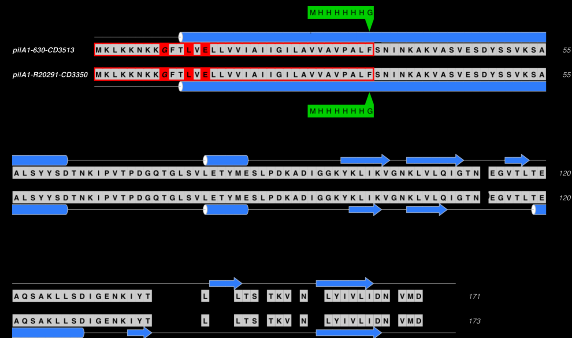
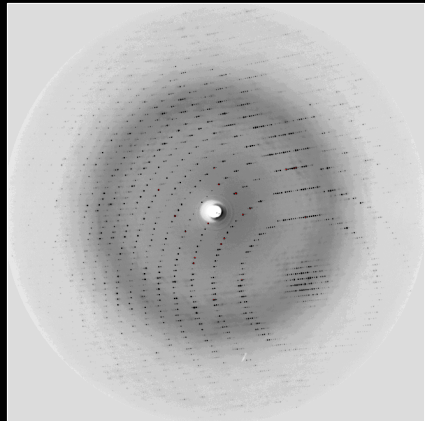


Molecular Replacement case studies

Adam Crawshaw
VMXm
Diamond Light Source

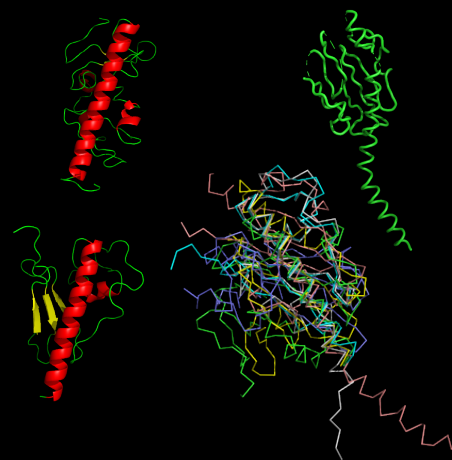


Search for homologues

5V0M	A	POBID	CHAIN	SEQUENCE	MERYRQK--GTFKTYDKNLQKMYFNVTLSKVSP--
5V23	A	POBID	CHAIN	SEQUENCE	MERYRQK--GTFKTYDKNLQKMYFNVTLSKVSP--
1QVE	A	POBID	CHAIN	SEQUENCE	VSDIFSQD--GSCPANTAATAGIEKDTIDNGKYVAK--
1R60	A	POBID	CHAIN	SEQUENCE	VSDIFSQD--GSCPANTAATAGIEKDTIDNGKYVAK--
2H12	A	POBID	CHAIN	SEQUENCE	VTEYLLNH--GQWENMTAGVASSPTDINGKRYKE--
5CFV	A	POBID	CHAIN	SEQUENCE	KNKHMND--TDYSIAEAAFNKGETAMTINGPAWSNII
5IHJ	A	POBID	CHAIN	SEQUENCE	KNKHMND--TDYSIAEAAFNKGETAMTINGPAWSNII
5VAM	A	POBID	CHAIN	SEQUENCE	KNKHMND--TDYSIAEAAFNKGETAMTINGPAWSNII
R209Z1					ALSYYSDT--NRIPVTPDGGTGLSVLETYMESLPDK---
630					ALSYYSDT--NRIPVTPDGGTGLSVLETYMESLPDK---
1792	A	POBID	CHAIN	SEQUENCE	LMYRLND--NRYPTTEQGLRAL--VSRPTVQPERNN--
3620	A	POBID	CHAIN	SEQUENCE	LMYRLND--NRYPTTEQGLRAL--VSRPTVQPERNN--
20PD	A	POBID	CHAIN	SEQUENCE	SKQFILKNPLDQNTIKSKLERFVSGYNNPKIAEK---
20PE	A	POBID	CHAIN	SEQUENCE	SKQFILKNPLDQNTIKSKLERFVSGYNNPKIAEK---
68BK	A	POBID	CHAIN	SEQUENCE	ASGAILLG---KEIVSATPKDTQYDIGFTSTLLD---
350T	A	POBID	CHAIN	SEQUENCE	VAGVQVGLD---KLTPDEANNGISDYGIGGATYS---
3HRV	A	POBID	CHAIN	SEQUENCE	ANGVSLG---KVSADKAKNFTGTAMGIFSPRNS---
350J	A	POBID	CHAIN	SEQUENCE	IONDINN---LDISQQTYYDTPTGVTVGTSTSG---
41XJ	A	POBID	CHAIN	SEQUENCE	SSKDMNEVTKVLENNKCKYFETPKCKSGGTYGATYD



Model preparation



Molecular replacement

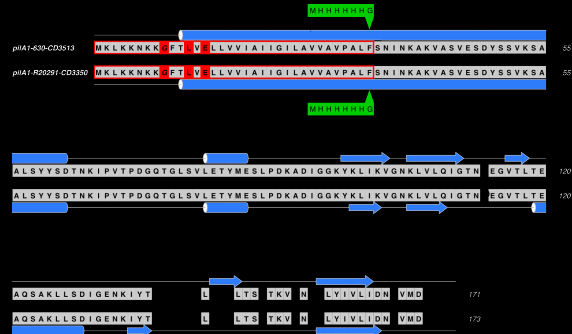
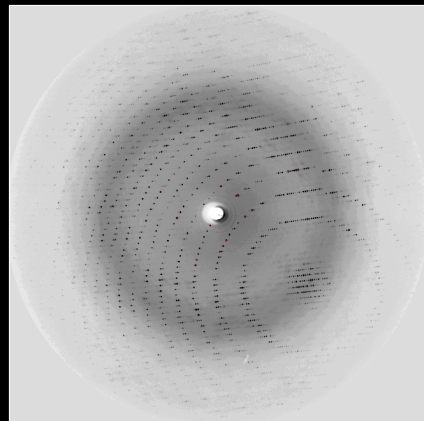


Model Building

Refinement

Completion Validation

Deposition



Search for homologues

PDB/
BLAST

ClustalW

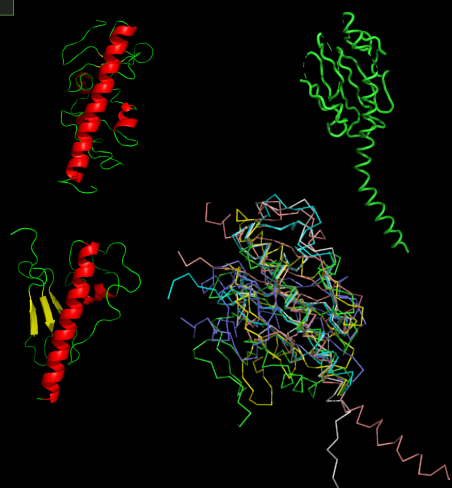
Chainsaw
Sculptor
AMPLE

Sequence alignment and structural diagrams. The top part shows a sequence alignment of two proteins, piliA7-630-CD3513 and piliA7-R20291-CD3300, with a green arrow pointing to a specific residue. Below this, there are two structural diagrams showing the arrangement of alpha-helices and beta-sheets, with labels for residues 171 and 173.



PDB
Phyre2

Model preparation



Molecular replacement

Mr BUMP
Phaser
Molrep
Arcimboldo



Model
Building

Buccaneer
ShelxE
ARPwARP

Coot

Refinement

Refmac

Completion
Validation

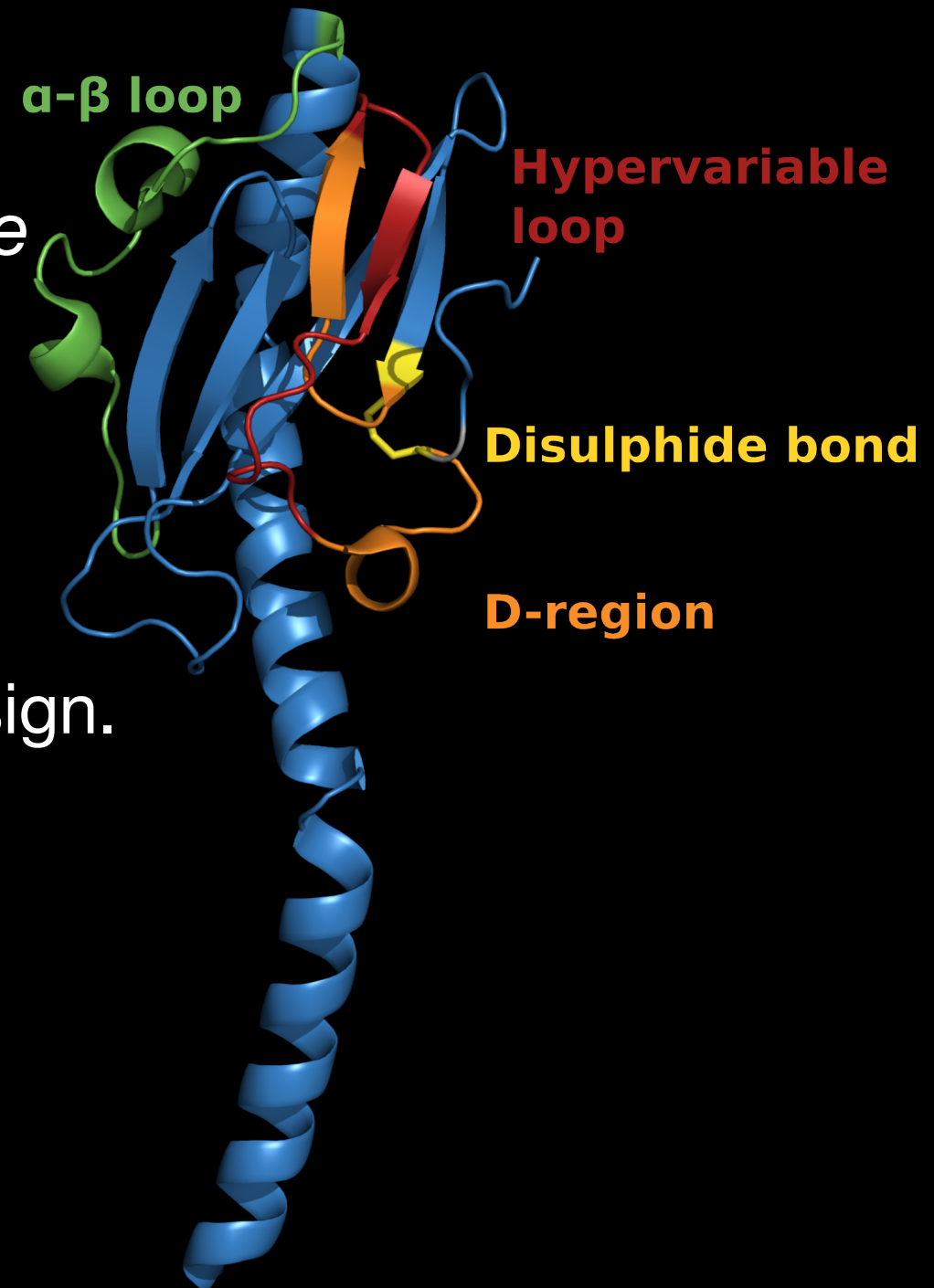
Molprobity
Coot

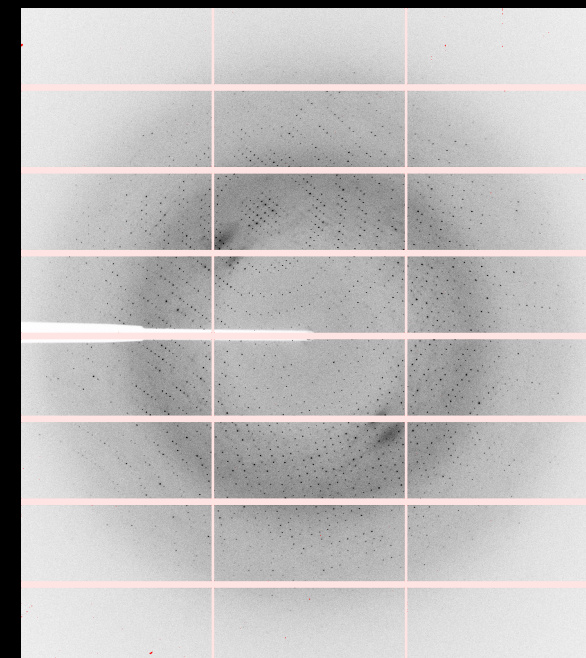
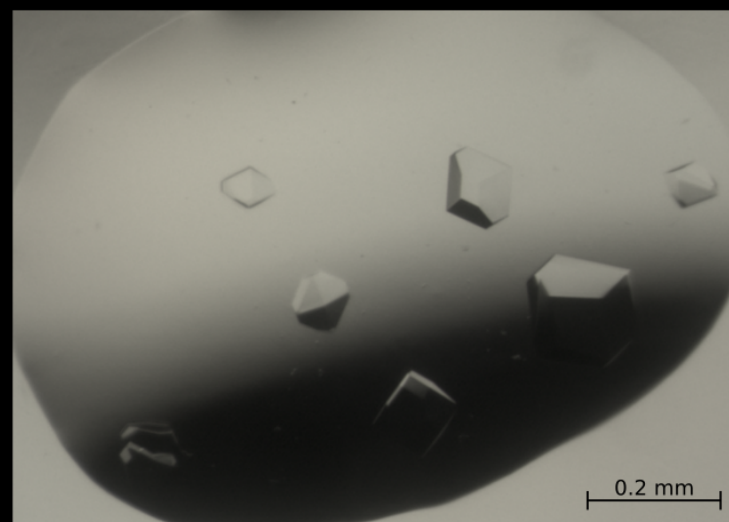
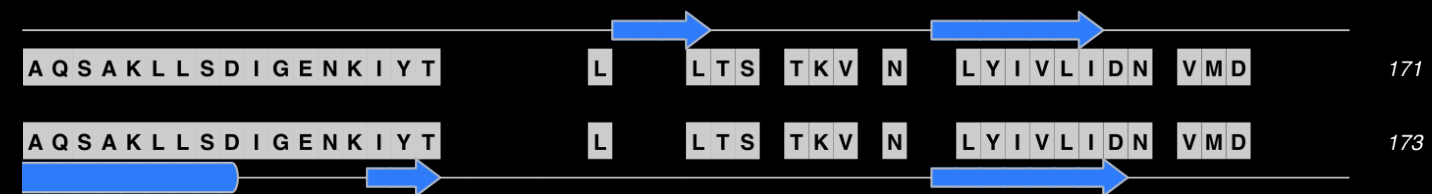
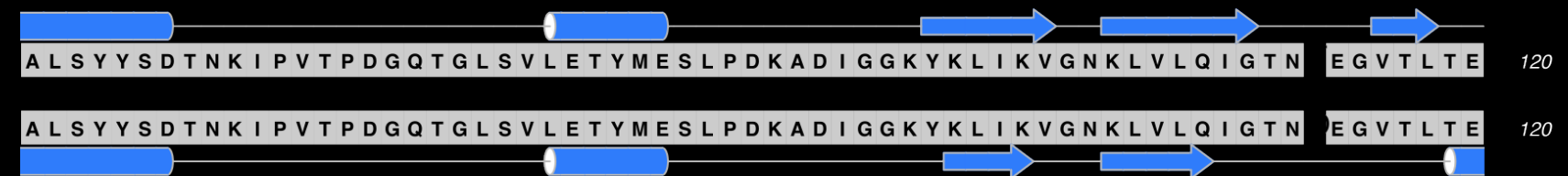
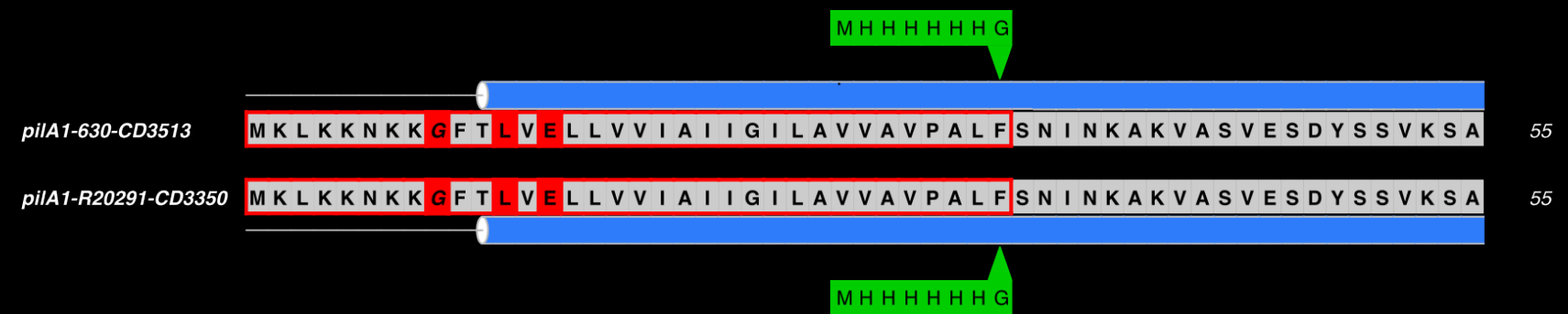
Deposition

PDBe

Biological Context - PilA1

- Type IV Pilin protein.
- Source Gram-positive *Clostridiodes difficile*
- The function of the protein.
 - Forms long fibres via long helix
- Target protein has truncated helix
- We know the sequence from construct design.
- Existing structure models of pilins and pseudopilins (Type II secretion)
- Structural bioinformatics information:
 - Secondary structure prediction





The data

- Single sweep of data
- Indexed and integrated with DIALS
- Scaled with Aimless
- Enantiomorphic ambiguity:
 - $P4_12_12$
 - $P4_32_12$

Source	Diamond Light Source I04-1
Wavelength (Å)	0.92
Resolution range (Å)	51.21-1.65
Unit-cell parameters	
a = b, c	102.57, 104.34
$\alpha = \beta = \gamma$ (°)	90
Spacegroup	$P4_12_12$ or $P4_32_12$
Total No. of reflections	964312 (39779)
No. of unique reflections	67404 (3396)
$\langle I/\sigma(I) \rangle$	12.0 (0.6)
CC1/2	0.999 (0.390)
R _{meas}	0.142 (3.961) [0.05†]
Multiplicity	14.3 (11.7)
Completeness (%)	100.0 (100.0)
Mosaicity(°)	0.12
Wilson B factor (Å ²)	27.1

A Manual approach

Search for homologues

- Using full-length protein sequence
- PDB sequence FASTA and PHMMER search functions
- Structure search tools:
 - HHPred



Ex. - hemoglobin, BRCA1_HUMAN

Search

Examples: [hemoglobin](#), [BRCA1_HUMAN](#)

[Advanced search](#)

[Feedback](#)

PDBE / SEARCH

Advanced search form



Phmmer sequence search

Example: ADKSDLGYTGLTDEQAQELHSVYMSGLWLFSAVAIVAH LAVYIWRPWF

MKLKKNKKGFTLVELLVVIAIIGILAVVAVPALFSNINKAKVASVESDYSSVKSAALSYSDTNKIPVTP
DGQTGLSVLETYMESLPDKADIGGKYKLIKVG NKLVLQIGTNDEGVTLTEAQS AKLLSDIGENKIYTSVT
ADNLGNPLTSNTKVDNKVLYIVLIDNTVMDSTK



Select to add a search field :

Select / type to filter options



Submit

Cancel

EMBL-EBI

[Services](#)

[By topic](#)

☐ [2pil](#)

Crystallographic Structure of Phosphorylated Pilin from Neisseria: Phosphoserine Sites Modify Type IV Pilus Surface Chemistry

Forest KT, Dunham SA, Koomey M, Tainer JA

Mol. Microbiol. (1999) [PMID: [10048019](#) [↗](#)]

Source organism: [Neisseria gonorrhoeae](#) [↗](#)

Assembly composition: protein only structure

Bound ligands: [PT](#) [HTO](#)

Modified residues: [MEA](#) [SEP](#)

Carbohydrate polymer components: **NEW**

Molecule 1 - [GLA\(1\)](#), [NAG\(1\)](#)

 PDBe-KB: [P02974](#) [PRO_0000024155](#)

[3D Visualisation](#)

[Download files](#)

Accuracy: 0.78

Identity percentage: 43.53%

Bit score: 31.24

Score: 31.8

E-value: 0.0000028

Similarity count: 52

Identity count: 37

Similarity percentage: 61.18%

Chain ID: A

Sequence Alignment:

Query :	FTLVELLVVIAIIGILAVVAVPALFSNINKAKVASVESDYSSVKSAALSYYSDTNKIPVTPDGGQTGLSVLETYNESLPDKADIGGKY.KL1
Match :	FTL+EL++VIAI+GILA VA+PA +A+V+ KSA YY + K P ++ T+ V S +DI GKY K +
Target :	FTLIELMIVIAIVGILAAVALPAYQDYTARAQVSEAILLAEGQKSAVTEYYLNHGKWP---ENNTSAGVA-----SP--PSDIKGKYVKEV

[Click here to check for other PDB entries with identical sequence](#)

X-ray diffraction

2.6Å resolution

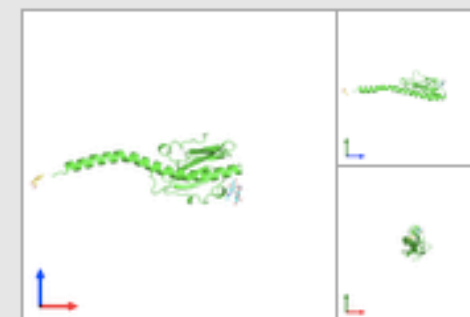
Released: 27 May 1998

Model geometry



Fit model/data

Data not deposited



☐ [2hi2](#)

Crystal structure of native Neisseria gonorrhoeae Type IV pilin at 2.3 Angstroms Resolution

Craig L, Arvai AS, Tainer JA

Mol. Cell (2006) [PMID: [16949362](#) [↗](#)]

Source organism: [Neisseria gonorrhoeae](#) [↗](#)

Assembly composition: protein only structure

Bound ligands: [HTO](#) [OPE](#)

Modified residues: [MEA](#)

Carbohydrate polymer components: **NEW**

Molecule 1 - [DT6\(1\)](#), [GLA\(1\)](#)

 PDBe-KB: [P02974](#) [PRO_0000024155](#)

[3D Visualisation](#)

[Download files](#)

Accuracy: 0.79

Identity percentage: 42.42%

Bit score: 31.13

Score: 31.8

X-ray diffraction

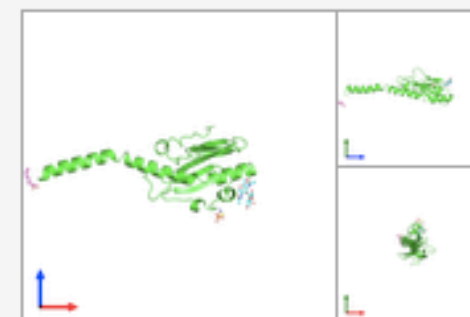
2.3Å resolution

Released: 12 Sep 2006

Model geometry



Fit model/data



[ID](#)
[Date](#)
[Tool](#)


[3212085](#)
[HHPR](#)


HHpred

[Input](#)
[Parameters](#)

```
>PilA1 Peptoclostridium difficile R20291
MKLKKNKKGFTLVLLVIAIIGILAVVAVPALFSNINKAKVASVESDYSSVKSAALSYSDTNKIPVTP
DGQTGLSVLETYMESLPDKADIGGKYKLIKVGKLVQLIGTNDGVTLTEAQSAKLLSDIGENKIYTSVT
ADNLGNPLTSNTKVDNKNVLYIVLIDNTVMDSTK
```

[Paste Example](#)
[Upload File](#)

Protein FASTA

Align two sequences/MSAs

☐

Select structural/domain databases

PDB_mmCIF70_14_Oct 

Select proteomes

Select options

Custom Job ID

Submit



4IXJ_A	Fimbrial protein (Pilin); filament formation, CELL ADHESION; HET: GOL; 1.983A {Clostridium difficile}; Related PDB entri	99.17	4.2e-10	85.38	10.7	98	283
2HI2_A	Fimbrial protein; TYPE IV PILIN, FIBER-FORMING PROTEIN; HET: HTO, GLA, DT6, MEA; 2.3A {Neisseria gonorrhoeae} SCOP: d.24	98.75	1.3e-7	65.17	9.3	63	158
6XXD_L	PilA; Type IV pilin glycosylation twitching; 3.22A {Thermus thermophilus (strain HB27 / ATCC BAA-163 / DSM 7039)}; Relat	98.74	1.3e-7	62.41	8.8	60	125
6GV9_I	Prepilin peptidase-dependent protein D; TYPE IV PILI EHEC T4P; 8.0A {Escherichia coli O157:H7}; Related PDB entries: 6GV	98.67	3.7e-7	61.01	9.6	66	140
2M7G_A	Geopilin domain 1 protein; pilin, PilA, bacterial nanowire, type; NMR {Geobacter sulfurreducens}	98.58	0.0000016	48.05	9.1	61	61
6XXE_O	Uncharacterized protein; Type IV pilin glycosylation twitching; 3.49A {Thermus thermophilus (strain HB27 / ATCC BAA-163	98.53	0.0000015	55.79	9.1	62	111
1OQW_B	Fimbrial protein; Type IV pilin, fiber-forming protein; 2.0A {Pseudomonas aeruginosa} SCOP: d.24.1.1; Related PDB entrie	98.35	0.0000051	55.84	8.6	60	144
3SOK_B	Fimbrial protein; pilus subunit, extracellular, CELL ADHESION; 2.3A {Dichelobacter nodosus} SCOP: d.24.1.0; Related PDB	98.32	0.0000094	55.15	9.3	63	151
5G23_A	TYPE-IV LIKE PILIN TTHA1219; STRUCTURAL PROTEIN, NATURAL COMPETENCE, TYPE; 1.85A {THERMUS THERMOPHILUS}; Related PDB ent	98.08	0.00003	57.33	8.3	58	236
6GUT_B	23S rRNA pseudouridine synthase D,PilA; BACTERIAL PROTEIN, SURFACE ADHESIN, PROT-E; HET: ACT, SO4; 1.63A {Haemophilus in	97.68	0.0001	55.47	5.4	67	262

Collection of homologues

Protein	Organism	PDB Code	Sequence Identity (%)	Method of identification
Psuedopilin	<i>Escherichia coli</i>	3g20	31.82	PDBe BLAST, HHPred
PulG	<i>Klebsiella pneumoniae</i>	1t92	27.18	PDBe BLAST, HHPred
PilJ	<i>Clostridioides difficile</i>	4ixj	19.86	PDBe BLAST, HHPred
TcpA	<i>Vibrio cholerae</i>	1oqv	17	HHpred
PilE1	<i>Neisseria gonorrhoeae</i>	2hi2	17	HHPred, Literature Search
PilE	<i>Francisella tularensis</i>	3soj	14.42	Literature Search
PilX	<i>Neisseria meningitidis</i>	2opd	14.06	PDBe BLAST, HHPred
PilA	<i>Acinetobacter baumannii</i>	5vaw	13.49	Literature Search
FimU	<i>Pseudomonas aeruginosa</i>	4ipu	13.51	HHPred



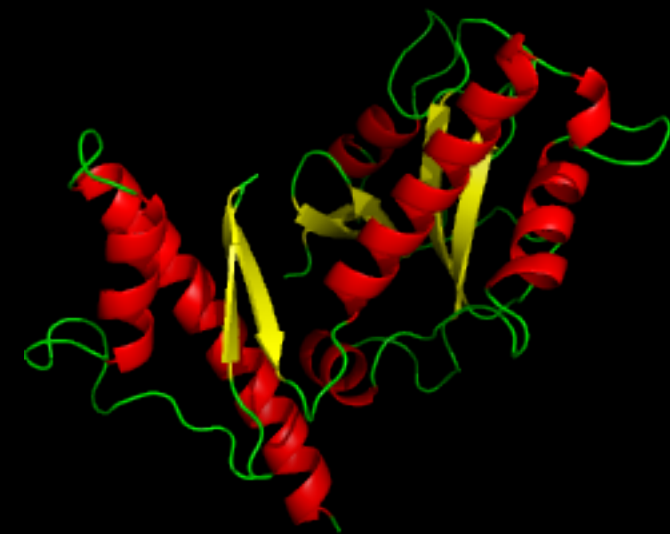
PilX: 2opd



Pseudopilin: 3g20



PulG: 3g20



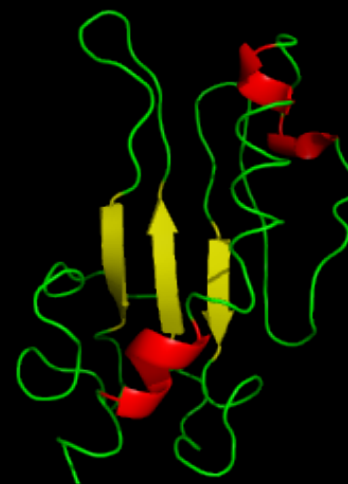
PilJ: 4ixj



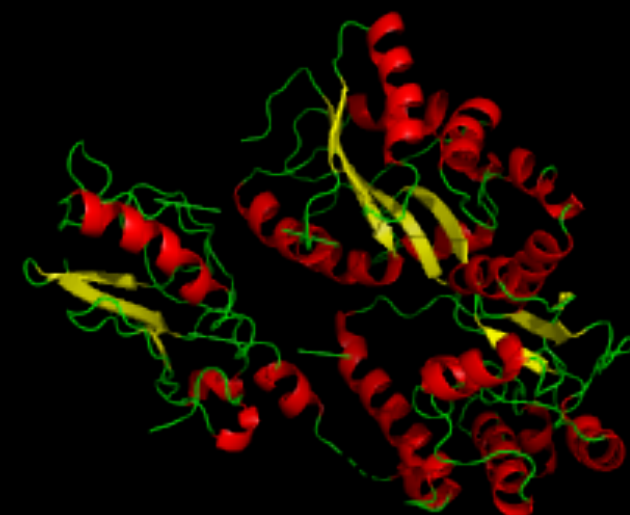
TcpA: 1oqv



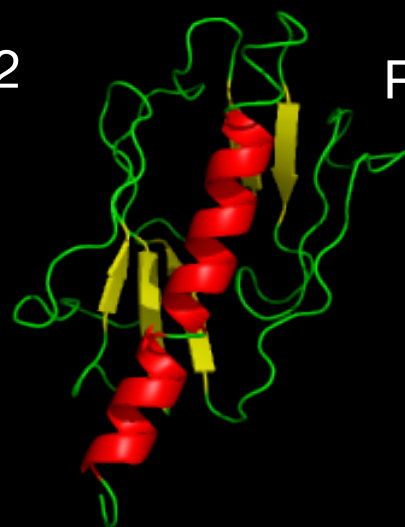
PilE1: 2hi2



PilE: 3soj



PilA1: 5vaw



FimU: 4ipu

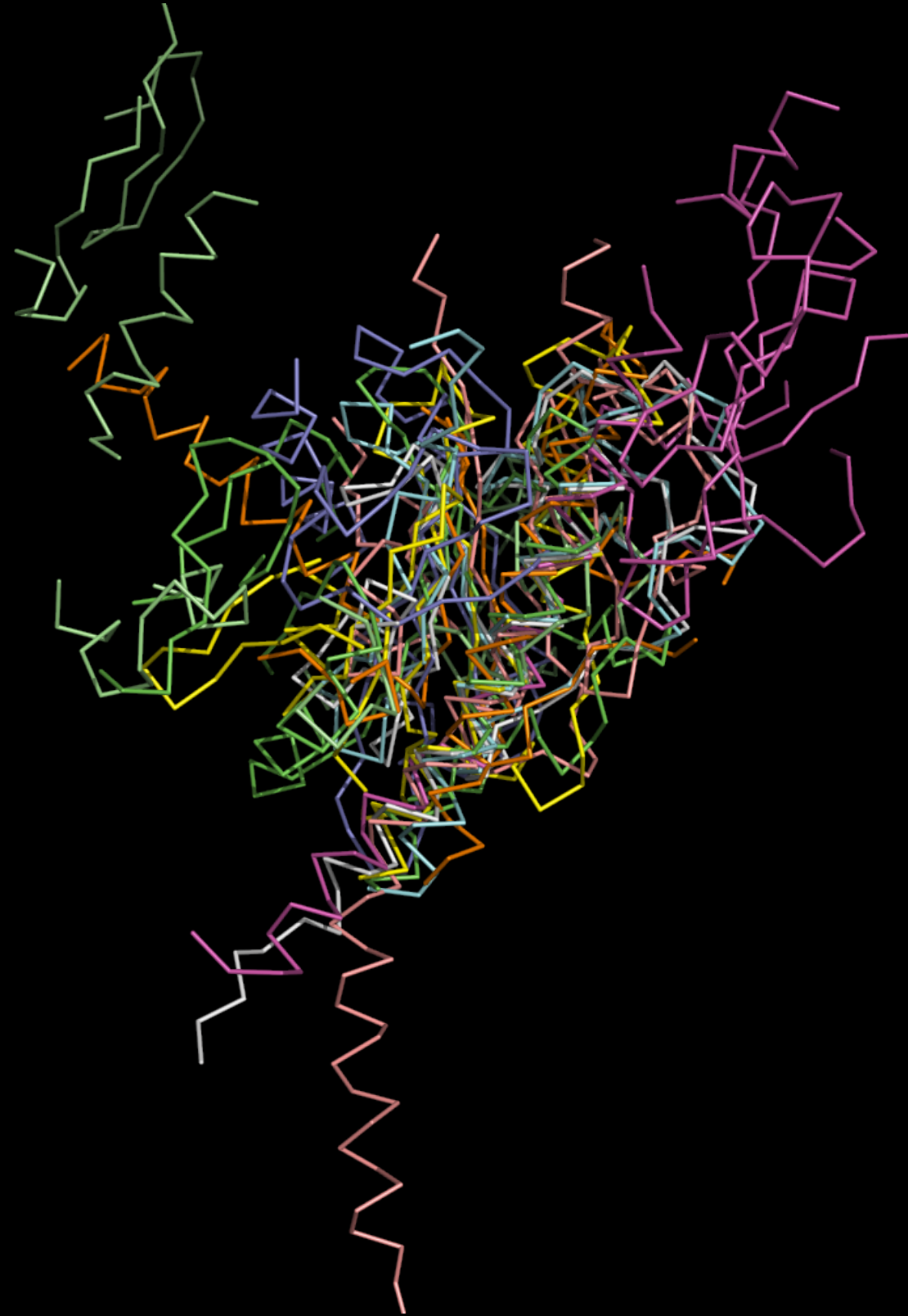
Model preparation

- Select a peptide chain from each search model
- Perform alignment with target for each search model
- Prepare a model using Sculpture

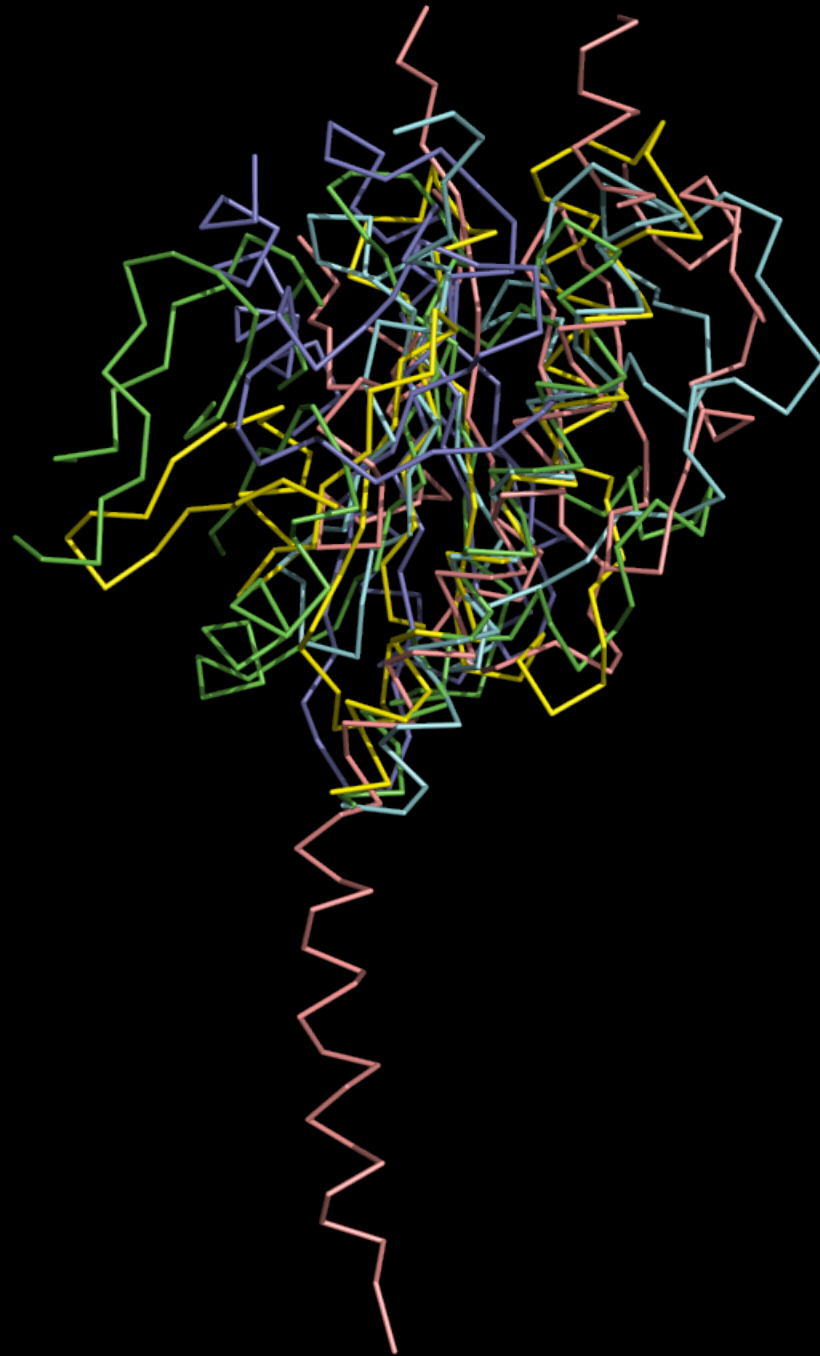


```
cat *_entry.fasta > combined.fasta
```

Build an ensemble



Build an ensemble



Input

Results

Comments

Log stream from phaser.ensembl

► Log stream from phaser.ensembl

Input Data

	Aligned sequence	3 Alignment: Aligned sequences loaded from models_pilins-noCDP1A1.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

Unit-cell contents

- Determine the number of copies within the unit cell
- Mathews co-efficient suggests 4 but 3 could also be likely

job from project: pila_ccp4

Job 3: Define AU contents The job is Finished

AU content - sequences Matthews analysis Run

▼ AU content - sequences

1 copies of pECC52
Description: pECC52

MHHHHHHGSN INKAKVASVE SDYSSVKSA LSYYSDTNKI PVTPDGGTGL SVLETYMESL
PDRADIGGKY KLIKVGKLV LQIGTNDGV TLTEAQSACL LSDIGENKIY TSVTADNLGN
PLTSNTKVDN KVLIVLIDN TVMDSTK

▼ Matthews analysis

Total molecular weight: 15890.7
Cell volume: 1098494.00

Number of copies	Solvent %	Matthews coefficient	Matthews probability
1	85.77	8.64	0.00
2	71.55	4.32	0.02
3	57.32	2.88	0.34
4	43.10	2.16	0.63
5	28.87	1.73	0.01
6	14.65	1.44	0.00
7	0.42	1.23	0.00

Name	Number of copies	Molecular weight
pECC52	1	15890.7

Input Data

Reflections 1 Reflections from AUTOMATIC_DEFAULT_free_1

Output Data

AU contents AU contents

► Job run details

Run

Questions to answer

- Spacegroup ambiguity
 - $P4_12_12$ or $P4_32_12$
- Number of molecules in the asymmetric unit
 - 3 or 4 copies

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

to view jobs and files

Current best solution 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

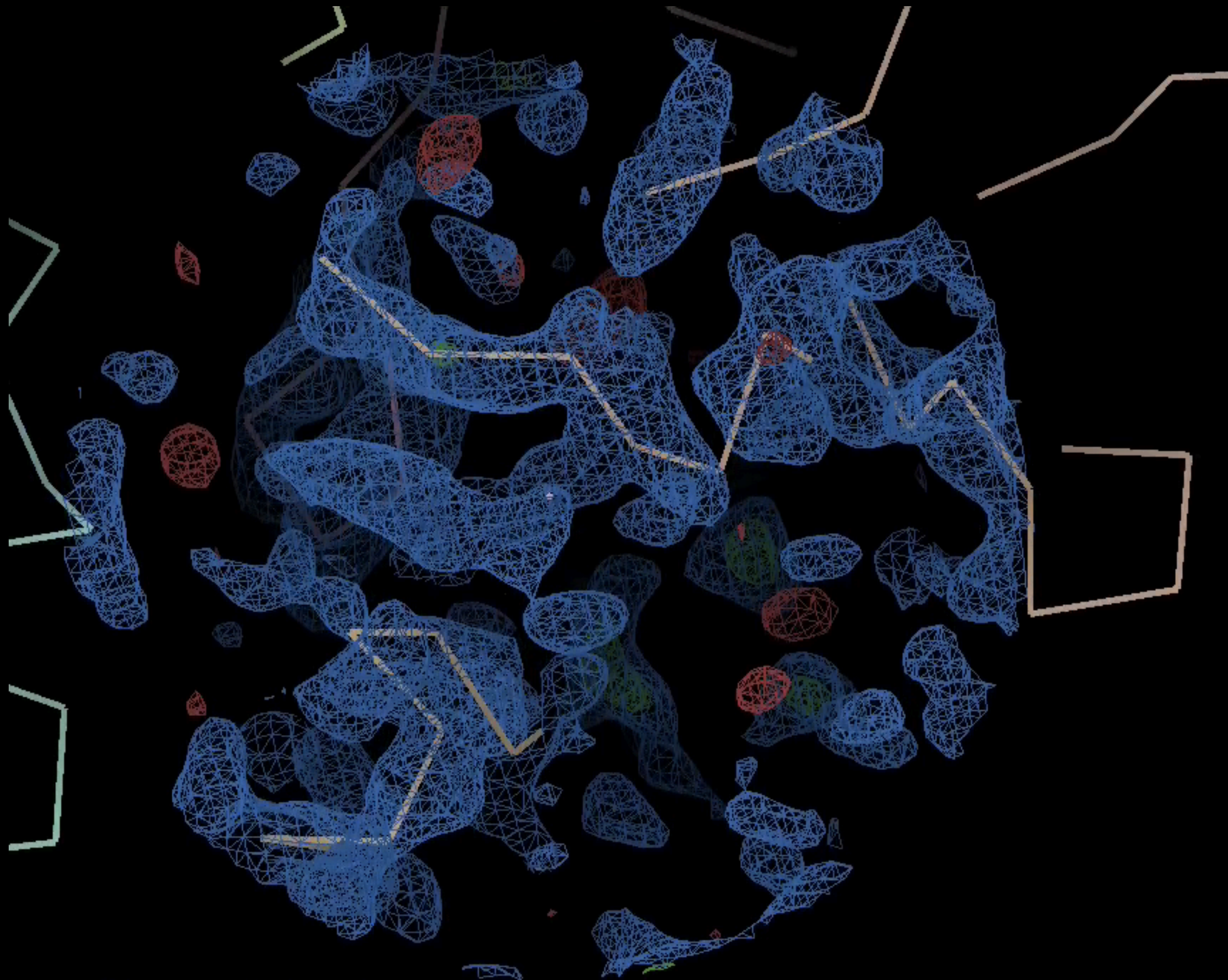
Solutions

Solution	Trans_Z-score	Refined_Trans_Z-score	Initial_LLГ	Refined_LLГ	Initial_Rfactor	Refined_Rfactor	Clashes
0	18.0	6.0	29.61	29.61	58.76	58.76	2.00
1	26.0	5.0	22.39	22.39	58.93	58.93	0.00
2	31.0	8.0	21.24	21.24	59.03	59.03	9.00
3	24.0	6.0	21.09	21.09	58.89	58.89	7.00
4	21.0	6.0	20.71	20.71	58.82	58.82	7.00
5	35.0	9.0	20.53	20.53	59.11	59.11	10.00
6	8.0	5.0	20.47	20.47	58.58	58.58	6.00
7	31.0	5.0	20.38	20.38	59.12	59.12	3.00
8	31.0	5.0	20.24	20.24	59.04	59.04	7.00
9	59.2	59.0	19.52	19.52	59.17	59.17	5.00

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

Success?

- TFZ score:
 - > 8 solution found
 - 5-7 possible solution
 - < 5 unlikely solved
- LLG
 - Value increases during process.
 - > 100
- How does the map look?



Automated Approach

AMPLE / MrBUMP

- AMPLE: automated search model building.
- MrBUMP: complete molecular replacement pipeline
 - Searches for suitable models and prepares them.
 - MR utilising Phaser and MOLREP
 - Phase improvement using SHELXE
 - Full model building with Arp/Warp and Buccaneer

MrBUMP

\$\$

\$TABLE: Final Summary:

\$GRAPHS: Final Rfactor/Rfree vs model :N:1,5,6:

: Buccaneer Final Rfactor/Rfree vs model :N:1,7,8:

: SHELXE CC vs model :N:1,9:

: SHELXE Average Chain Length :N:1,10:

\$\$

Model	Model_Name	MR_Program	Solution_Type	Phaser_LLQ	Phaser_TFZ	final_Rfact	final_Rfree	Bucc_final_Rfact	Bucc_final_Rfree	ARP_final_Rfact	ARP_final_Rfree	SHELXE_CC	SHELXE_ACL
-------	------------	------------	---------------	------------	------------	-------------	-------------	------------------	------------------	-----------------	-----------------	-----------	------------

\$\$

1	4tsm_C_CHNSAW	MOLREP	GOOD	-- --	0.292 0.304 0.265 0.283	-- --	45.410 43						
2	4tsm_C_CHNSAW	PHASER	GOOD	2741.000 4.500	0.298 0.313 0.265 0.285	-- --	45.400 79						
3	1qve_A_CHNSAW	PHASER	POOR	55.000 6.500	0.531 0.533 0.529 0.537	-- --	10.060 11						
4	2m7g_A_CHNSAW	MOLREP	POOR	-- --	0.548 0.539 0.526 0.556	-- --	9.960 12						
5	3g20_B_CHNSAW	PHASER	POOR	23.000 6.200	0.543 0.541 0.533 0.560	-- --	10.280 11						
6	1zwt_A_CHNSAW	MOLREP	POOR	-- --	0.525 0.541 0.523 0.561	-- --	10.820 10						
7	3gn9_A_CHNSAW	MOLREP	POOR	-- --	0.538 0.541 0.523 0.545	-- --	11.280 10						
8	6gms_A_CHNSAW	PHASER	POOR	93.000 8.100	0.553 0.542 0.542 0.556	-- --	10.230 12						
9	3fu1_A_CHNSAW	PHASER	POOR	56.000 6.600	0.542 0.543 0.526 0.545	-- --	10.640 13						
10	3hrv_A_CHNSAW	PHASER	POOR	62.000 6.400	0.532 0.543 0.525 0.549	-- --	11.370 11						
11	3hrv_A_CHNSAW	MOLREP	POOR	-- --	0.519 0.546 0.526 0.554	-- --	10.780 10						
12	4d40_B_CHNSAW	PHASER	POOR	6.000 6.500	0.564 0.546 0.543 0.561	-- --	9.380 11						
13	3g20_B_CHNSAW	MOLREP	POOR	-- --	0.529 0.547 0.520 0.557	-- --	10.260 9						
14	3fu1_A_CHNSAW	MOLREP	POOR	-- --	0.543 0.548 0.522 0.557	-- --	9.620 11						
15	1t92_B_CHNSAW	PHASER	POOR	46.000 6.500	0.543 0.548 0.526 0.558	-- --	10.850 11						
16	1zwt_A_CHNSAW	PHASER	POOR	64.000 6.300	0.527 0.549 0.530 0.559	-- --	10.940 10						
17	1oqv_C_CHNSAW	MOLREP	POOR	-- --	0.522 0.550 0.524 0.551	-- --	10.190 12						
18	1t92_B_CHNSAW	MOLREP	POOR	-- --	0.538 0.551 0.520 0.549	-- --	11.700 12						
19	5v0m_A_CHNSAW	MOLREP	POOR	-- --	0.556 0.551 0.531 0.571	-- --	8.970 9						
20	2m7g_A_CHNSAW	PHASER	POOR	50.000 6.600	0.544 0.553 0.518 0.573	-- --	9.550 13						
21	5o2y_A_CHNSAW	PHASER	POOR	45.000 6.300	0.551 0.553 0.533 0.562	-- --	9.950 10						
22	3gn9_A_CHNSAW	PHASER	POOR	63.000 6.300	0.551 0.553 0.523 0.554	-- --	11.190 10						
23	5wda_O_CHNSAW	MOLREP	POOR	-- --	0.541 0.553 0.530 0.557	-- --	10.230 10						
24	4d40_B_CHNSAW	MOLREP	POOR	-- --	0.563 0.554 0.550 0.570	-- --	11.350 9						
25	5v23_A_CHNSAW	PHASER	POOR	100.000 7.300	0.543 0.554 0.526 0.555	-- --	10.280 10						
26	2kep_A_CHNSAW	MOLREP	POOR	-- --	0.526 0.554 0.525 0.562	-- --	11.590 10						
27	6gms_A_CHNSAW	MOLREP	POOR	-- --	0.552 0.558 0.521 0.554	-- --	10.240 10						
28	1qve_A_CHNSAW	MOLREP	POOR	-- --	0.550 0.559 0.529 0.564	-- --	10.200 11						
29	5o2y_A_CHNSAW	MOLREP	POOR	-- --	0.533 0.559 0.526 0.556	-- --	10.080 10						
30	5v0m_A_CHNSAW	PHASER	POOR	87.000 6.400	0.562 0.561 0.544 0.566	-- --	10.150 10						
31	4us7_A_CHNSAW	MOLREP	POOR	-- --	0.571 0.561 0.537 0.571	-- --	10.120 12						
32	1oqv_C_CHNSAW	PHASER	POOR	40.000 6.400	0.550 0.562 0.531 0.553	-- --	10.560 10						
33	2kep_A_CHNSAW	PHASER	POOR	43.000 6.200	0.536 0.566 0.531 0.553	-- --	10.470 9						
34	4us7_A_CHNSAW	PHASER	POOR	90.000 7.600	0.551 0.574 0.522 0.572	-- --	10.370 12						
35	5wda_O_CHNSAW	PHASER	POOR	22.000 6.500	0.560 0.579 0.538 0.562	-- --	9.730 10						
36	5v23_A_CHNSAW	MOLREP	POOR	-- --	0.579 0.585 0.529 0.562	-- --	10.120 12						

\$\$

#####

<!--SUMMARY_END-->

AMPLE

- Input:
 - Scaled data
 - Sequence
 - Decoys (fragments) generated by Quark

<i>ensemble_name</i>	<i>MR_program</i>	<i>Solution_Type</i>	<i>PHASER_LLG</i>	<i>PHASER_TFZ</i>	<i>REFMAC_Rfact</i>	<i>REFMAC_Rfree</i>	<i>SHELXE_CC</i>	<i>SHELXE_ACL</i>
c1_t24_r3_polyala	PHASER	POOR	42.0	7.5	0.5573	0.5527	10.56	11

Another way...



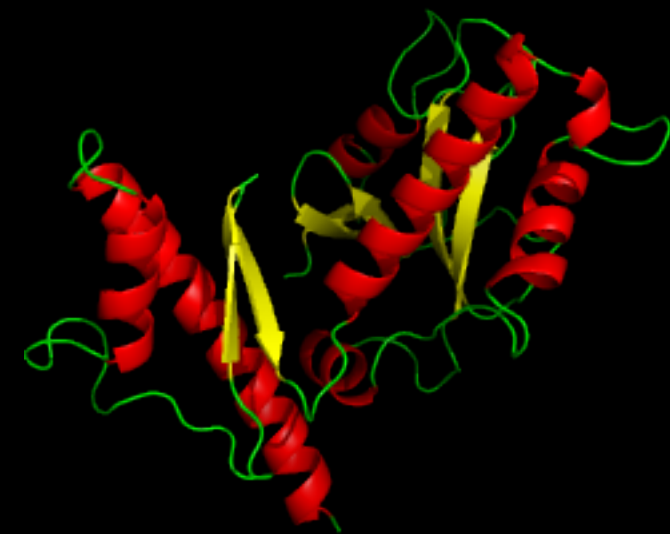
PilX: 2opd



Pseudopilin: 3g20



PulG: 3g20



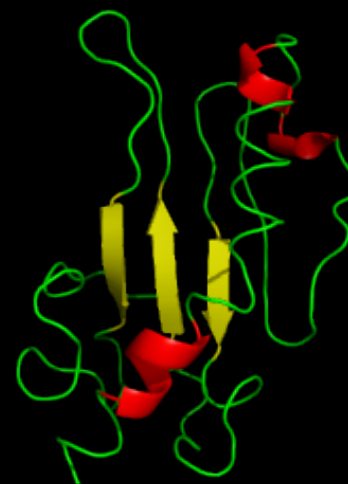
PilJ: 4ixj



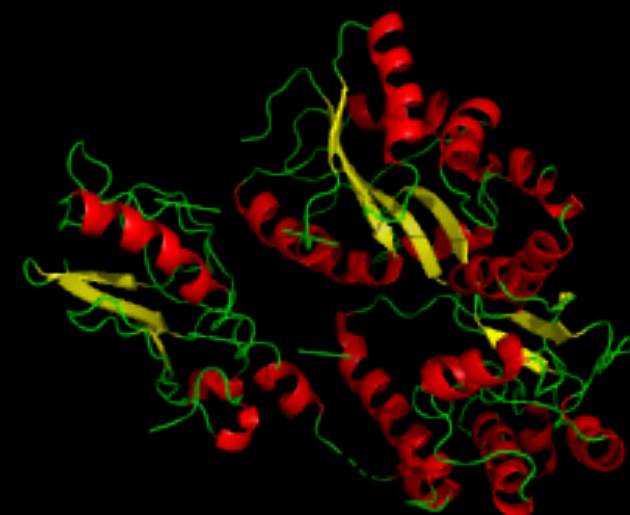
TcpA: 1oqv



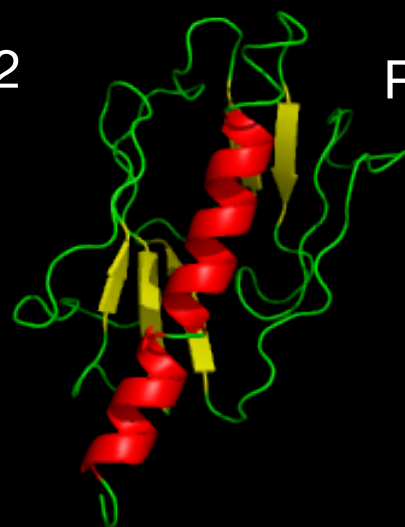
PilE1: 2hi2



PilE: 3soj



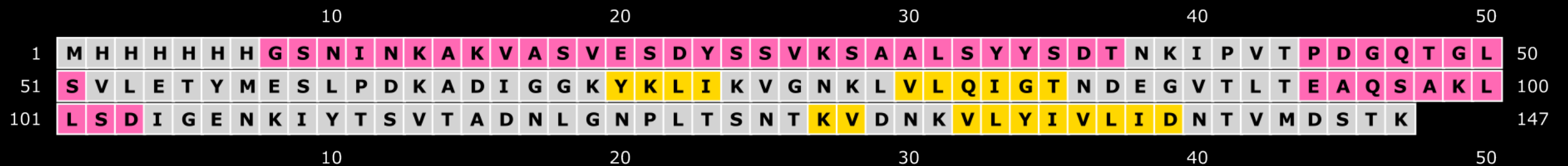
PilA1: 5vaw



FimU: 4ipu

Ab initio MR

- Arcimboldo - employs Phaser and SHELXE
 - Data extends beyond 2 Å
- Secondary structure prediction: PSI-PRED
- 30 residue helix



Arcimboldo

Input data

Advanced data

Job title

Arcimboldo



Use data from job

No

as input below..

Run ARCIMBOLDO

LITE

on

this machine

☐ Run in coil coiled mode

Input data



Reflections

pila1-arcimboldo 1 /pila/4

Asymmetric unit contains

3

components of molecular weight

15890.0

Daltons

Model

Use

one or more copies of a helix

assuming rmsd from target

0.2

Å

Search for

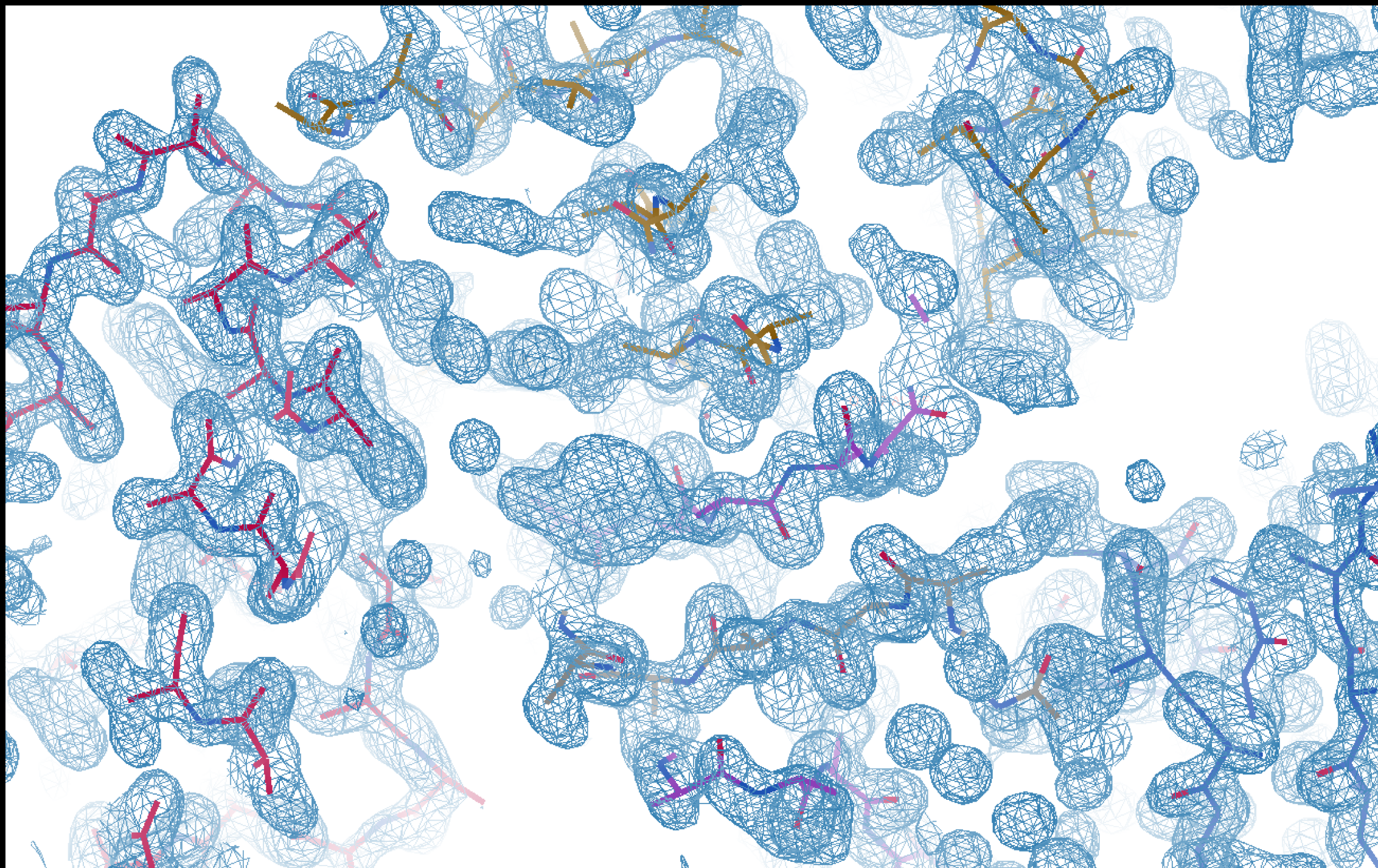
3

copies of a helix containing

30

residues

Arcimboldo



Model Building

Results

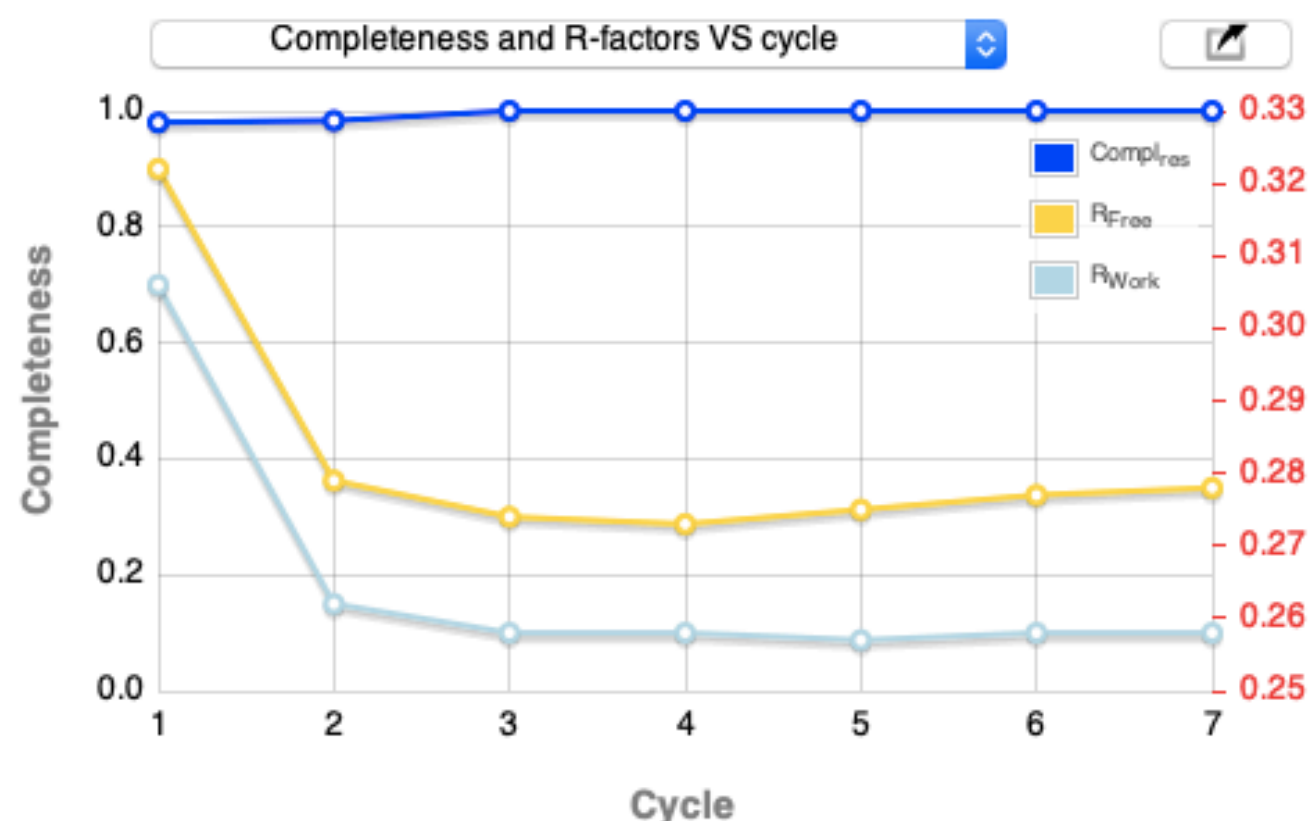
The final model is taken from cycle 4 as this had the lowest free-R factor.

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 0.26, and the free-R factor is 0.27. The RMS bond deviation is 0.009 Å.
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	138
Number of fragments	3
R_{Work}	0.258
R_{Free}	0.273
$\text{RMS}_{\text{Bonds}}$	0.009
$\text{RMS}_{\text{Angles}}$	1.901

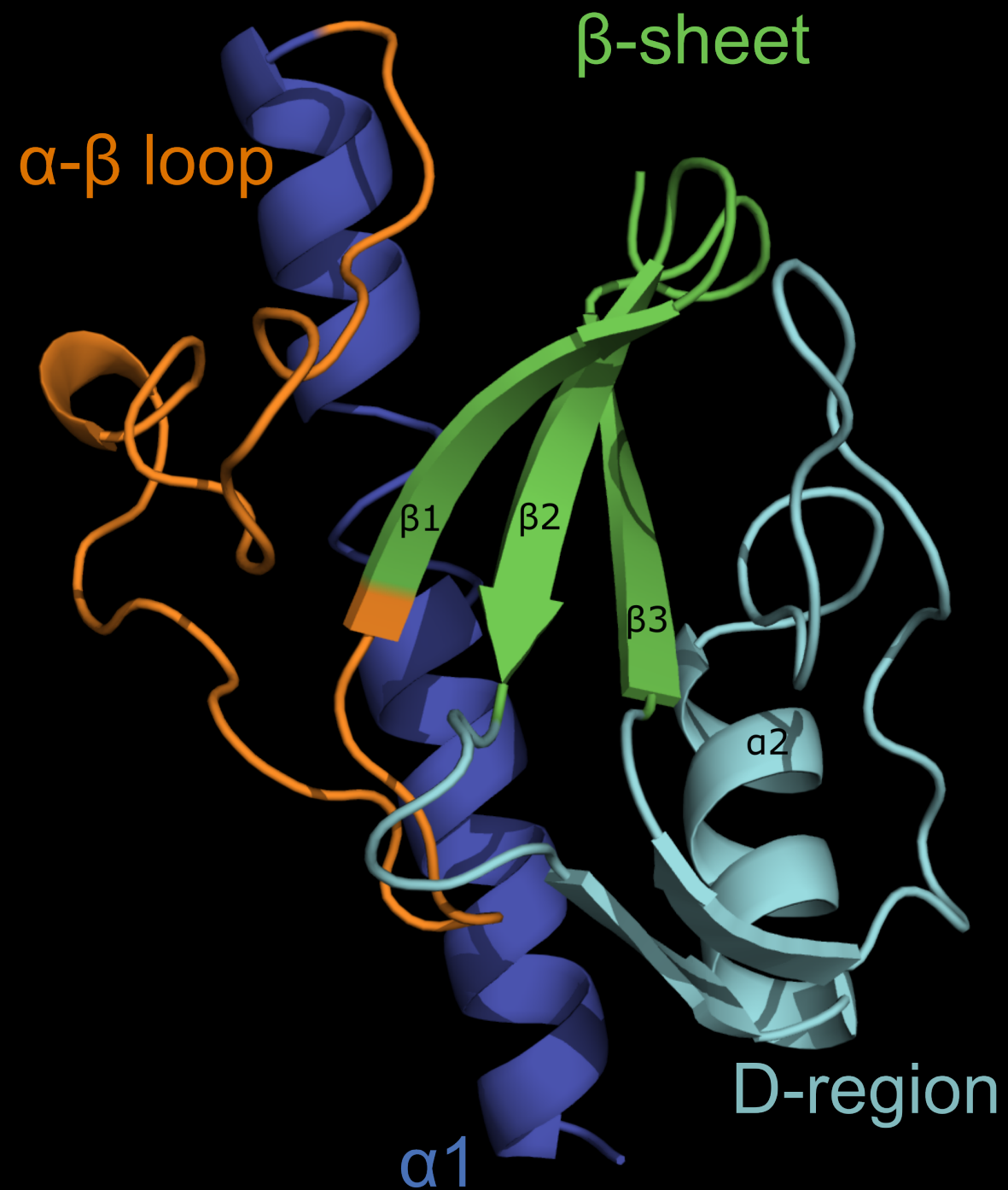


Refinement

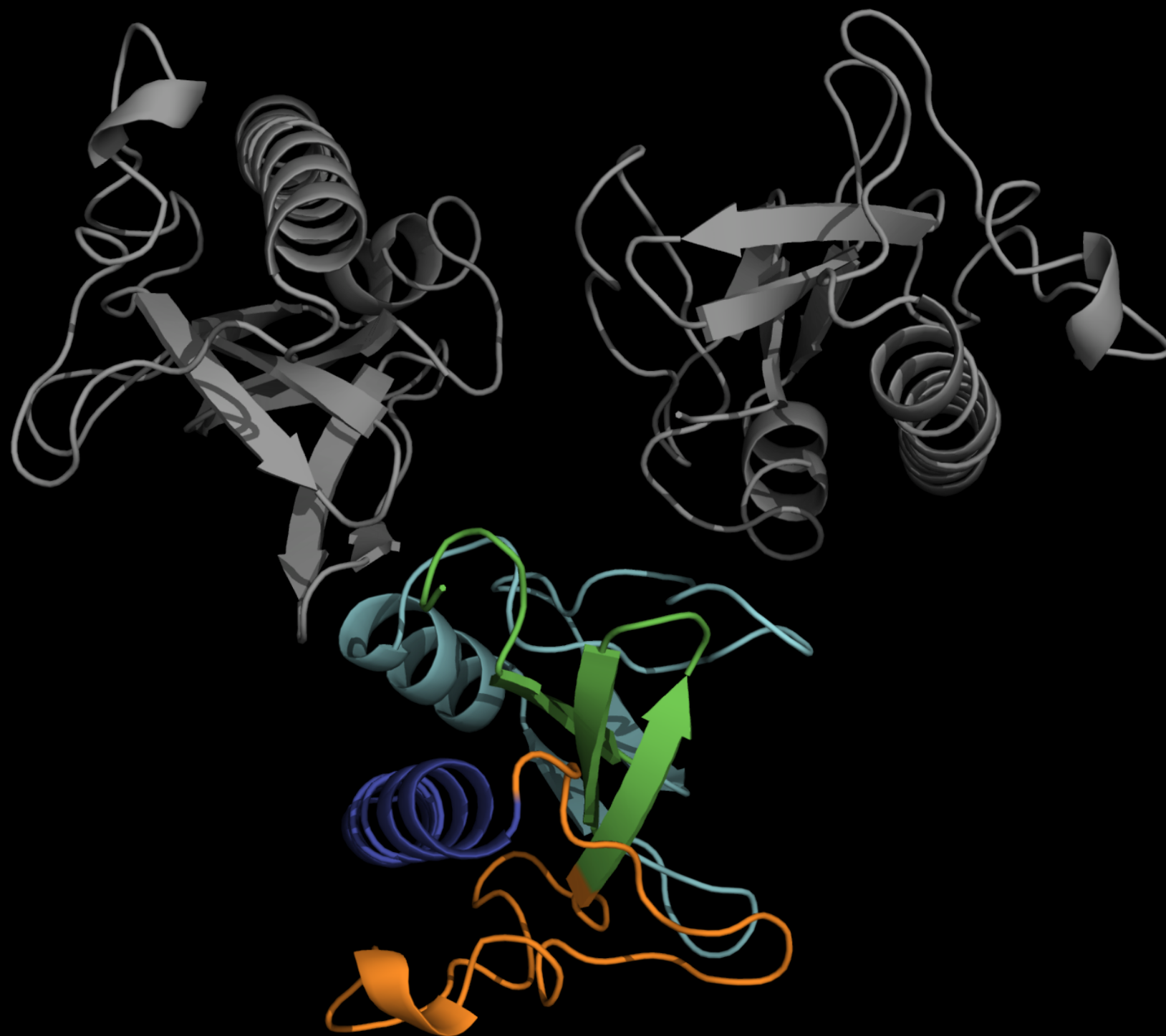
- Multiple rounds of refinement with Refmac and Coot.
- Weighting term of 0.1 applied
- Water placement using Coot
- Clashes and outliers detected with Molprobity
- Validated using the PDBe validation server

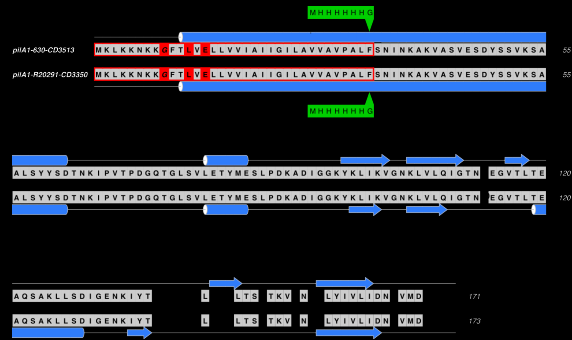
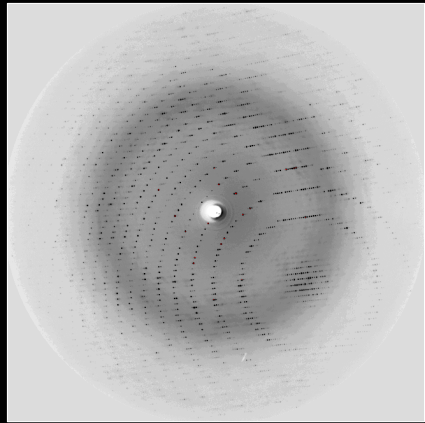
Spacegroup	P4 ₁ 2 ₁ 2
Rwork (%)	19.4
Rfree (%)	21.4
No. Of non-H atoms	
Protein	3069
Solvent	254
No. Of residues	400
R.M.S deviations	
Bond angles (°)	1.30
Bond length (Å)	0.01
Average B factor (Å ²)	
Overall	28.00
Protein	28.53
Water	34.38
Ramachandran plot	
Most favoured regions (%)	97.21
Allowed regions (%)	2.57
Outliers (%)	0
Clashscore	0.48
MolProbity clashscore	0.81

Final Model



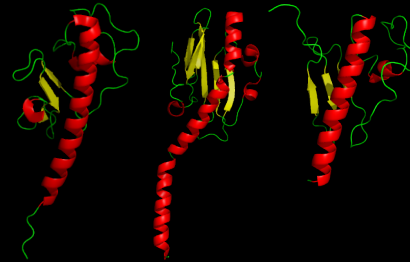
Final Model



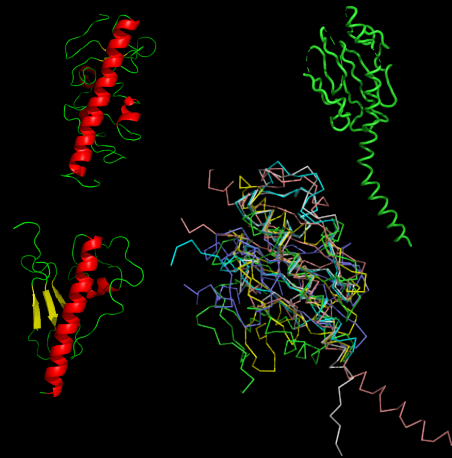


Search for homologues

5V0M A|PDBID|CHAIN|SEQUENCE MERYRQK--GTFKTYDKNLQKMYFNVTLSKVSP--
5V23 A|PDBID|CHAIN|SEQUENCE MERYRQK--GTFKTYDKNLQKMYFNVTLSKVSP--
1QVE A|PDBID|CHAIN|SEQUENCE VSDIFSQD--GSCPANTAATAGIEKDTIDNGKYVAK--
1R60 A|PDBID|CHAIN|SEQUENCE VSDIFSQD--GSCPANTAATAGIEKDTIDNGKYVAK--
2H12 A|PDBID|CHAIN|SEQUENCE VTEYLLNH--GQWENMTAGVASSPTDINGKRYKE--
5CFV A|PDBID|CHAIN|SEQUENCE KNKHMND--TDYSIAEAAFNKGETAMTINGPAWSNII
5IHJ A|PDBID|CHAIN|SEQUENCE KNKHMND--TDYSIAEAAFNKGETAMTINGPAWSNII
5VAM A|PDBID|CHAIN|SEQUENCE KNKHMND--TDYSIAEAAFNKGETAMTINGPAWSNII
R209Z1 ALSVYSDT--NRIPVTPDGGTGLSVLETYMESLPDK--
630 ALSVYSDT--NRIPVTPDGGTGLSVLETYMESLPDK--
1792 A|PDBID|CHAIN|SEQUENCE LMYKLNN--SRVPTDQGLAL--VSAFSAEPHAR--
3620 A|PDBID|CHAIN|SEQUENCE LMYRLNN--NRYPTEQGLAL--VSRPTVQPERN--
20PD A|PDBID|CHAIN|SEQUENCE SKQFILKNPLDQQTISKLERFVSGYNNPKIAEK--
20PE A|PDBID|CHAIN|SEQUENCE SKQFILKNPLDQQTISKLERFVSGYNNPKIAEK--
68BK A|PDBID|CHAIN|SEQUENCE AEAILEG---KEIVSATPKDTQYDIFGTSTLLD--
350T A|PDBID|CHAIN|SEQUENCE VAGVQLG---KLTPDEANQISGDTIGIGGATYS--
3HRV A|PDBID|CHAIN|SEQUENCE ANGLVSLG---KVSADKAKNFTGTAMGIFSPRNS--
350J A|PDBID|CHAIN|SEQUENCE IONDINNN---LDISQQTYYDTPTGVTVGTSTG--
41XJ A|PDBID|CHAIN|SEQUENCE SSKDKNEVTKVLENNQKTYFETPKCKSGGTYGATYD



Model preparation



Molecular replacement



Model Building

Refinement

Completion Validation

Deposition

Summary

- Many avenues to pursue...
- Take into account biological context, particularly with model building
- Try the automated pipelines
- If data extend to 2 Å attempt *ab initio MR*

Acknowledgements

Newcastle University

Paula Salgado

Arnaud Baslé

Carys Davies

Imperial College

Neil Fairweather

Edward Couchman