4mg | ViewHKL | Display

[0257] Structure Revision

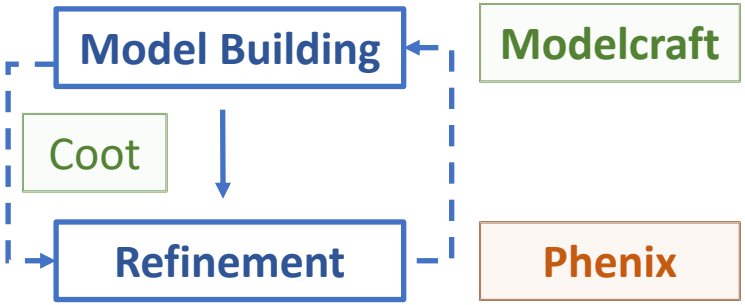
[0257] Quality Assessment

- B-Factors Analysis
- Electron Density Fit Analysis
- Molprobity Analysis
- Ramachandran Plot



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

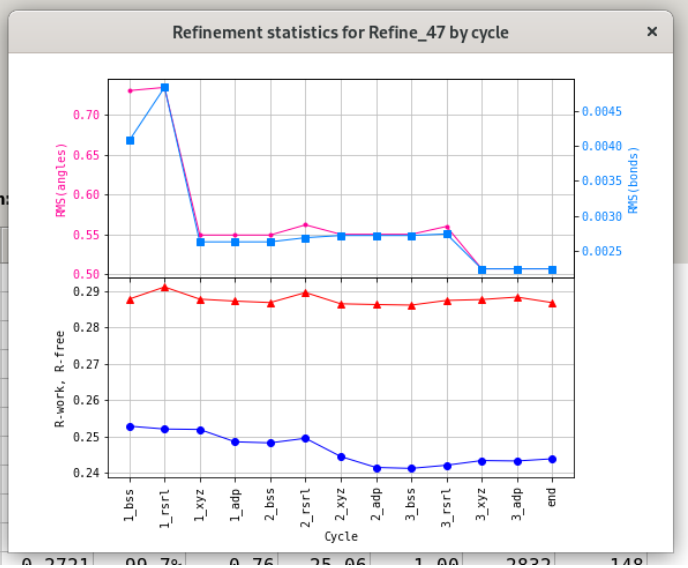


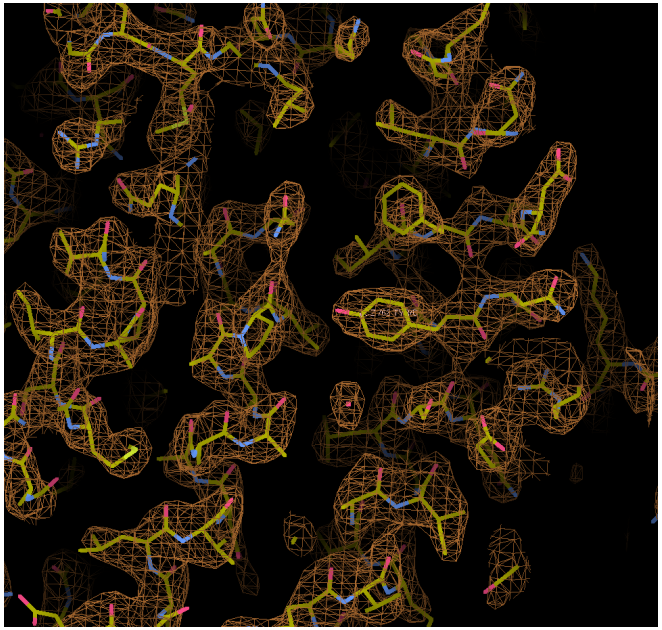
Before and after refinement:

	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



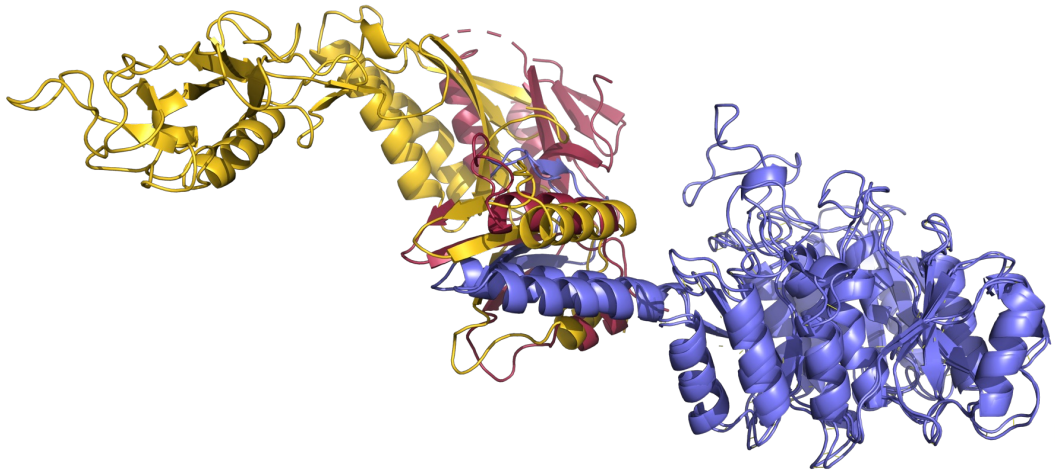


Basic statistics for rev189_refine_47.pdb:

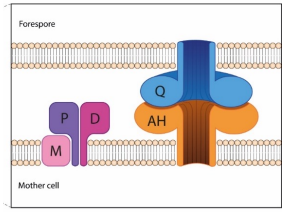
Ramachandran outliers:	0.19%	(Goal : < 0.2%)	Ramachandran favored:	96.76% (Goal : > 98%)
Rotamer outliers:	1.28%	(Goal : 1%)	C-beta outliers:	0 (Goal : 0)
Clashscore:	10.75		Overall score:	1.83

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

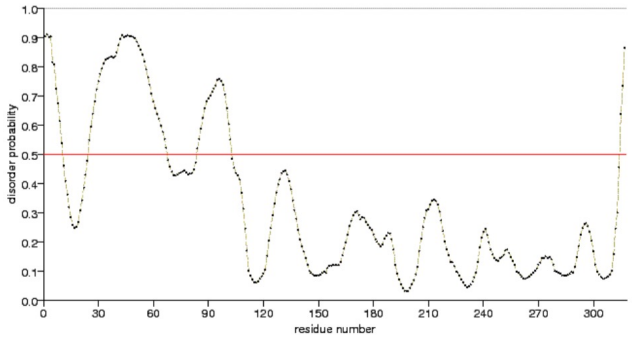
whole (N = 1080):	-1.50 (0.24)
helix (N = 409):	-0.93 (0.24)
sheet (N = 134):	-0.49 (0.42)
loop (N = 537):	-1.08 (0.25)

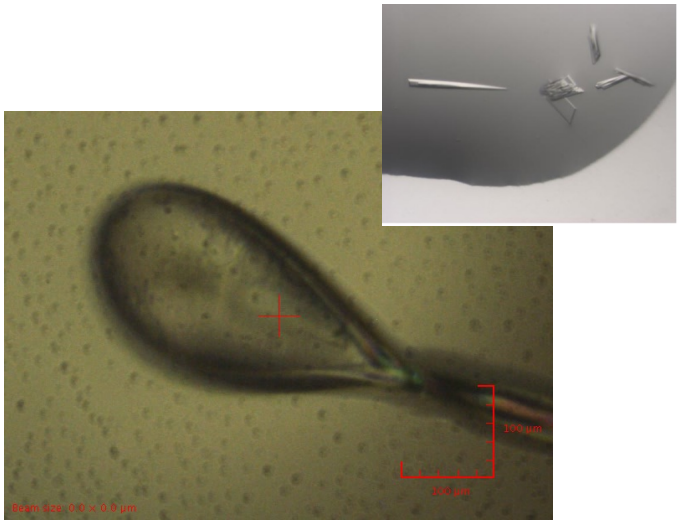


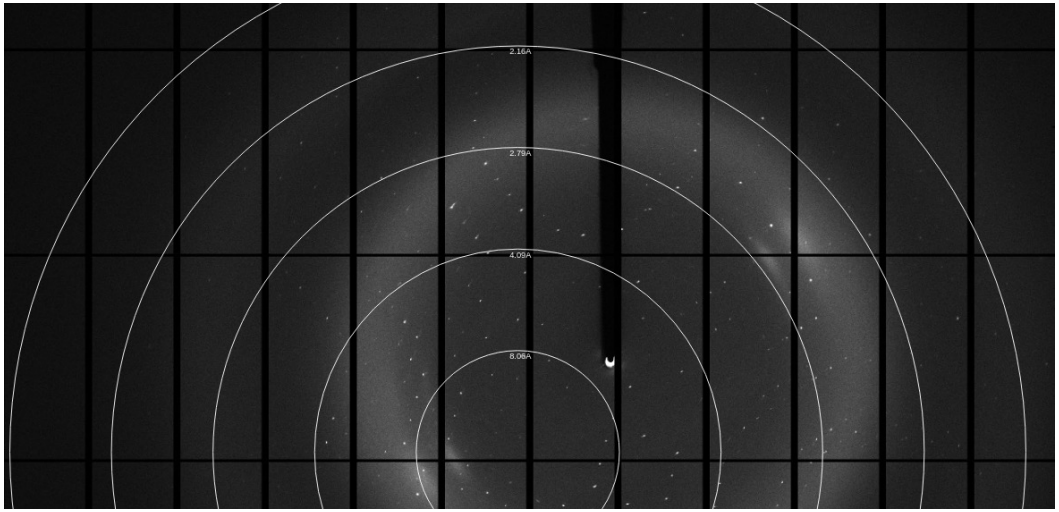
Case study 3: *C. difficile* PG hydrolase



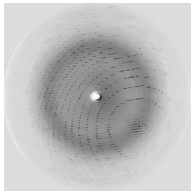

```
1   GAMGNQDDFL  KFLVNSSYPE  AKVEGNDTEN  KKNNKNKETS  KENKEESKEE  50
51  NTKSKDASKV  DNKKESEKEY  IKLYVGKENV  PDIESKNSDT  TETNTTSSSD  100
101 YKDDLRVTKE  NPRILIYHTH  GCETYSNSPD  GNYHSRDKKN  SVMEVGSALT  150
151 SALDSKGWGV  VHTTKYHDYP  SYNNSYASSL  KTIQSILPKY  NSVDIAIDLH  200
201 RDARDLTNPA  TKEKDHLKYT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

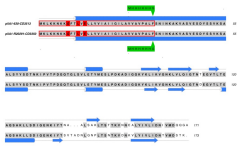


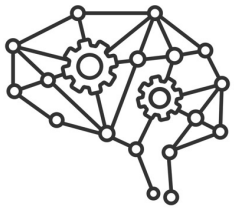




	Overall	High res.
Space group	P2 2_1 2_1	
Unit cell		
a, b, c (Å)	47.62, 51,88, 81.80	
α=β=γ (°)	90.0	
Resolution (Å)	43.81 - 1.56	1.59 - 1.56
Rmerge	0.148	1.644
I/σI	7.1	1.2
Half-set correlation CC(1/2)	0.996	0.548
Completeness %	100	99.9







Model generation

AlphaFold2
RosettaFold

Edit B-factors

Model preparation

Model Building

Buccaneer

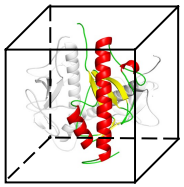
Coot

Refinement

Refmac

Phaser

Molecular replacement



Completion
Validation

Molprobity
Coot

Deposition

PDBe

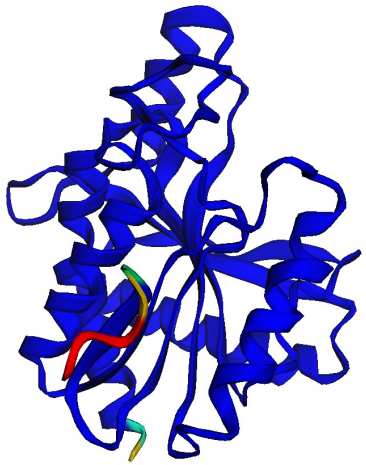
Model generation

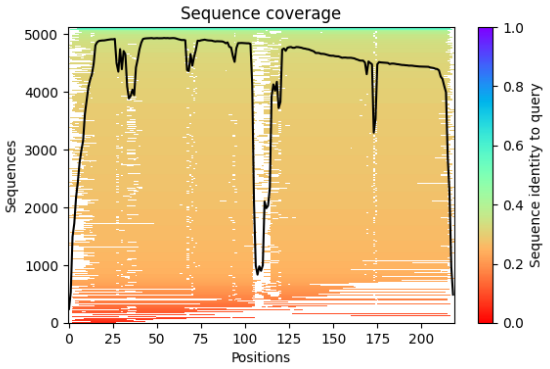
AlphaFold2

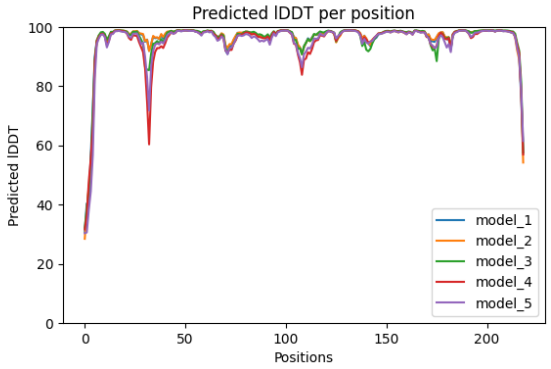
AlphaFold2.ipynb

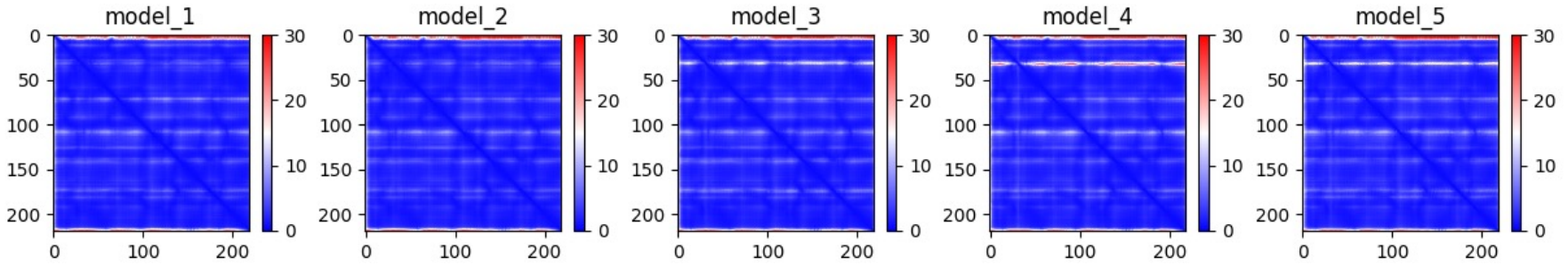
File Edit View Insert Runtime Tools Help Cannot save changes

Code + Text Copy to Drive









Model preparation

Molecular replacement

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

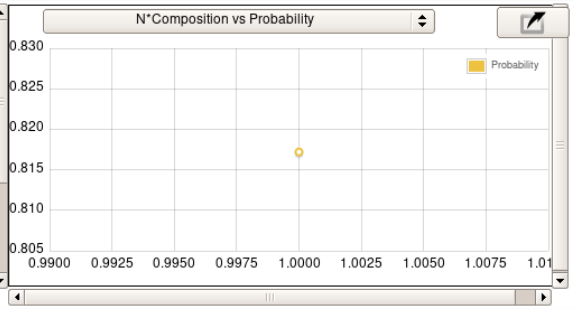
▶ Analysis of composition and data

▶ COM file for this run

▶ Search strategy employed by PHASER

▼ Plots from PHASER output

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 #1 (Ensemble_1)
Translation Function Component #1 (Ensemble_1)
Refinement After Placing Component #1 (Ensemble_1 5...
Refinement After Placing Component #1 (Ensemble_1 1...
Refinement After Placing Component #1 (Ensemble_1 1...
Cycle 1. Rfactor analysis, F distribution v resln

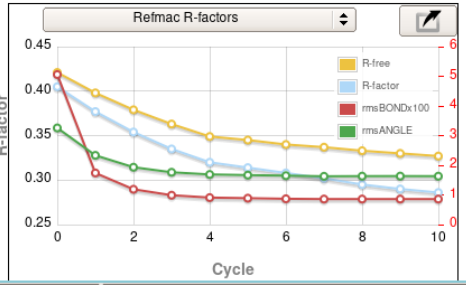


▶ Shift field refinement

▼ 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)

Download



Edit B-factors

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

Input | Results | Comments

Input Data

Job title **Edit B-Factors**

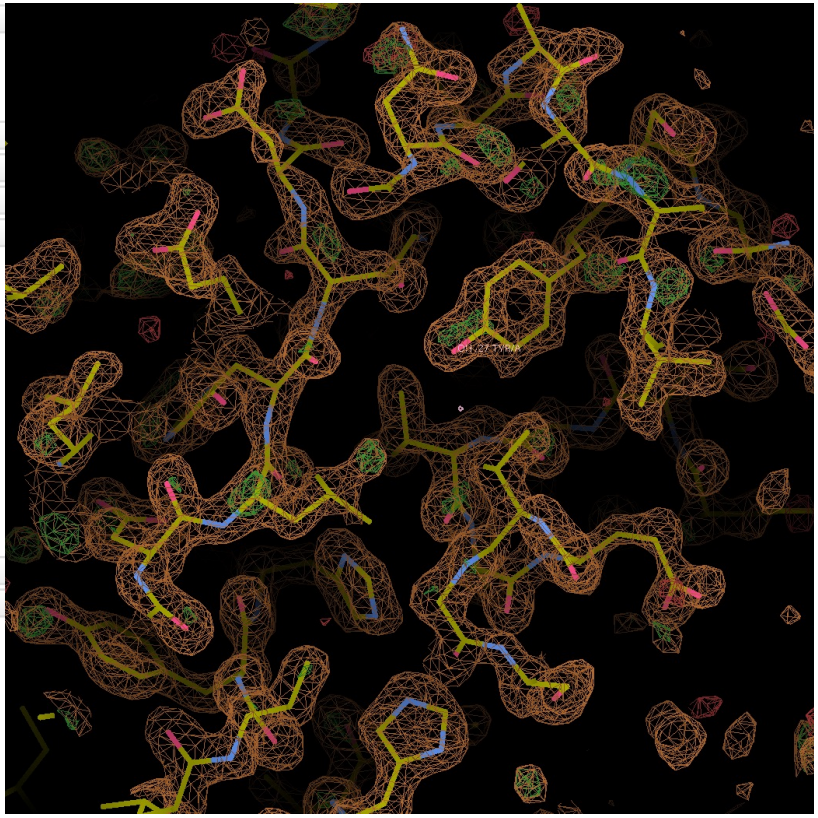
Atomic model

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

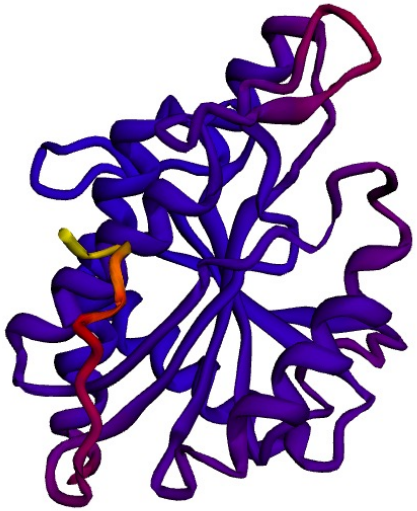


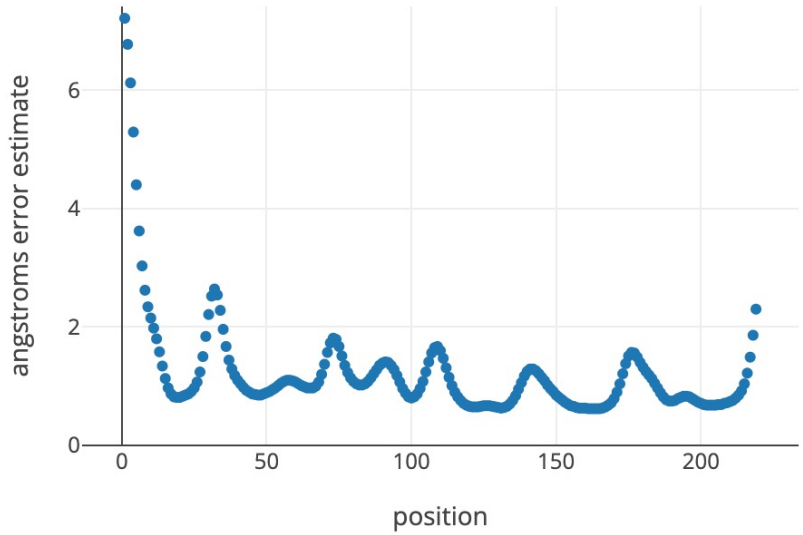
Model generation — RosettaFold

↓

Model preparation — Edit B-factors

Model 4 Model 5





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

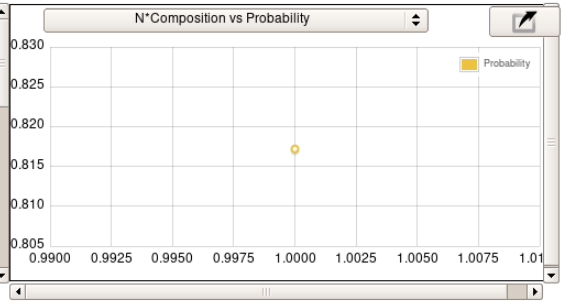
▶ Analysis of composition and data

▶ COM file for this run

▶ Search strategy employed by PHASER

▼ Plots from PHASER output

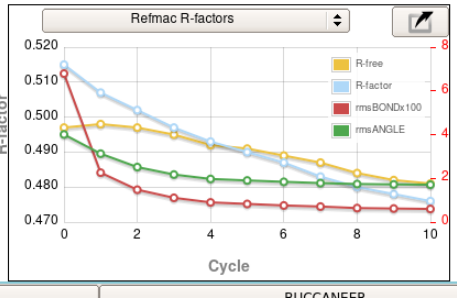
- Cell Content Analysis
- Intensity distribution for Data
- Weighted second moments for Data
- Intensity distribution after Anisotropy Correction
- Weighted second moments after Anisotropy Correction
- Cell Content Analysis
- Cell Content Analysis
- Cell Content Analysis
- Rotation Function Component #1 (Ensemble_1)
- Cell Content Analysis
- Translation Function Component #1 (Ensemble_1)
- Cell Content Analysis



▶ Shift field refinement

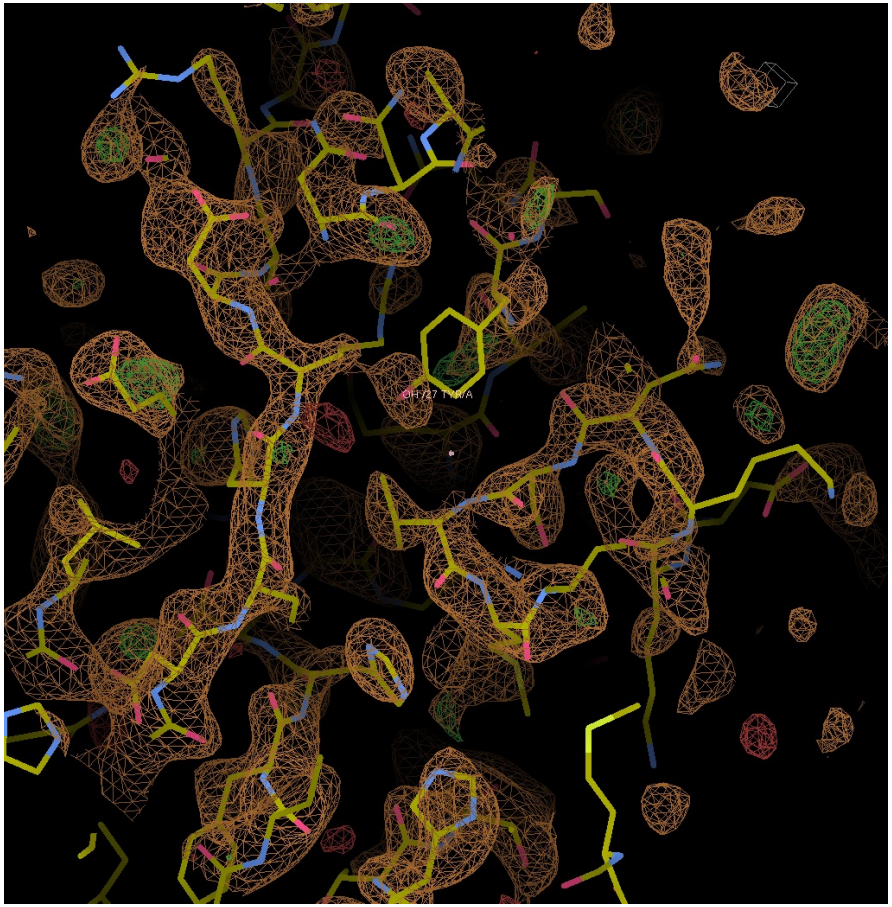
▼ Summary of refinement

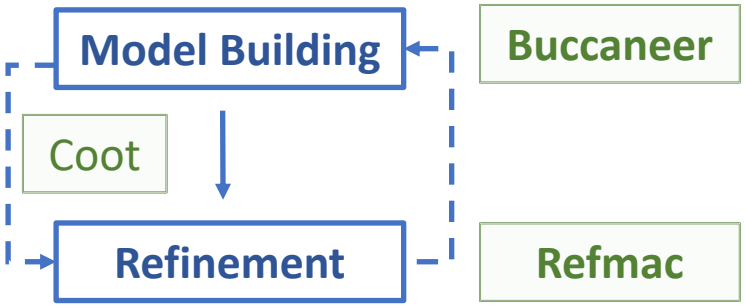
Statistic	Value
Resolution	43.85-1.56
No. reflections all/free	29547 / 1422
R-factor/R-free	0.476 / 0.481
RMS Deviations	
Bonds	0.0060
Angles	1.709
Chain B-factors	*mean B (#atoms)*
AAA	38.3(3404)

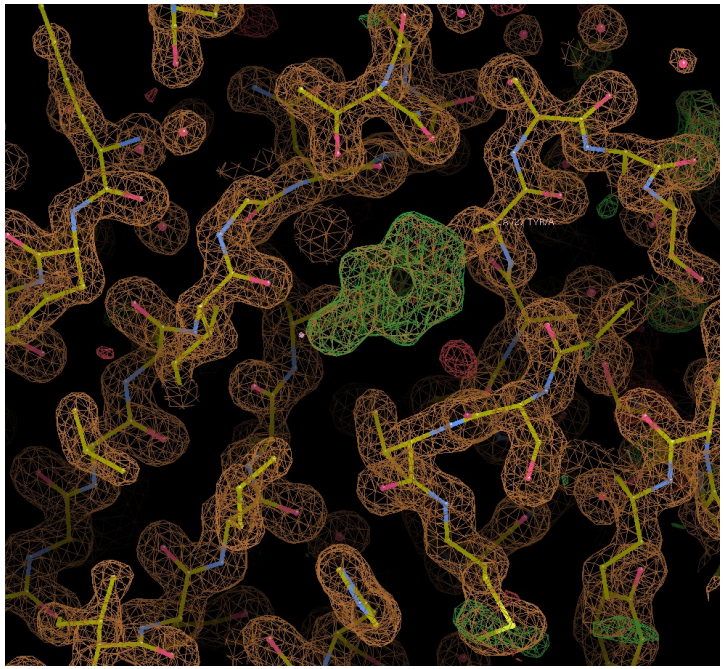


Download

REFMAC5 BUCCANEER







BUCCANEER from phaser RF1 xtal 14

15:20 02-Dec-2021

Results

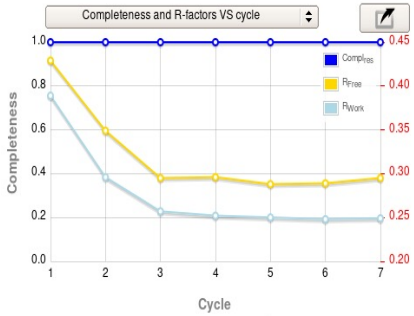
The final model is taken from cycle 5 as this had the lowest free-R factor.

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

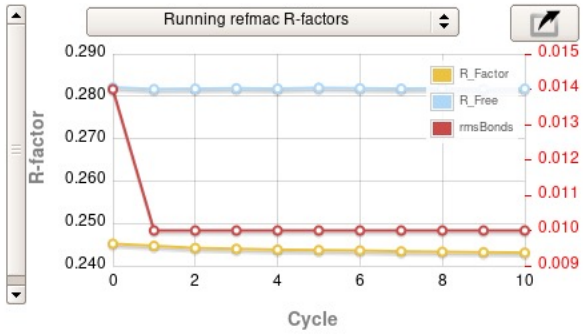
Autobuilt structure

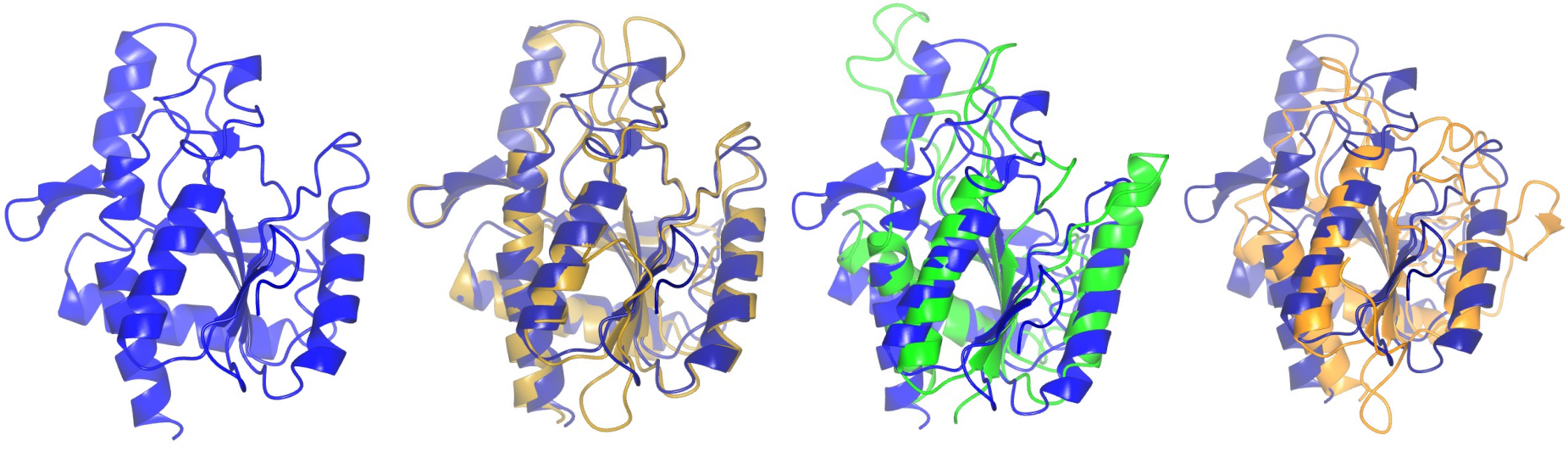
REFMAC5 after bucanner 35 RF1 xtal14

▼ Refinement

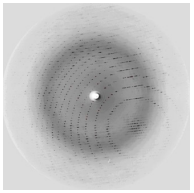
Statistic	Value
Resolution	43.85-1.56
No. reflections all/free	29547 / 1422
R-factor/R-free	0.243 / 0.282
RMS Deviations	
Bonds	0.0097
Angles	1.805
Chain B-factors	*mean B (#atoms)*
AAA	23.2(3284)
CCC	28.5(155)

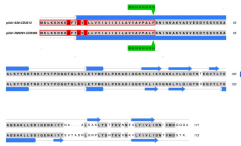
Download





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 (seq ID <20%)	Phaser: TFZ – 3.7 No solution (seq ID <20%)
Buccaneer: 215 aa R = 26% Rfree = 31%	Buccaneer: 214 aa R = 25% Rfree = 29%	-	-
Refmac: R = 25% Rfree = 30%	Refmac: R = 24% Rfree = 28%	-	-



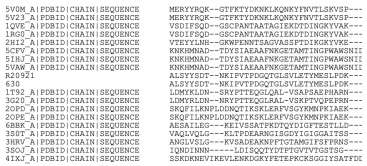


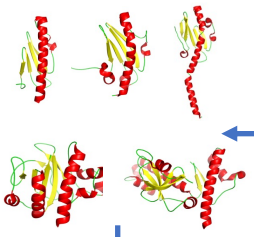
Search for homologues

PDB/ BLAST

AlphaFold2
RosettaFold

ClustalW





PDB
Phyre2
SWISS-MODEL
i-TASSER

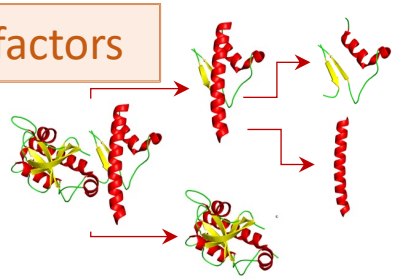
Chainsaw
Sculptor
Ample

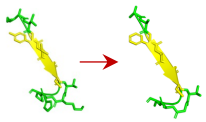
Model preparation

Buccaneer
ShelxE
ARPwARP

Model Building

Edit B-factors





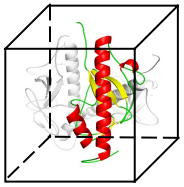
Coot

Refinement

Refmac

Mr Bump
Phaser
Molrep
Arcimboldo

Molecular replacement



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

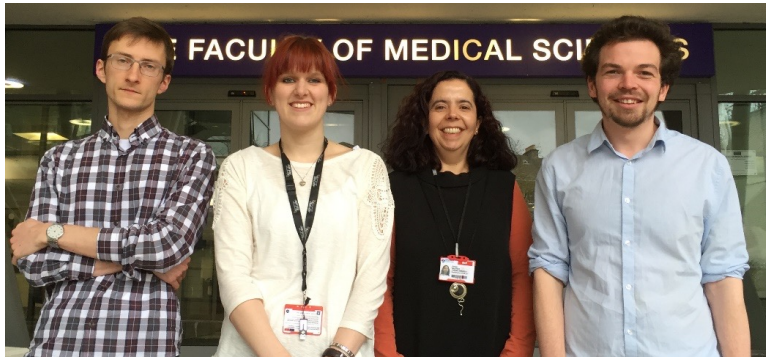
Kamel El Omari
Armin Wagner

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